Machine learning tools for pattern recognition in polar climate science

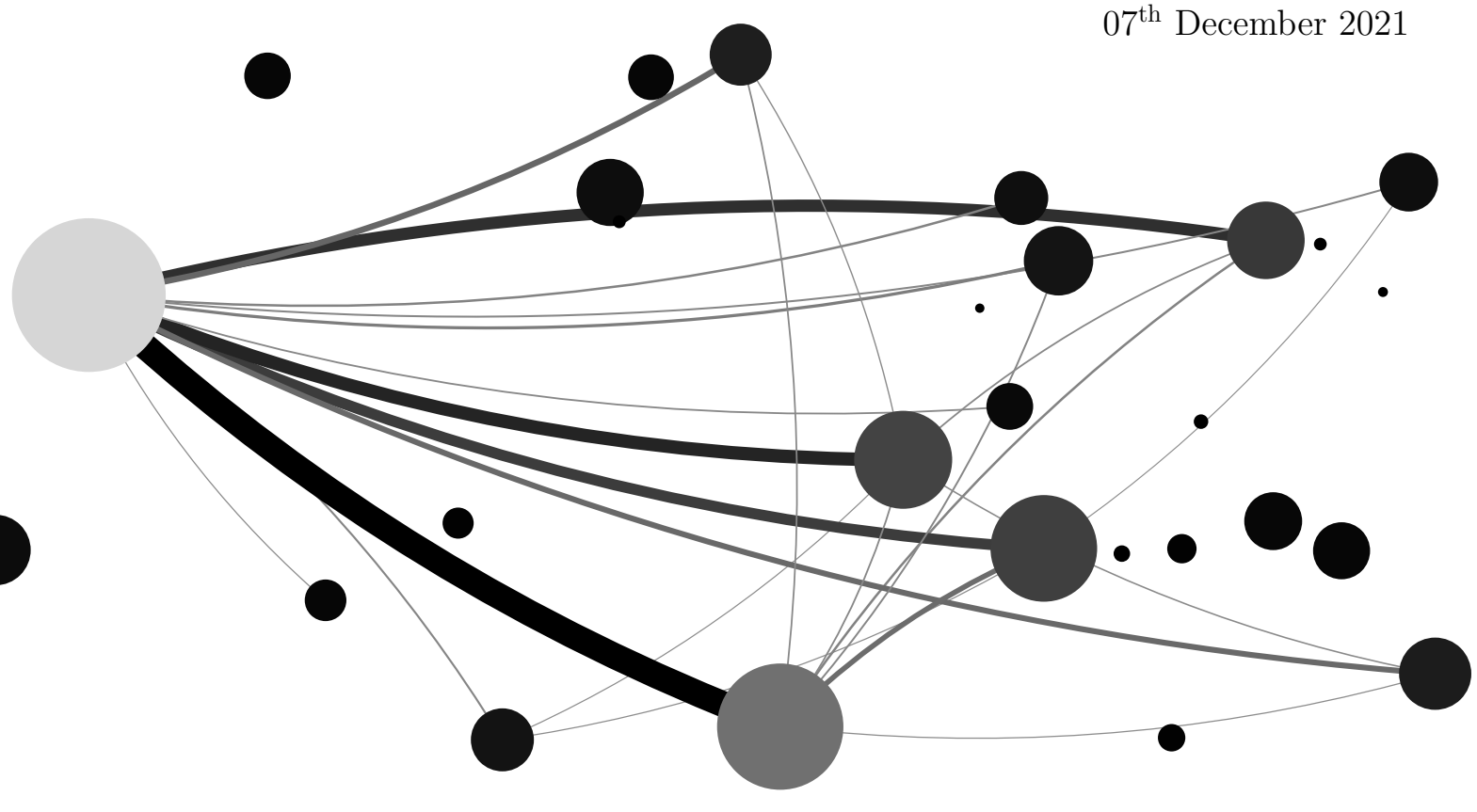
William J. Gregory

A thesis submitted in partial fulfilment
of the requirements of the degree of

Doctor of Philosophy

University College London
Department of Earth Sciences

07th December 2021



I, William Gregory, confirm that the work presented in this thesis is my own. All images and results were created through Python code which I have written myself, and which is publicly available within various repositories at the following GitHub profile: https://github.com/William-gregory. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

07/12/2021

Abstract

This thesis explores the application of two novel machine learning approaches to the study of polar climate, with particular focus on Arctic sea ice. The first technique, complex networks, is based on an unsupervised learning approach which is able to exploit spatio-temporal patterns of variability within geospatial time series data sets. The second, Gaussian Process Regression (GPR), is a supervised learning Bayesian inference approach which establishes a principled framework for learning functional relationships between pairs of observation points, through updating prior uncertainty in the presence of new information. These methods are applied to a variety of problems facing the polar climate community at present, although each problem can be considered as an individual component of the wider problem relating to Arctic sea ice predictability.

In the first instance, the complex networks methodology is combined with GPR in order to produce skilful seasonal forecasts of pan-Arctic and regional September sea ice extents, with up to 3 months lead time. De-trended forecast skills of 0.53, 0.62, and 0.81 are achieved at 3-, 2- and 1-month lead time respectively, as well as generally highest regional predictive skill (> 0.30) in the Pacific sectors of the Arctic, although the ability to skilfully predict many of these regions may be changing over time.

Subsequently, the GPR approach is used to combine observations from CryoSat-2, Sentinel-3A and Sentinel-3B satellite radar altimeters, in order to produce daily pan-Arctic estimates of radar freeboard, as well as uncertainty, across the 2018–2019 winter season. The empirical Bayes numerical optimisation technique is also used to derive auxiliary properties relating to the radar freeboard, including its spatial and temporal (de-)correlation length scales, allowing daily pan-Arctic maps of these fields to be generated as well. The estimated daily freeboards are consistent to CryoSat-2 and Sentinel-3 to within < 1 mm (standard deviations < 6 cm) across the 2018–2019 season, and furthermore, cross-validation experiments show that prediction errors are generally ≤ 4 mm across the same period.

Finally, the complex networks approach is used to evaluate the presence of the winter Arctic Oscillation (AO) to summer sea ice teleconnection within 31 coupled climate models participating in phase 6 of the World Climate Research Programme Coupled Model Intercomparison Project (CMIP6). Two global metrics are used to compare patterns of variability between observations and models: the Adjusted Rand Index and a network distance metric. CMIP6 models generally over-estimate the magnitude of sea-level pressure variability over the north-western Pacific Ocean, and under-estimate the variability over the north Africa and southern Europe, while they also under-estimate the importance of regions such as the Beaufort, East Siberian and Laptev seas in explaining pan-Arctic summer sea ice area variability. They also under-estimate the degree of covariance between the winter AO and summer sea ice in key regions such as the East Siberian Sea and Canada basin, which may hinder their ability to make skilful seasonal to inter-annual predictions of summer sea ice.

Impact statement

Concerted efforts have been made in recent decades to better understand how climate-change-induced sea ice loss is manifested within the various regions which make up the Arctic Ocean, and furthermore how these changes affect our ability to predict the state of sea ice cover on time-scales ranging from a few days to a few years. This sea ice prediction problem is multifaceted, and is an important area of research for a variety of end-users including indigenous communities, industrial shipping and coastal resupply vessels, as well as eco-tourism. The results presented in this thesis represent contributions to this on-going area of research.

On the one hand, our ability to make reliable sea ice predictions depends on our understanding of how different climatic processes combine together to drive sea ice variability on different time-scales. Secondly, we require high-resolution data sets relating to these different climatic processes in order to provide the best possible observations for our forecast model to learn from in a statistical sense, and also to initialise our model with, in a dynamical sense. Finally, our forecast model itself must be sophisticated enough so that it can separate the key predictive patterns from noise within our training data, and capture the large signature of variability that exists within the observational sea ice record.

At the time of publication of the results presented in chapter 3, the outlined forecast model had produced the highest forecast skill score of pan-Arctic September sea ice extent anomalies (up to 3-month lead time), of any statistical or dynamical model presented in the literature at that time. This model has also been used each year since 2019 to provide periodic submissions of seasonal September sea ice extent forecasts to the Sea Ice Prediction Network's on-going Sea Ice Outlook (SIO) program. The SIO is a community-led, platform which is constantly changing to meet end-user needs, and combines individual sea ice forecasts to provide the most probable picture of end-of-summer ice conditions, as well as better understanding the drivers of inter-annual sea ice variability.

The results presented in chapter 4 outline the first observation-based daily estimates of pan-Arctic radar freeboard, which can be considered the base product for deriving sea ice thickness. This product could potentially be a big step forward for assessing how synoptic scale weather patterns drive sea ice thickness variability on sub-weekly timescales, and furthermore it could also provide an uplift to users of dynamical sea ice forecast models and sea ice data assimilation systems, given that key predictive information of the sea ice cover resides in its thickness distribution.

Finally, the results presented in chapter 5 provide critical information relating to perhaps why the latest generation of coupled climate models fail to accurately reflect one of the major atmospheric processes that drives summer sea ice variability, and subsequently, why they under-perform at operational seasonal predictions of summer Arctic sea ice, thus potentially bridging the gap between potential and operational dynamical model forecasts.

Acknowledgements

I feel undeniably privileged to say that this PhD has been one of the most enjoyable experiences of my life so far, and I can only say this thanks to a key group of people who have been instrumental in aiding its completion, and to whom I would like to give a special mention to here.

First, my deepest thanks go to my supervisors Michel Tsamados and Julienne Stroeve. Michel, you are the epitome of a research scientist, with your meticulous attention to detail, scientific integrity, and overall zeal for developing new ideas. Working through theory and code, and bouncing countless ideas off the walls with you in the early stages, is everything I had hoped scientific research would entail, and I am forever grateful for your dedication and seemingly endless capacity for patience and kindness. I want to thank you for taking a chance on someone without any previous experience in this research area, for the emotional support you provided through the difficult times, and most of all for your friendship.

Julienne, thank you for making this project possible and for your invaluable insights along the way. Your ability to always see the bigger picture when either Michel or I were lost in the small details has often been the key to pushing various aspects of this thesis forward, and I can only hope that I will carry all that I have learned from your guidance over to future research projects.

I would also like to extend my thanks to Peter Sollich. Peter, I can't express my gratitude enough for your patience in teaching Michel and I all the theory that went into our forecasting paper. Learning everything from scratch was a challenge that, at times, felt insurmountable, and any journey I make into the world of statistics and machine learning beyond this thesis, I owe all to you.

To everyone who has been part of the UCL staff, PhD, and research cohort; Rosie, Harry, Carmen, Tom Johnson, Almu, Blaise, Lucia, Frans, Gab, Giles, Mark, Chris, Nic, Will, Bobby, Kate, Jonathan, Alan, Danuta, Jen, Leisa. Through all the various times we've either exchanged ideas and code, or whenever I've bugged you for technical or admin support, or simply all the times we may have shared a pint or two, each one of you has contributed in some way to this thesis, and I am truly grateful.

To Giulia, Ashley, Jamie, and James Gilbert. The friendships and memories I have gained in each of you over the years are something I deeply cherish and are what I will always look back on fondly when I think about the PhD. Each of you holds truly endearing qualities which never fail to lift my spirits whenever I see any one of you, and which always make me gravitate to you in a room full of people. I hold you all closely in my heart and it has been a pleasure to have shared parts of this journey with each of you.

To Isobel and Becks, the two of you are ultimately responsible for shaping a significant portion of my PhD experience and for giving me some of the fondest memories of the last four years. Isobel, what more could one ask for out of a PhD than to come to work every day and be in the company of someone so kind, and so full of life and unparalleled empathy. You made me feel so welcome among the UCL PhD cohort

when I started, you taught me how to code, and also what a privilege it has been to have co-written a paper together. You kept my head above water so many times both academically and personally, and I'll never forget it.
Becks, you have been a substantial influence on me over the last few years. Being in your company was always an effective means of washing away any feelings of burden from life and the PhD, through your gentle spirit and outlandish sense of humour. Your ability to dive into new experiences with gusto and an open mind is something I've always found inspirational, and I try every day to live by your example. Thank you for the best of times during the PhD.
To James and Luke, thank you for always providing a hub of serenity in any place that we have lived together over the last six years – even Bromley. There has never been a day where I didn't look forward to coming home at the end of it, and that is all down to the two of you. I have never for a moment tired of your company, and I couldn't have asked for two better brothers to live with over the course of my PhD and through the various covid-related lockdowns. The memories we have created together, along with Emma, Edie, and of course Abbey and Silky, will stay with me forever.
To my family, thank you for supporting me in any decision I've ever made, including the decision to start over and do a PhD. There has never been any pressure to do anything other than what I enjoy and what makes me happy, and for that I am eternally grateful. Mum, these past few years have been tough on you personally, and it is a testament to your character that even at the lowest points you still find the space to worry about others. You are one of the most selfless people I have ever known and I'm proud of everything that you have overcome and achieved.
To some my oldest and dearest friends who have been with me since time immemorial; Callum, James Davies, Tom and Bill Robinson, Ash, Blake, Frank, Fabio, Sylvain, and Rachel. Thank you for always supporting me through the highs and lows, and I look forward the best of times still to come with each and every one of you.

This thesis was made possible through funding from the U.K. Natural Environment Research Council (NERC) London Doctoral Training Partnership (DTP), grant NE/L002485/1.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past four decades, the inexorable growth in technology and subsequently the availability of Earth-observation and model data has been unprecedented. With each passing day, measurements from satellite, airborne, and ground-based missions extend our observational record of the changes occurring across the various components of the climate system, and technological advancements allow us to model numerous realisations of the Earth's climate, with increasingly high resolution and fidelity. Hidden within these data are the fingerprints of the physical processes that govern climate variability over a wide range of spatial and temporal scales, and it is the task of the climate scientist to separate these patterns from noise. Given the wealth of data now at our disposal, machine learning methods are becoming the tools of choice in climate science for a variety of applications ranging from data assimilation, to sea ice feature detection from space. In this thesis, the application of machine learning methods to the study of polar climate is explored, with particular focus on Arctic sea ice, however it is worth noting that the tools which will be presented are quite generic and can, in principle, be applied to most geospatial time series data sets.
This chapter presents a general overview of Arctic sea ice, discussing recent and projected changes, while taking insights from both observational and model data. Chapter 2 then introduces the machine learning concepts which will be used in this thesis, and furthermore outlines the research objectives pertaining to each subsequent chapter.

1.1 Arctic sea ice

Arctic sea ice is a major component of the northern-hemisphere polar climate system. Climatologically, it acts as a barrier which both reflects incoming solar radiation (Light et al., 2008) and regulates the rate of energy exchange between the

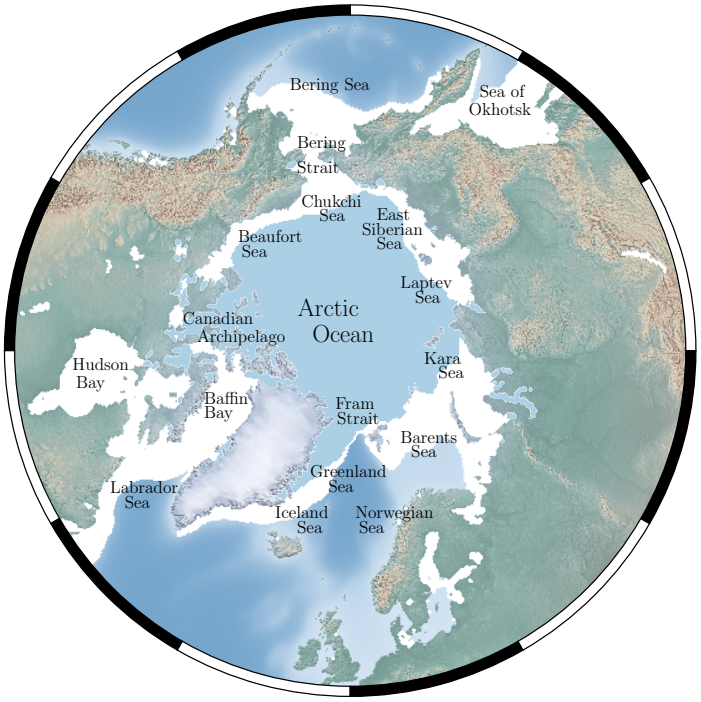


Figure 1.1: Map of the Arctic seas with the average seasonal maximum sea ice extent (March) between 1979 and 2020 overlain in cream (15.3×10^6 km^2), and the average seasonal minimum sea ice extent (September) between 1979 and 2020 overlain in blue (6.0×10^6 km^2). Data from the National Snow and Ice Data Center (NSIDC; Cavalieri et al. 1996).

atmosphere and ocean (Maykut 1978; Kurtz et al. 2011), and furthermore acts to re-circulate salt around the upper ocean (Toudal & Coon, 2001). Culturally, it is an integral platform for travel and subsistence (Aporta 2009; Dammann et al. 2018; Panikkar et al. 2018), and ultimately for maintaining strong relationships between local communities and the surrounding landscape (Segal et al., 2020). Ecologically, it provides a natural habitat for various endemic species (Moore & Huntington, 2008) and is thought to play a crucial role in the efficiency of both primary (Stroeve et al., 2021) and secondary (Sakshaug et al. 1994; Stirling 1997) productivity rates.

The launch of the Nimbus-7 NASA satellite in late 1978 was a remarkable step forward for polar monitoring. Fitted with a multichannel microwave radiometer, its ability to measure passive microwave emissions from orbit meant that, for the first time, pan-Arctic (basin-wide) measurements up to 84.5° N of Arctic sea ice concentration (the area fraction of sea ice for a given pixel) could be obtained every two days (Cavalieri et al., 1996). This then advanced to daily observations up to 87.2° N in July 1987, and finally 89.18° N in January 2008 with the launch of the Defense Meteorological Satellite Program (DMSP) F-8 and F-17 satellites respectively. This has provided a continuous record of sea ice concentration, sea ice area (concentration multiplied by pixel area), sea ice extent (generally defined as the sum of pixel areas for which concentration values exceed 15%), and subsequently the dramatic seasonal

cycle of sea ice growth and retreat that occurs in the Arctic each year (see Figure 1.1). During this cycle, the ice cover shrinks to over half its winter extent each summer, and grows to over twice its summer extent each winter, by a magnitude which is primarily controlled by two opposing thermodynamic feedback mechanisms that operate over either the spring–summer period, or autumn–winter period. As the sun rises in the spring and atmospheric temperatures begin to increase, the Arctic melt season begins. Incoming (short-wave) solar radiation is absorbed by the sea ice and areas of open ocean, and the melting of snow on the top-surface of the sea ice facilitates the formation of pools of surface melt water, known as *melt-ponds*. These melt-ponds dramatically alter the surface-albedo characteristics of the sea ice, which initiates a positive feedback mechanism as the lower-albedo melt-ponds are able to absorb more short-wave radiation, which acts to precipitate increased melt until the underlying ocean is exposed. This then leads to larger heat uptake by the ocean, causing enhanced warming and further melt of sea ice from the bottom (Curry et al., 1995). Following large summer sea ice melt events however, the timing of ice growth is delayed (Markus et al., 2009), and subsequently the ocean is able to radiate more heat to the atmosphere (Holland et al., 2011). When ice growth does begin, the new ice is relatively thin, and so a stabilising negative feedback mechanism is established by which the sea ice is able to grow more easily over winter (Notz & Bitz, 2017). New sea ice which grows during this winter period, and which subsequently melts in the following summer is known as *first-year* ice, whereas ice which persists through the melt season is known as *multi-year* ice. Satellite-derived ice motion vectors allow us to track the life cycle of sea ice as it circulates around the Arctic Ocean, thus giving an indication of ice age (i.e., first-year or multi-year ice; Maslanik et al. 2011; Tschudi et al. 2019). On long time-scales this dynamic ice circulation is primarily driven by two major ocean currents that exist within the Arctic (Figure 1.2). These include the Beaufort Gyre; an anticyclonic (clockwise) current which circulates large quantities of freshwater around the western Arctic (Proshutinsky et al. 2009; Giles et al. 2012) and simultaneously advects sea ice in and out of the Beaufort Sea (Kwok et al. 2013; Petty et al. 2016a), and the Transpolar Drift Stream; an ocean current which transports sea ice from the Eurasian seas out of the Fram Strait southward – a passage where approximately 10% of the Arctic's sea ice area budget inside the Arctic basin is exported each year (Smedsrud et al., 2017). The strength of these ocean currents is largely modulated by large-scale atmospheric pressure patterns, which drive surface winds during different phases of the North Atlantic Oscillation (Kwok, 2000), and subsequently the Arctic Oscillation (see section 1.1.3). Figure 1.3a shows the average spatial distribution of autumn–spring Arctic sea ice age derived from ice motion vectors, where we can see that the older multi-year ice (> 1 year) is generally situated along the western Arctic, north of Greenland and within the Beaufort and Chukchi seas, whereas first-year ice (< 1 year) generally makes up the

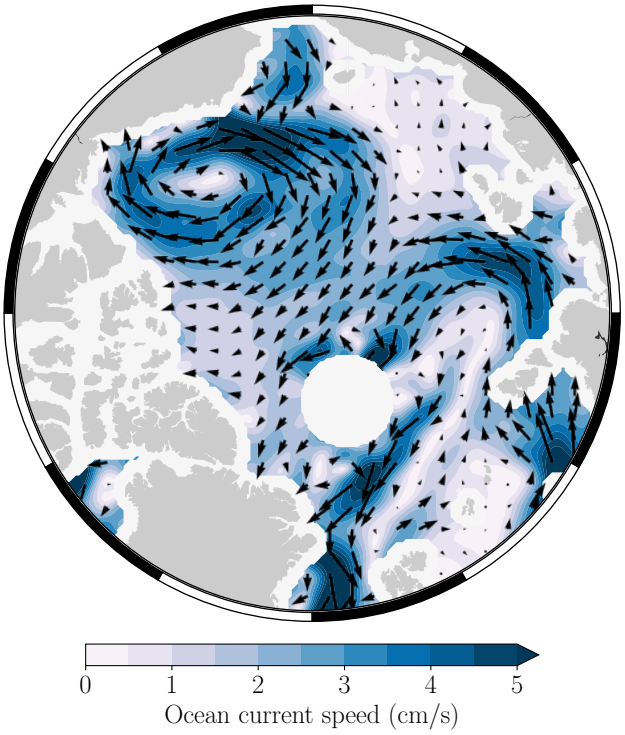


Figure 1.2: The average circulation of Arctic Ocean surface currents between 2011 and 2020. These data contain a polar hole at $\sim$ 88° N, which the limit of coverage for this particular satellite (CryoSat-2) due to its orbit inclination. Data from the Center for Polar Observation and Modelling (CPOM; internally sourced).

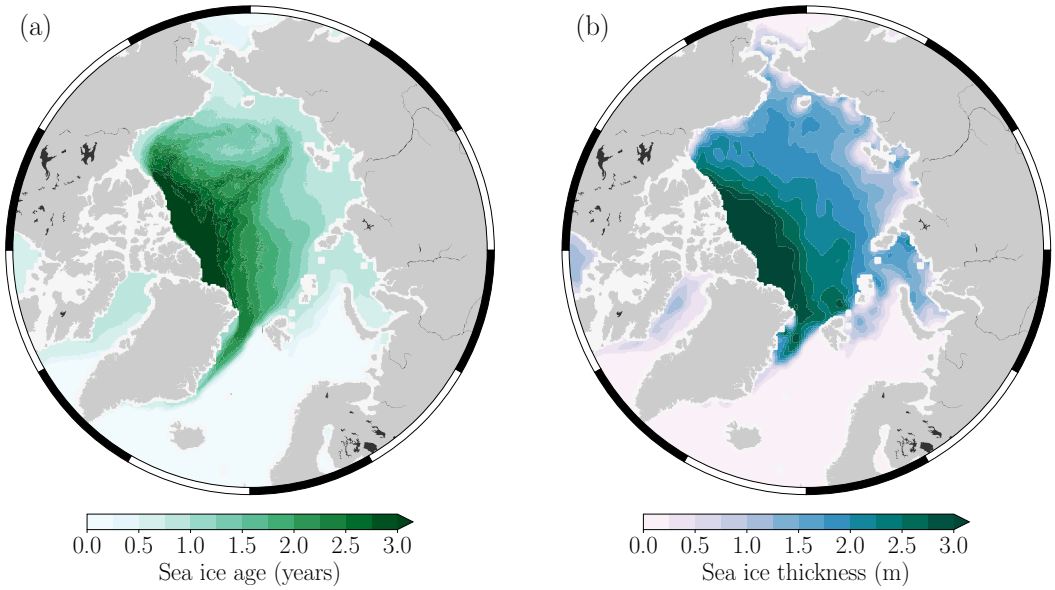


Figure 1.3: The average autumn–spring (October–April) Arctic (a) sea ice age and (b) thickness between 2010/11 and 2019/20. Sea ice age and thickness data from NSIDC (Tschudi et al., 2019) and the Alfred Wegener Institute (AWI; Ricker et al. 2017) respectively.

Eurasian and peripheral seas. In Figure 1.3b we also notice how thicker ice generally coincides with older multi-year ice and vice versa (Tschudi et al. 2016; Liu et al. 2020). In this example, the average autumn–spring thickness of multi-year ice between 2010/11 and 2019/20 was 2.2 metres, while first-year ice was 0.85 metres. Basin-wide observations of sea ice thickness were made possible by the launch of satellite radar and laser altimeters, beginning with the European Space Agency's (ESA) ERS-1 and -2 radar altimeters in 1991, which provided thickness estimates up to 81.5° N (e.g., Laxon et al. 2003). Since then, several altimeters have been instrumental in advancing the spatio-temporal availability of thickness observations, to the current date where basin-wide estimates are available up to 88° N for each month between October and April (see section 4.1 for further details).

1.1.1 Recent trends and variability

After the formation of the Intergovernmental Panel on Climate Change (IPCC) in 1988, a worldwide collaborative initiative began, with a focus on improving our understanding of Earth's climate system, as well as the rate and scale of climate change and its possible drivers. Research facilities from around the globe periodically provide evidence based on observational data and model studies, which are then collated into a series of assessment reports e.g., AR4 (Solomon et al., 2007), AR5 (Stocker et al., 2013), and AR6 (Arias et al., 2021), which summarise the changes occurring across the various components of the climate system. This section outlines some of the key changes in the Arctic sea ice cover that have occurred over recent decades, taking particular examples from observational data sets.

Figure 1.4 shows the daily evolution of pan-Arctic sea ice extent between 1979 and 2020, along with the long-term trend associated with each day. Noticeably, the observed trend across all days is negative, meaning that the Arctic sea ice cover has been diminishing over the past four decades. The largest rate of decline is seen in the summer and autumn months, with an average maximum trend of $\sim -84,000$ km^2 per year in October, and the smallest rate of decline in May ($\sim -36,000$ km^2 per year). Looking at the regional and seasonal patterns of sea ice concentration trends (Figure 1.5), we see that sea ice is in decline in nearly all regions of the Arctic (for trends with a magnitude larger than 0.5% per year), and for all seasons. The dominant regions of winter–spring sea ice loss have occurred in the Sea of Okhotsk, Bering, Labrador, Barents and Greenland seas, and in summer–autumn the Beaufort, Chukchi, East Siberian, Kara and Laptev seas. Previous studies (Onarheim et al. 2018; Stroeve & Notz 2018) analysed regional trends in March and September sea ice extent, and found that September trends in the East Siberian Sea explain as much as 27% of the total trend in pan-Arctic September sea ice extent, and that the March trends in the Sea of Okhotsk and Barents Sea together explain 54% of

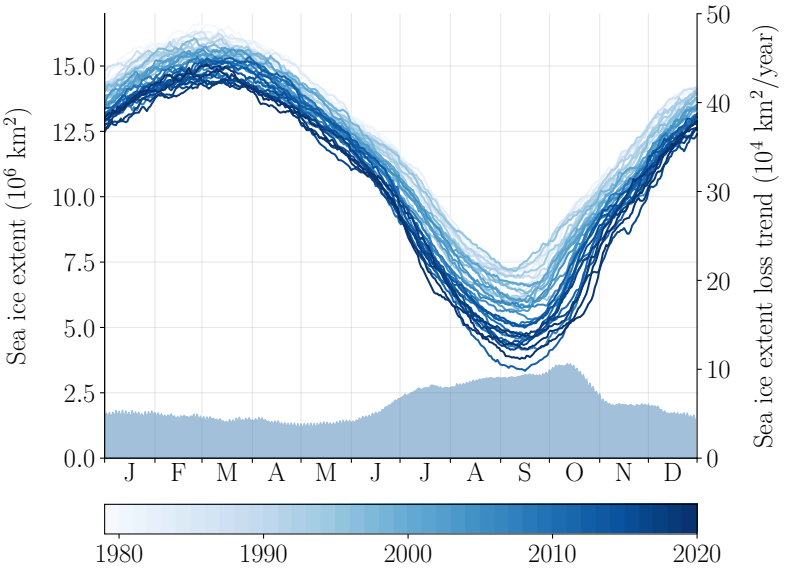


Figure 1.4: Daily pan-Arctic sea ice extent between 1979 and 2020. The shaded region represents the trend in sea ice extent for each day of the year. Data from NSIDC (Fetterer et al., 2017).

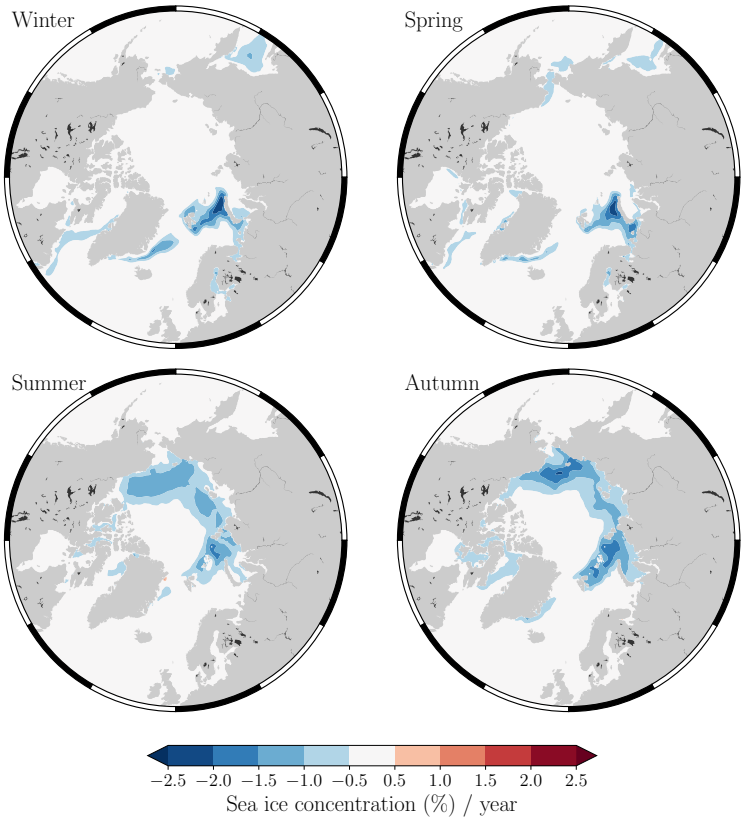


Figure 1.5: Linear trends in sea ice concentration between 1979 and 2020, for each season. Winter = December–March; spring = April–May; summer = June–September; autumn = October–November. Data from NSIDC (Cavalieri et al., 1996).

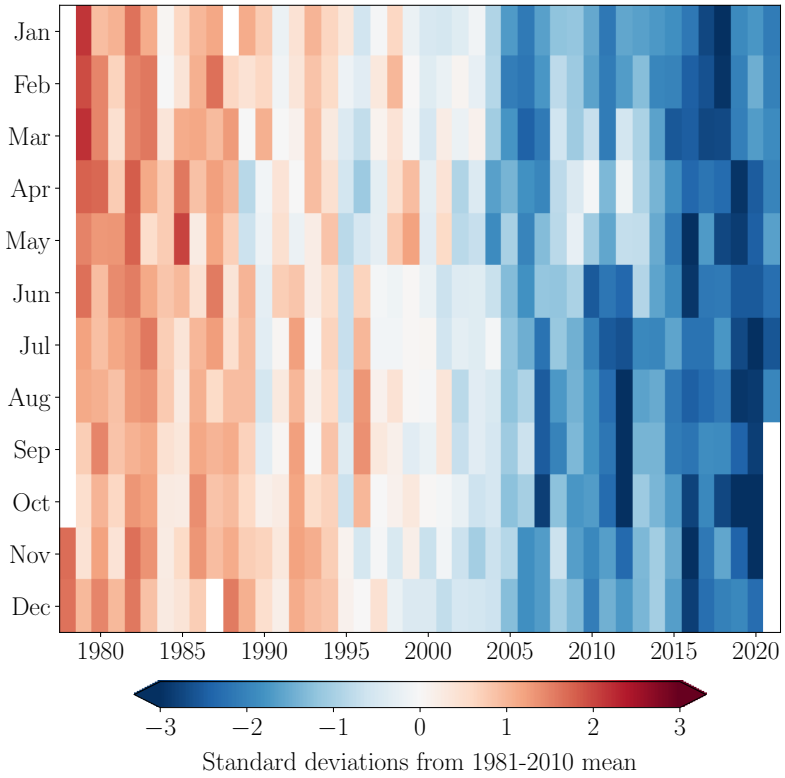


Figure 1.6: Z-score of pan-Arctic sea ice extent anomalies between 1978 and 2021. Plot updated from Stroeve & Notz (2018). Data from NSIDC (Fetterer et al., 2017).

the total trend in pan-Arctic March sea ice extent. Furthermore, relative to the period 1979–1989, regions such as the East Siberian, Kara, Laptev, Chukchi seas, and Hudson Bay have lost over 80% of their September sea ice cover, while areas such as the Barents Sea and Sea of Okhostk have lost almost 50% and 30% of their March ice cover respectively (Stroeve & Notz, 2018).

Sea ice extent also exhibits significant inter-annual variability. The year 2007 for example was record-breaking in terms of its anomalously low summer (September) sea ice cover, falling 2.5 standard deviations below the 1981–2010 long-term mean, which was then broken in September 2012 by an extent which fell 3 standard deviations below the mean (Stroeve & Notz, 2018). Despite this, the sea ice cover made remarkable recoveries in the following years, falling 2 and 1.5 standard deviations below the 1981–2010 mean in September 2008 and 2013 respectively (Figure 1.6). Figure 1.6 is an updated version of the analysis presented by Stroeve & Notz (2018), which shows the number of standard deviations (i.e., Z-score) that each monthly mean pan-Arctic sea ice extent value fell above or below the 1981–2010 reference mean, for all years between 1978 and 2021. The largest deviation from the long-term average to date was in October 2020, for which sea ice extent fell over 3.7 standard deviations below its 1981–2010 mean. Regionally, Onarheim et al. (2018) showed that the Beaufort, Chukchi, East Siberian, Laptev and Kara seas together explain

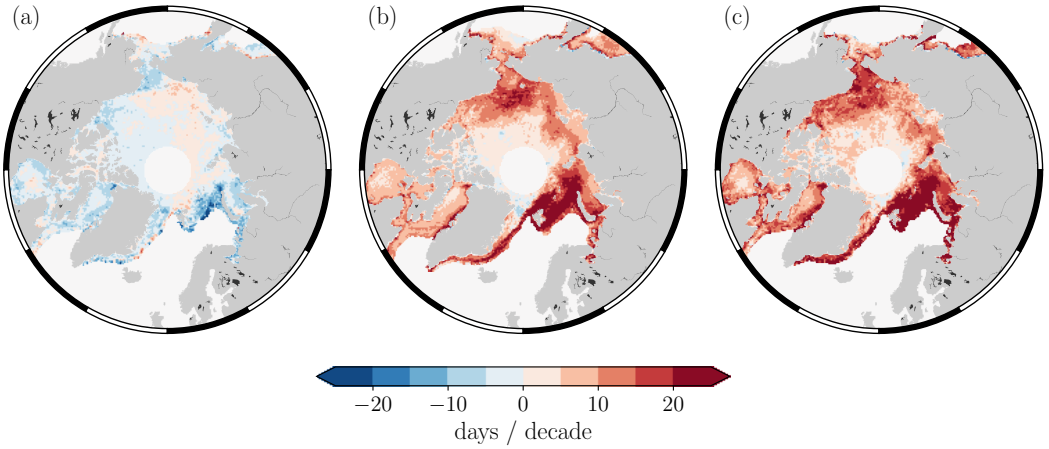


Figure 1.7: Linear trends in the date of (a) melt onset and (b) freezeup, and (c) the length of the melt seasonal (freezeup − melt onset) between 1979 and 2017. Data from NSIDC (Steele et al., 2019).

88% of the inter-annual variability in pan-Arctic September sea ice extent between 1950 and 2013, and the Sea of Okhotsk, Baffin Bay, Barents and Greenland seas explain 86% of the inter-annual variability in pan-Arctic March sea ice extent over the same period – chapters 3 and 5 take a closer look at the spatio-temporal patterns of inter-annual sea ice variability.

Seasonal changes in the ice cover also include a lengthening of the Arctic melt season (Markus et al. 2009; Stroeve et al. 2014a; Bliss & Anderson 2014). Figure 1.7 shows trends associated with the date of sea ice melt onset, freezeup and subsequently the length of the melt period between 1979 and 2017 (negative trends indicate the event at a given pixel, e.g., melt onset, is occurring earlier, and positive trends indicate the event is occurring later). On average we can see that melt onset is occurring earlier in the year, with an average trend of approximately −3 days per decade, and that the date of ice-freezeup is occurring later, with an average trend of approximately +7 days per decade. This has resulted in an overall lengthening of the melt season by approximately +10 days per decade. Some of the largest trends have occurred in the Barents Sea, with average trends of roughly −8, +15, and +23 days per decade for the date of melt onset, freezeup, and the length of the melt season respectively. While observations of sea ice thickness are sparse in space and time, Kwok (2018) collated a wide array of observational data sets spanning 1958–2018, including in-situ buoy data, upward-looking sonar, and satellite altimetry and scatterometer data to deduce regional trends in sea ice thickness, volume, and age. They found that Arctic sea ice has undergone basin-wide thinning over this period, having lost over 50% of its multi-year ice cover since 1999, whilst also reporting downward trends in winter ice volume between 2003–2018 of almost -3000 km^3 per decade (Kwok, 2018). Figure 1.8 shows trends in winter sea ice age between 1984–2020, where the largest trends occur in the Canada basin, reaching magnitudes as high as −2.7 years per

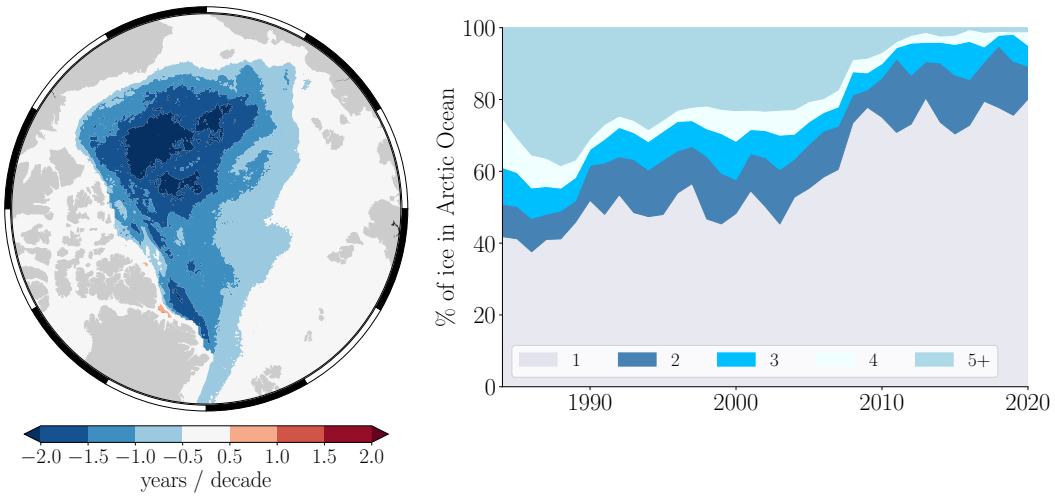


Figure 1.8: Linear trends in winter (December–March) sea ice age within the Arctic Ocean between 1984 and 2020 (left), and the cumulative percentage of sea ice within the Arctic Ocean belonging to each age category, in years (right). Plot inspired by Stroeve & Notz (2018). Data from NSIDC (Tschudi et al., 2019).

decade. We also notice the relative proportion of winter sea ice older than 5 years has reduced from over 30% in the 1980s, to now less than 2%, and that the Arctic has shifted to a predominantly seasonal ice cover, with approximately 80% of ice now being less than 1 year old.

1.1.2 Drivers of sea ice trends

Deducing the drivers behind observed changes in any component of the climate system requires careful consideration of cause and effect. While causal inference methods can in some cases elucidate physical drivers (e.g., Runge et al. 2019a), such approaches generally do not take into account physical laws, and so one must make certain that any analysis components are not confounded (i.e., there is not some unknown 'third-party' variable driving the relationship we see between two or more other variables), and also that the direction of causality is clearly understood (Pearl & Mackenzie, 2018). Climate models offer a natural pathway to studying causality as we can control for the influence of a given process when performing a particular climate simulation, and then assess any changes in subsequent experiments by allowing for its influence. One common example which has been highlighted in previous IPCC assessment reports is whether the increase in anthropogenic (human-produced) CO_2 emissions since the pre-industrial era has directly caused global surface temperatures to increase. This has been tested in climate models by controlling for this anthropogenic forcing and assessing the evolution of surface temperature anomalies between 1850 and the present day (e.g., Gillett et al. 2021). Figure 1.9 illustrates this by

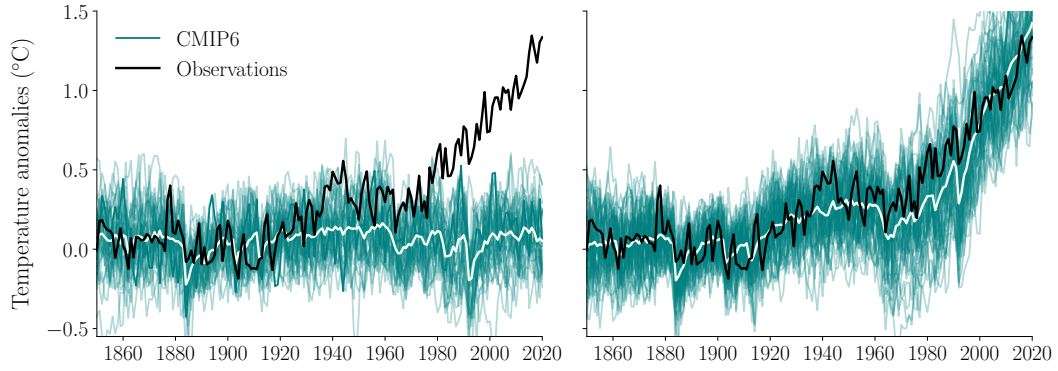


Figure 1.9: Evolution of global mean surface temperature anomalies relative to the period 1880–1919, under natural forcing conditions (left) and natural + anthropogenic CO_2 forcing (right). The shaded teal region represents the range of uncertainty projected by individual model runs from CMIP6. The pale green line represents the multi-model mean. Observations from the Met Office Hadley Centre (HadCRUT5; Morice et al. 2021).

taking a number of model runs from 20 of the latest generation of climate models submitted to phase 6 of the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016), where we notice that the natural forcing scenario alone (Figure 1.9 left) cannot explain the increasing trend in global mean surface temperature anomalies seen in the observational data – note that natural forcing implies that atmospheric CO_2 concentrations only change through natural processes such as changes in incoming solar radiation and volcanic activity. On the other hand, when models also take into account the anthropogenic forcing (Figure 1.9 right), they are, on average, able to reproduce the observed trends. Subsequent analysis of this result in AR6 recently led to the conclusion that "it is unequivocal that human influence has warmed the global climate system since pre-industrial times." (IPCC, 2021).

In the context of Arctic air temperatures, warming has occurred at over twice the rate of the global average (Serreze & Francis 2006; Screen & Simmonds 2010), and seasonally, lower troposphere temperatures have increased by at least four times as much in winter as they have in summer (Bintanja & Van der Linden, 2013). This *Arctic amplification* is thought to be due to changes in the surface albedo characteristics over the Arctic, caused by the continued sea ice decline, as well as other factors including (but not limited to) increased cloud cover (Kay & L'Ecuyer 2013; Jun et al. 2016), atmospheric heat transport to the Arctic (Zhang et al., 2008), and downwelling longwave radiation from increased CO_2 emissions (Notz & Stroeve, 2016) – see also Stroeve & Notz (2018) for further details. This warming pattern over the Arctic is directly correlated with increased sea ice loss in both observations and models (e.g., Gregory et al. 2002; Stroeve & Notz 2015), which indirectly points to the increase in anthropogenic CO_2 emissions being the primary driver of sea ice loss over the last century. From these findings, Notz & Stroeve (2016) regressed

cumulative CO_2 emissions against observed September sea ice area and found that each metric ton of CO_2 released to the atmosphere corresponds to approximately 3 ± 0.3 m^2 of sea ice loss in September. A further study by Niederdrenk & Notz (2018) framed this in the perspective of global mean surface temperatures, and found that each 1°C of temperature increase corresponds to between 3 and 4 million km^2 of sea ice loss in summer.

Another factor which can cause changes in Earth's climate is internal variability, i.e., natural changes as a result of the chaotic nature of the interactions between different components of the climate system that occur in the absence of any radiative forcing (Hawkins & Sutton, 2009). This variability can therefore act to augment or suppress any underlying trends caused by anthropogenic forcing, particularly on shorter time-scales (Swart et al. 2015; Notz 2015). For example, in terms of global surface temperature anomalies, IPCC (2021) concluded that the contribution of internal variability to the temperature trends between 1850–2019 was likely between −0.2°C and 0.2°C, and as such the warming trend cannot be explained by internal variability alone. On the other hand, Ding et al. (2017) estimated that between 30–50% of the trend in September sea ice area can be explained by internal variability, and Notz (2017) stated that the trends in sea ice decline between 2003–2012 were amplified by rapid ice loss events such as September 2007, which are considered to be extreme events compatible with internal variability.

Sea ice loss into the 21st century Much recent attention has gone into estimating the approximate point at which the Arctic will become seasonally ice-free[1] (e.g., Stroeve et al. 2007; Jahn 2018; Notz & Community 2020; Årthun et al. 2021) as atmospheric temperatures continue to increase and the ice cover continues to thin and retreat. In the context of CO_2 emissions, Notz & Stroeve (2016) estimated that the Arctic is likely to become ice-free between August and September, with an additional 800 ± 300 Gt of anthropogenic CO_2 emissions, and furthermore ice-free between July and October with an additional 1400 ± 300 Gt of CO_2. In terms of an ice-free date, climate models are the obvious tool of choice for constraining the uncertainty in future projections. In the previous section, internal variability was highlighted as one of the major sources of uncertainty in short-term climate predictions, however possibly the largest uncertainty in long-term (e.g., 50–100 year) future projections comes from the uncertainty in future anthropogenic CO_2 emissions; commonly referred to as *scenario uncertainty*[2]. In the IPCC assessment report AR5, this uncertainty was described in terms of the amount of radiative forcing (in Wm^{-2}) that the Earth will experience by the year 2100, with model experiments outlining three major

[1] Ice-free conditions generally refers to less than 1 million km^2 of ice area.

[2] A third factor which also contributes to model uncertainty and which dominates over decadal time scales is the response of a given model to climate fluctuations due to model physics; often referred to as *response uncertainty*. See Figure 2 of Hawkins & Sutton (2009).

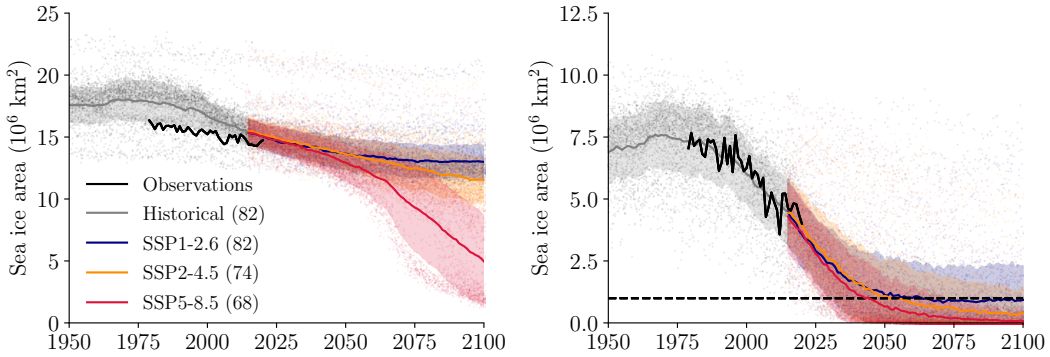


Figure 1.10: March (left) and September (right) sea ice area evolution between 1950 and 2100 from 30 CMIP6 models following different shared socioeconomic pathways (SSPs). The solid lines correspond to the mean across all model ensembles, and the shaded areas correspond to ±1 standard deviation around the mean. The number of ensembles used to generate each curve is given in the parentheses in the plot legend. The dashed horizontal line corresponds to 'ice-free' conditions (1×10^6 km^2). Plot adapted from Notz & Community (2020).

scenarios, or Representative Concentration Pathways (RCPs): RCP2.6, RCP4.5, and RCP8.5 (Van Vuuren et al., 2011). The development of these scenarios however, is an on-going process, with RCPs just the first stage. In AR6, model projections from CMIP6 now follow the next phase of emissions scenarios, which have been termed Shared Socioeconomic Pathways (SSPs; O'Neill et al. 2014). As well as environmental changes, SSPs aim to incorporate changes in societal factors including governance, technology and demographic changes, and furthermore describe how society adapts to and/or mitigates the effects of climate change. The current SSPs range from 1 to 5, where SSP1 is considered a sustainable future with increased emphasis on human well-being and consumption oriented towards low material growth, while in SSP5 social and economic growth are tied to continued fossil-fuel development, such that energy-intensive lifestyles expand across the globe (see Riahi et al. 2017 for further details). To allow for consistency between previous assessment reports, these SSPs have been integrated with RCPs, with some example scenarios given as: SSP1-2.6, SSP2-4.5, and SSP5-8.5 (Van Vuuren et al., 2014).

Figure 1.10 shows the evolution of pan-Arctic sea ice area between 1950 and 2100 for the months of March and September, from 30 different models submitted to CMIP6. In terms of September sea ice, we can see that on average the models report that the Arctic will be seasonally ice-free by the middle of the 21st century, albeit with significant uncertainty. In fact, Jahn (2018) stated that the uncertainty on ice-free dates can be as large as 6 decades in the lower forcing scenarios owing to enhanced internal variability, and also in the same study showed that limiting the degree of surface warming from 2°C to 1.5°C could reduce the probability of ice-free summers in 2100 from 100% to 30%. In the March panel in Figure 1.10, the ice cover survives until the end of the century for SSP1-2.6 and 2-4.5, although shows

significant decline between ∼2060–2100 for SSP5-8.5.

1.1.3 Drivers of sea ice variability

Predicting the state of the Arctic sea ice cover on time-scales ranging from a few days to a few years can be equally as much of a challenge as modelling its long-term evolution into the 21st century, and ultimately hinges upon our understanding of the various physical processes that occur on different time-scales, but function in a combined way to produce the eventual state of sea ice. These processes can be generally categorised into the following:

- A long-term external driver (e.g., anthropogenic CO_2 emissions) which impacts many aspects of the cryosphere (Hanna et al. 2012; Mernild et al. 2013; Notz & Stroeve 2016).
- Annual- to sub-seasonal ($>$ 1 year to $<$ 2 months) scale mechanistic drivers that contain some level of predictability, including: snow cover (Romanovsky et al. 2010; Crawford et al. 2018), spring-time humidity (Kapsch et al., 2013), melt-ponds (Schröder et al., 2014), atmospheric pressure patterns (Cohen et al. 2014, 2020), and sea-surface temperatures (Venegas & Mysak 2000; Vinje 2001; Balmaseda et al. 2010; Zhang 2015).
- Chaotic events with predictability similar to that of weather (on the order of 2 weeks), e.g., the *Great Arctic Cyclone* of 2012 (Overland et al., 2012).

Chapters 3 and 5 of this thesis focus on the inter-annual patterns of sea ice variability, and subsequently explore seasonal mechanisms for sea ice predictability. These mechanisms are often referred to as *teleconnections* in the climate science community, and can be considered as significant correlations between different components of the climate system over a wide range of spatial and temporal scales (Wallace & Gutzler, 1981). Specifically, the mechanism of sea ice persistence is utilised in chapter 3 to explore seasonal sea ice forecasting of the September sea ice minimum. Persistence refers to an inherent memory of sea ice extent/area anomalies whereby anomalies of the same sign recur, generally over a period of a few months – or even up to a year when considering thickness or volume persistence (Guemas et al., 2016). This is illustrated in Figure 1.11 by computing lagged correlations of pan-Arctic sea ice extent anomalies for each month of the year, where we notice the high degree of correlation in the first few leading months. What is also noticeable is that, after the point of maximum correlation loss (marked by the crosses), there is a re-emergence pattern (circles), at which point correlations begin to increase. For initial months in spring, there is a loss of correlation over the summer and a re-emergence of correlation in the proceeding autumn, and vice versa for autumn anomalies through winter.

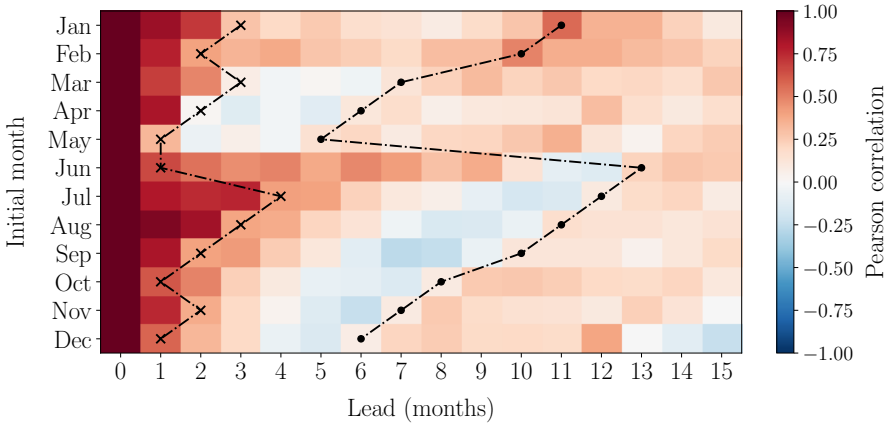


Figure 1.11: Lagged correlations (persistence) of pan-Arctic sea ice extent anomalies between 1979 and 2020. Lead refers to moving forwards in time, hence lead 1 for the initial month of January refers to the correlation between January and the proceeding February. Crosses denote the point of maximum de-correlation for each initial month, and the circles the point of re-emergence (see main text). Data from NSIDC (Fetterer et al., 2017).

This *sea ice re-emergence* feature occurs due to the sea ice cover 'imprinting' spring anomalies of the opposite sign in sea-surface temperatures around the location of the ice edge, which then persist through the summer and are returned to the sea ice as it grows in autumn – and similarly for autumn-to-spring/summer re-emergence (see Blanchard-Wrigglesworth et al. 2011 and Day et al. 2014 for further details).

In chapter 5, the role of large-scale winter atmospheric pressure patterns in driving the inter-annual variability of summer sea ice area is explored. Related to this pressure pattern is a feature known as the *Arctic Oscillation* (AO; Thompson & Wallace 1998), which corresponds to the leading mode of variability (see section 2.3.1) in mean sea-level pressure data north of 20° N. Previous studies (Rigor et al. 2002; Williams et al. 2016) have highlighted the pathway which connects the winter AO to summer sea ice as follows: a positive winter AO (anomalously low mean sea-level pressure) is coincident with (a) a weakening of the Beaufort Gyre, which reduces the amount of west-to-east ice advection, (b) a strengthening of the Transpolar Drift Stream, which increases ice export out of the Fram Strait, and (c) an increase in cyclonic ice motion in the Eurasian–Pacific sectors of the Arctic, which causes increased ice divergence and facilitates new ice formation. Once the melt season then begins, these expanses of relatively thin ice are more susceptible to melting, thus generally leading to anomalously low sea ice area by the end of summer.

Exploiting the predictive power associated with these various teleconnections requires tools which can isolate regions of variability and model their connectivity structure, whilst also accounting for any prior understanding of how we expect these patterns to emerge. The next chapter provides a background to some of the core principles of machine learning, and subsequently outlines the specific methods used to accomplish these objectives.

Chapter 2

Machine learning principles

2.1 Introduction

Machine learning is a field dedicated to the development of generalised models which are able to identify regularities, or *patterns*, within a given data set. The following chapter outlines two novel techniques for exploiting such patterns within climate time series data; techniques belonging to the sub-fields of supervised and unsupervised learning, and which have been widely applied in a number of fields beyond polar climate science. First, a brief discussion on the general aims of supervised learning is outlined, before moving on to a comprehensive overview of regression analysis. From first principles, linear Ordinary Least Squares (OLS) regression is introduced as the foundation from which all regression models inherently stem, before subsequently building towards the first novel technique of this thesis, Gaussian Process Regression (GPR). Although excellent works on regression analysis already exist in the literature (see e.g., Bishop 2006 and Rasmussen & Williams 2006), the methods are presented here as a self-contained piece of text in order to show how GPR is in fact just an advanced non-parametric adaptation of the simple linear OLS model, yet is a powerful tool for learning relatively complex functions. The chapter then turns to a discussion on unsupervised learning, where Principal Component Analysis (PCA) is outlined as the standard tool for investigating spatio-temporal patterns within climate time series data, and sets the scene for the second novel technique of this thesis, which is an alternative method for dimensionality reduction based on cluster analysis. The final section of this chapter shows how clustering leads naturally to a complex network representation of the climate system, and subsequently outlines the advantages of using networks as graphical representations of real physical systems, arguing that the approach provides advantages over PCA in terms of interpretability and flexibility for being incorporated into other models.

2.2 Supervised learning

Consider the case where we have two sets of (one-dimensional) observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$, such that together they comprise a data set $\mathcal{D}$ of n pairs of observations $\mathcal{D} = (\mathbf{x}, \mathbf{y}) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. These observations could correspond to measurements at n different points in time, or similarly at n different spatial locations at a given instant in time. Now suppose that we wish to learn about the process, or *function*, which relates each pair of observation points:

$$\mathbf{y} = f(\mathbf{x}). \tag{2.1}$$

This is the fundamental basis of supervised learning, where typically $\mathbf{x}$ is defined as the set of *inputs*, $\mathbf{y}$ as the set of *outputs* (or *targets*), such that collectively $\mathcal{D}$ becomes the *training set* of observations which enables us to learn about the function f. In climate science we are often interested in f as it (a) gives us some insight into the relationship between two or more physical processes and (b) allows us to make predictions about them. For example, what do I expect the value of y_{n+1} to be, given $\mathcal{D}$? This is essentially the goal of time series forecasting, and once we are confident in our model $f(\mathbf{x})$, answering this question is relatively straight forward. We simply supply the learnt function with a *test input*, e.g., $x_* = x_{n+1}$, in order to produce the equivalent function value at the $n^{\text{th}} + 1$ point in time:

$$f_* = f(x_*). \tag{2.2}$$

Supervised learning can further be sub-divided into the frameworks of *regression* and *classification*. Regression problems concern output variables which are continuous by nature and as such include almost all available climate time series data products. Conversely, classification problems handle discrete output variables, which are typically generated by the assignment of *class labels* that may correspond to the state of the output variable at a given instant in time. Figure 2.1 shows an example time series measuring sea ice concentration at some location in the Arctic Ocean. Sea ice concentration is a continuous variable and as such this time series could easily correspond to the outputs of a regression problem. On the other hand, we could discretise this time series by defining class labels which correspond to different ranges of sea ice concentration values, for example:

$$y_i = \begin{cases} 1 & \text{if } a_i \geq 15\% \\ 0 & \text{otherwise} \end{cases},$$

where a_i is the sea ice concentration value at time i. The vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ then becomes the set of binary values which, in this case, describe the presence 1 or

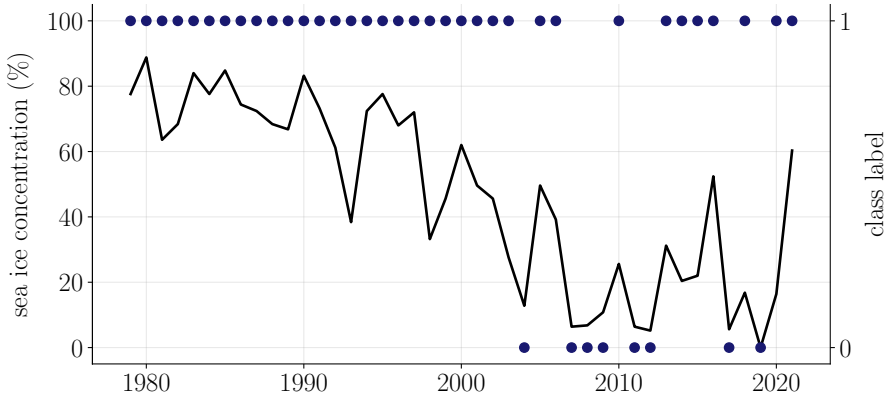


Figure 2.1: Sea ice concentration represented as both a continuous (black) and discrete (blue) variable. A class label is assigned for sea ice concentration values such that if the sea ice concentration value is $\geq 15\%$, then the class label = 1, and subsequently the class label = 0 otherwise.

absence 0 of sea ice at a particular location in the Arctic Ocean, and could therefore be used as the outputs in a (binary) classification problem.
The remainder of this chapter focuses on regression problems, where output variables are continuous.

2.2.1 Regression analysis

The aim of regression analysis is to learn about functional relationships between inputs and continuous outputs, and to subsequently make predictions of the outputs based on this learning. The notation hereafter follows that of the input-output pairs defined in section 2.2 although, for convenience, an additional column vector of ones is introduced to the inputs, such that $\mathbf{x}^{\mathrm{T}}$ becomes a $n \times 2$ matrix.

Linear ordinary least squares regression Let us begin by considering a simple parametric linear regression model, where it is assumed that input-output pairs (x_i, y_i) are linearly related by a set of regression parameters $\boldsymbol{\beta}$, but have been corrupted by some amount of unobserved random noise ε_i, such that Equation (2.1) becomes:

$$\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\varepsilon} = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.3}$$

or equivalently

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

In this case, the outputs are simply linear functions of the inputs such that the model contains two parameters β_0 and β_1, which govern the offset and gradient of the function values respectively. This type of model has desirable properties in that the optimal estimated solution for the parameters $\boldsymbol{\beta}$ is analytically tractable via the method of Ordinary Least Squares (OLS) regression, which can be understood as the model which chooses the values of $\boldsymbol{\beta}$ that minimise the error function:

$$\begin{aligned} E_{OLS}(\hat{\boldsymbol{\beta}}) &= (\mathbf{y} - \mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}\mathbf{y} + \hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}\mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}. \end{aligned} \tag{2.4}$$

Here the notation $\hat{\boldsymbol{\beta}}$ is to differentiate the *estimated* regression parameters through OLS regression, from the *true* parameters $\boldsymbol{\beta}$. From Equation (2.4) we notice that it is possible to find a unique solution for the values of $\hat{\boldsymbol{\beta}}$ for which $E_{OLS}(\hat{\boldsymbol{\beta}})$ is minimum, by first deriving the partial derivatives with respect to $\hat{\boldsymbol{\beta}}$:

$$\frac{\partial E_{OLS}(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{x}\mathbf{y} + 2\mathbf{x}\mathbf{x}^{\mathrm{T}}\hat{\boldsymbol{\beta}}, \tag{2.5}$$

and then setting this solution to zero, so we arrive at $(\mathbf{x}\mathbf{x}^{\mathrm{T}})\hat{\boldsymbol{\beta}} = \mathbf{x}\mathbf{y}$ – the *normal equations* for linear OLS regression. Finally, if the inverse $(\mathbf{x}\mathbf{x}^{\mathrm{T}})^{-1}$ exists, then simply computing the scalar product between the normal equations and this inverse gives the solution for the optimal estimated regression parameters:

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}\mathbf{x}^{\mathrm{T}})^{-1}\mathbf{x}\mathbf{y}. \tag{2.6}$$

The solution above constitutes the *learning* phase of regression analysis, however beyond the actual values of $\hat{\boldsymbol{\beta}}$, we are often interested in *predicting* the function values at new test points $\mathbf{x}_*$ (e.g., where we do not have observations of the output variable). As outlined in section 2.2, this is achieved by providing the model with a set of test inputs:

$$\boldsymbol{f}_* = f(\mathbf{x}_*) = \mathbf{x}_*^{\mathrm{T}}\hat{\boldsymbol{\beta}}. \tag{2.7}$$

The Gauss-Markov theorem states that the OLS solution for linear regression in Equation (2.6) is the *best linear unbiased estimator*. This means that if we consider the training outputs to be a set of random samples $\mathbf{y}_j = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j$ drawn from some overall population $\mathbf{y} = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$, then OLS regression essentially finds the parameters $\hat{\boldsymbol{\beta}}_j$ which minimise Equation (2.4) for the given sample set j, not for the population itself. OLS regression is *unbiased* because if this process was repeated for an infinite number of independently drawn (random) samples, such that $j = \{1, 2, \ldots, \infty\}$, the expected value (average) of the OLS regression parameters $\mathbb{E}[\hat{\boldsymbol{\beta}}]$ would be identical to that of the true population $\mathbb{E}[\hat{\boldsymbol{\beta}}] \equiv \boldsymbol{\beta}$. It is also the *best* because the variance of $\hat{\boldsymbol{\beta}}$, across all j, would be the smallest compared to any other linear unbiased estimator.

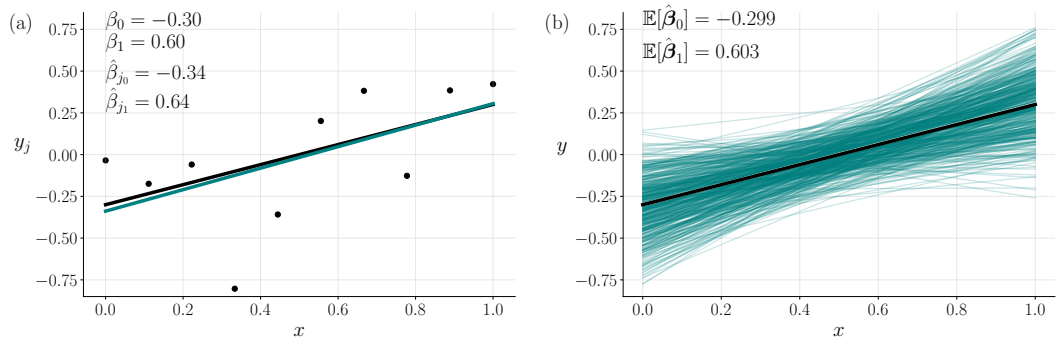


Figure 2.2: Synthetic example of 1-D linear OLS regression. (a) The black line shows the function from which the data are generated (before being corrupted by Gaussian noise). The teal line is the estimated function after linear OLS regression $\boldsymbol{f}_*$. (b) Repeating the process in (a) 10,000 times for samples drawn from the same function, but with independent noise contents. For visualisation purposes the input-output pairs are not shown, and also only the first 500 (out of 10,000) derived regression lines are shown.

In other words, any other model we might use to estimate $\hat{\boldsymbol{\beta}}$ that found a smaller variance $\sigma^2_{\hat{\boldsymbol{\beta}}}$ than the OLS solution would be biased ($\mathbb{E}[\hat{\boldsymbol{\beta}}] \neq \boldsymbol{\beta}$), and one which was unbiased ($\mathbb{E}[\hat{\boldsymbol{\beta}}] \equiv \boldsymbol{\beta}$) would have a larger variance than the OLS solution. This is commonly referred to as the *bias-variance trade-off*.

Figure 2.2 provides a synthetic example of linear OLS regression to illustrate the concepts outlined above. In Figure 2.2a, a set of 10 evenly spaced values are generated between 0 and 1 to represent the inputs $\mathbf{x}$, and subsequently the set of random outputs $\mathbf{y}_j$ are generated by passing $\mathbf{x}$ through the linear function $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$, where $\beta_0 = -0.30$ and $\beta_1 = 0.60$, before finally adding zero-mean Gaussian noise to each of the outputs in order to simulate noisy real-world data. The aim is to then recover the values of $\boldsymbol{\beta}$ using the training set $\mathcal{D}_j = (\mathbf{x}, \mathbf{y}_j)$ in a linear OLS regression framework. Using Equation (2.6), the magnitude of the linear regression parameters is slightly overestimated, with $\hat{\beta}_{j_0} = -0.34$ and $\hat{\beta}_{j_1} = 0.64$, however by repeating this process 10,000 times (Figure 2.2b), we find a much closer estimate of the true values by averaging across all samples, with $\mathbb{E}[\hat{\boldsymbol{\beta}}_0] = -0.299$ and $\mathbb{E}[\hat{\boldsymbol{\beta}}_1] = 0.603$. We can also see in this case how multiple data sets allow us to build up an estimate of uncertainty in an OLS regression model, which can be quantified by computing the standard deviation of the regression parameters across all 10,000 samples: $\sigma_{\hat{\boldsymbol{\beta}}_0} = 0.179$ and $\sigma_{\hat{\boldsymbol{\beta}}_1} = 0.301$. This is of limited use for observational products however, as in reality we often only have access to one data set and so cannot perform regression analysis on thousands of random samples. In any case, later sections will show how, by adopting a Bayesian approach, uncertainties can be generated with the predictions even for one data set. For now, let us advance the current framework, so that the OLS model is able to model non-linear relationships between inputs and outputs.

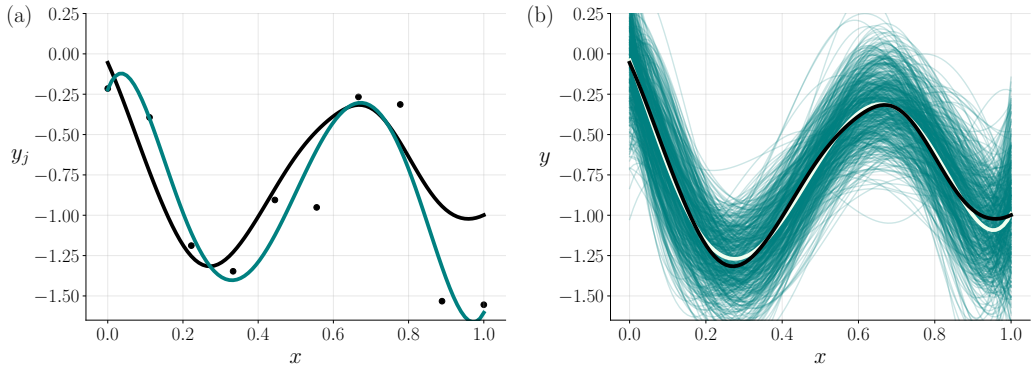


Figure 2.3: Synthetic example of 1-D non-linear OLS regression using polynomial basis functions with $m = 5$. The colours are consistent with the linear example, although the pale green curve in (b) represents the mean across all 10,000 estimated functions (cannot be seen in the linear case as the OLS solution very closely matches the true regression parameters).

Non-linear ordinary least squares regression One of the assumptions of the Gauss-Markov theorem is that the outputs are linear functions of the inputs; an example of which was shown in Figure 2.2. This poses a potential limitation, for if the training data exhibit non-linear relationships then this type of model will fail to accurately learn the desired function f. A way to free this type of model from such limitations is to map the inputs to a higher dimensional feature space using a set of fixed *basis functions* $\phi(\mathbf{x})$, such that Equation (2.3) becomes:

$$\mathbf{y} = f(\phi(\mathbf{x})) + \boldsymbol{\varepsilon} = \phi(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.8}$$

These basis functions allow non-linear relationships to be modelled between inputs and outputs whilst remaining linear in the parameters, and hence adhering to the linearity assumption of the Gauss-Markov theorem. One example of this mapping is the polynomial basis function: $\phi(x_i) = (1, x_i, x_i^2, x_i^3, \ldots, x_i^m)^{\mathrm{T}}$, such that $\phi(\mathbf{x})$ is a Vandermonde matrix which projects the inputs to the space of powers, up to the m^{th} order. Note however that in this instance, the number of parameters increases from two in Equation (2.3) to $m + 1$:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(\phi(x_1)) \\ f(\phi(x_2)) \\ \vdots \\ f(\phi(x_n)) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & x_2^3 & \ldots & x_2^m \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \ldots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Figure 2.3 shows an example of OLS regression where a non-linear relationship exists between inputs and outputs. In this case, the same data generating process is repeated as for the linear case, except the function f now draws samples from

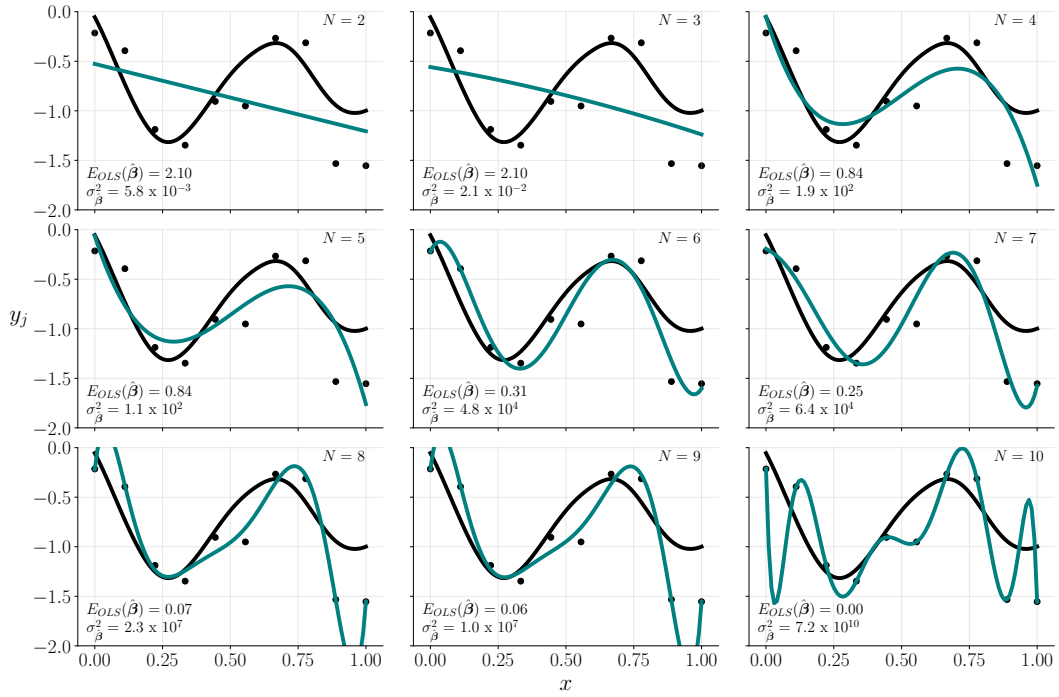


Figure 2.4: Synthetic examples of 1-D OLS polynomial regression for a variety of model complexities (increasing order of polynomial basis functions). As the number of model parameters increases the model begins to fit to the noise content of the data, until $N = 10$ where the model passes through every training point.

some multivariate Gaussian distribution (Gaussian distributions are explored in detail later). This time, learning the function values is achieved by mapping the inputs using a set of polynomial basis functions where $m = 5$ (hence the model has $N = 6$ parameters), and implementing the parametric OLS regression model as before with Equation (2.6). We can now see that although the model is linear in the parameters, the estimated function values are non-linear functions of the inputs. One important point to note however, is what happens when the number of parameters is varied in an OLS regression model. With $N = 6$ parameters in Figure 2.3a, the predictions are somewhat poor in areas with larger noise content. This is commonly referred to as *over-fitting*, when a model fits to the random noise in the data, thus leading to biased predictions. Over-fitting is a specific limitation of the OLS method, which in this case exacerbated by the complexity of the model – large N[1], however note how in Figure 2.3b the bias is reduced by averaging across 10,000 samples. Figure 2.4 illustrates the over-fitting problem with regards to model complexity by attempting to learn the same function as in Figure 2.3, except with a variety of polynomial basis functions with increasing number of parameters from $N = 2$ to $N = 10$. In the case where $N = 2$ (i.e., linear), the model is not flexible

[1]Over-fitting can also be a consequence of small n. With sufficient training points, OLS regression will reduce the severity of over-fitting for a given model complexity. See Figure A.1.

enough to explain the non-linear relationship between inputs and outputs, however as the number of parameters increases, we notice a better fit for the model. If a preference were to be given for a particular model based on $E_{OLS}(\hat{\boldsymbol{\beta}})$, then it would ultimately be the case where $N = 10$, as the function values pass through every training point such that $E_{OLS}(\hat{\boldsymbol{\beta}}) = 0$. This 'perfect' fit however, has come at a price. With $N = 10$ there is significant over-fitting and subsequently the predictions do not give a good representation of the true function f. Clearly then there is a trade-off between minimising $E_{OLS}(\hat{\boldsymbol{\beta}})$ and mitigating over-fitting. The next section turns to regularised least squares regression as an alternative parametric regression model which attempts to minimise the effect of over-fitting by exploiting the bias-variance trade-off in order to reduce the complexity of the model.

Regularised least squares regression In the previous section, we examined how over-fitting in OLS regression can produce statistical models that do not generalise well. This section continues with the feature space projection of the inputs, although for convenience hereafter uses the notation $\boldsymbol{\phi} = \phi(\mathbf{x})$. Regularised Least Squares (RLS) regression amends the OLS error definition in Equation (2.4) by which the sum of the squared residuals are minimised, subject to some condition:

$$E_{RLS}(\hat{\boldsymbol{\beta}}) = (\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}}) \quad \text{subject to} \quad \|\hat{\boldsymbol{\beta}}\|_p^p \leq \alpha. \tag{2.9}$$

Here $\|\cdot\|_p$ represents the $\mathcal{L}_p$ vector norm, and α is a constant. Geometrically, $\|\hat{\boldsymbol{\beta}}\|_p^p \leq \alpha$ can be visualised as a penalty constraint in the space of parameters, where the distance in parameter space to the OLS derived solution $\hat{\boldsymbol{\beta}}$ must not be greater than some distance α. Revisiting the over-fitting problem from Figure 2.4 and comparing the variance of the regression parameters $\sigma^2_{\hat{\boldsymbol{\beta}}}$ for each model, we notice that as the model complexity increases, so does $\sigma^2_{\hat{\boldsymbol{\beta}}}$, as the estimated regression parameters become very large. This is a specific case of Runge's phenomenon in polynomial regression (Runge, 1901), however large $\sigma^2_{\hat{\boldsymbol{\beta}}}$ can also occur in OLS regression when using multiple predictor variables (inputs) which are collinear. For example, if $\mathbf{x}^{\mathrm{T}}$ were a $n \times D$ design matrix, where $D < n$, collinearity would exist if a given column of $\mathbf{x}^{\mathrm{T}}$ could be linearly predicted by another column of $\mathbf{x}^{\mathrm{T}}$. RLS regression therefore aims to mitigate this problem by *shrinking* the values of $\hat{\boldsymbol{\beta}}$ until $\|\hat{\boldsymbol{\beta}}\|_p^p \leq \alpha$. Common choices of p are $p = 1$ or $p = 2$, such that distances are given by

$$\|\hat{\boldsymbol{\beta}}\|_1^1 \equiv |\hat{\boldsymbol{\beta}}| = |\hat{\beta}_0| + |\hat{\beta}_1| + , \ldots, + |\hat{\beta}_m|$$
$$\|\hat{\boldsymbol{\beta}}\|_2^2 \equiv \hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\beta}} = \hat{\beta}_0^2 + \hat{\beta}_1^2 + , \ldots, + \hat{\beta}_m^2.$$

In the case where $p = 1$ we arrive at Lasso regression (Tibshirani, 1997), and similarly Ridge regression for $p = 2$ (Hoerl & Kennard, 1970). Figure 2.5 shows the geometric

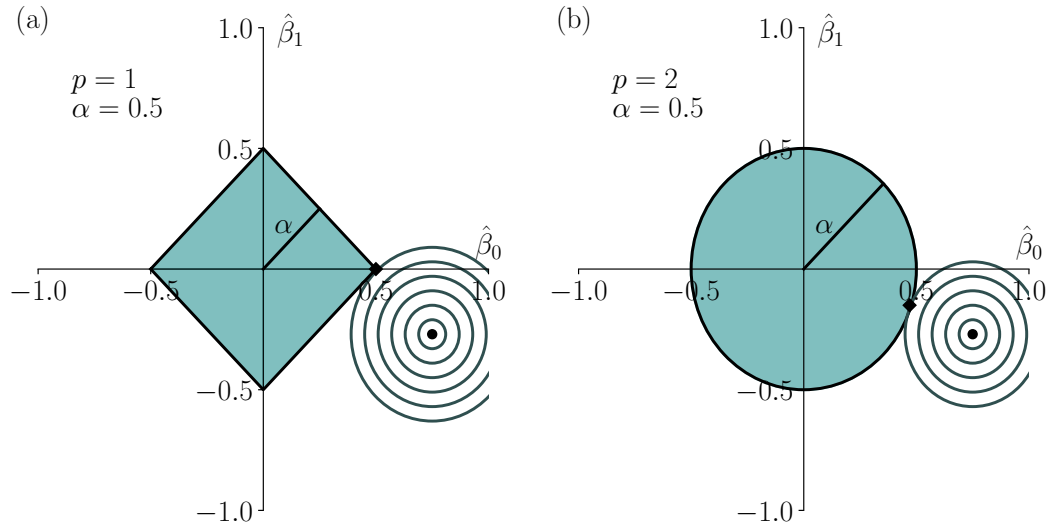


Figure 2.5: Parameter shrinkage for the cases of (a) Lasso and (b) Ridge regression. The circular scatter point is the OLS derived parameter solution $\hat{\boldsymbol{\beta}}$, with the $E_{OLS}(\hat{\boldsymbol{\beta}})$ contours shown by the teal circles. The shaded regions represent a distance α from the origin, according to the $\mathcal{L}_p$ norm. RLS regression shrinks the values of $\hat{\boldsymbol{\beta}}$ along the OLS contours until reaching the shaded region. The new parameters are given by the diamond scatter point.

representation of the penalty constraint for both Lasso and Ridge models, where the model contains 2 regression parameters (note that the same principles apply to any number of dimensions). We can see how Lasso regression is able to shrink parameters all the way to zero, whereas Ridge regression will typically shrink values close to zero (due to the geometric shape of the $\mathcal{L}_p$ vector norm). This *parameter shrinkage* has the effect of reducing the effective model complexity, and hence over-fitting.

By using Lagrange multipliers it is possible to derive an analytical solution for Ridge regression, although note that Lasso regression does not have an equivalent closed form solution. The expression of Equation (2.9) in the Lagrangian form is given as:

$$E_{RLS}(\hat{\boldsymbol{\beta}}, \lambda) = E_{OLS}(\hat{\boldsymbol{\beta}}) + \lambda F(\hat{\boldsymbol{\beta}}), \quad \text{where} \quad F(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \alpha$$

and λ is the Lagrange multiplier (also known as the *regularisation coefficient*). Expanding the notation above, we then arrive at:

$$\begin{aligned} E_{RLS}(\hat{\boldsymbol{\beta}}, \lambda) &= (\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}}) + \lambda(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \alpha) \\ &= \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\phi}\mathbf{y} + \hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}} + \lambda\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \lambda\alpha. \end{aligned} \tag{2.10}$$

The regularisation coefficient λ can be seen as the parameter which minimises over-fitting by controlling the effectiveness of the quadratic penalty term $\hat{\boldsymbol{\beta}}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$, where generally an appropriate value can be estimated through cross-validation (see

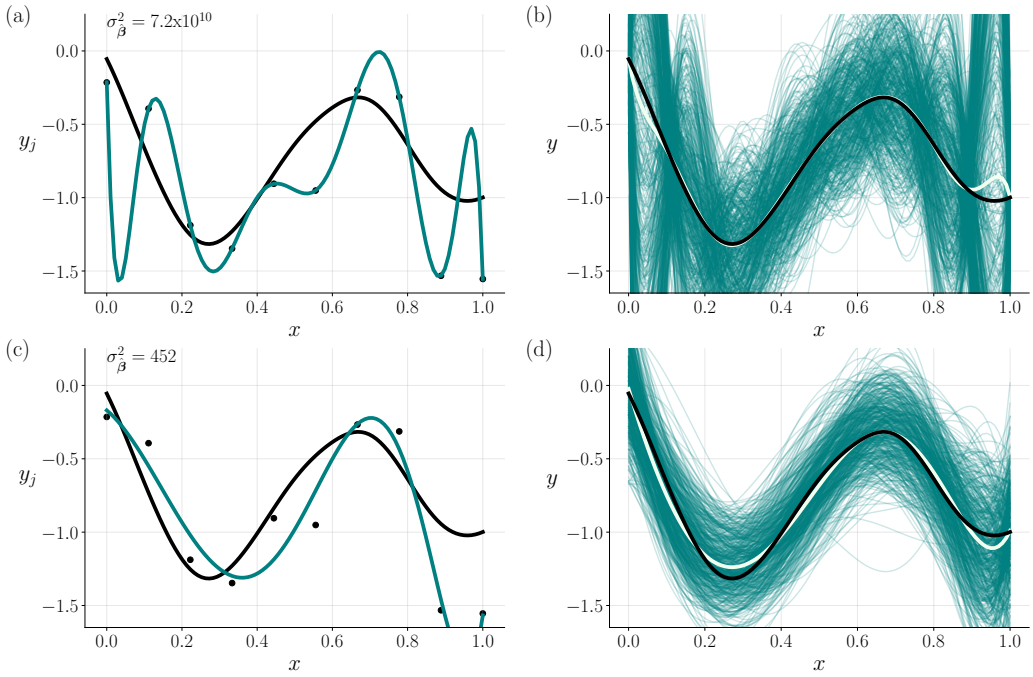


Figure 2.6: Synthetic example of 1-D (a,b) OLS and (c,d) Ridge polynomial regression ($N = 10$). The regularisation coefficient is set to $\ln \lambda = -12.2$ for the Ridge regression case. Colours are consistent with Figures 2.2 and 2.3.

appendix A.1). Section 2.2.2 however will outline a more principled and efficient Bayesian approach for estimating such 'free' parameters. In any case, the partial derivatives of Equation (2.10) with respect to $\hat{\boldsymbol{\beta}}$ are given as:

$$\frac{\partial E_{RLS}(\hat{\boldsymbol{\beta}}, \lambda)}{\partial \hat{\boldsymbol{\beta}}} = -2\boldsymbol{\phi}\mathbf{y} + 2\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\hat{\boldsymbol{\beta}} + 2\lambda\hat{\boldsymbol{\beta}}, \tag{2.11}$$

whereby setting to zero and re-arranging for $\hat{\boldsymbol{\beta}}$ leads to the optimal estimated Ridge regression parameters:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}} + \lambda\boldsymbol{I})^{-1}\boldsymbol{\phi}\mathbf{y}, \tag{2.12}$$

where $\boldsymbol{I}$ is the identity matrix. Figure 2.6 compares OLS and RLS (Ridge) regression, using a set of polynomial basis functions where $N = 10$. Immediately we notice how the over-fitting problem is minimised in the RLS case, as the variance of the regression parameters has been reduced considerably. We also notice how the average across 10,000 samples has a slightly higher bias in the RLS case, although has significantly smaller variance. In this case the bias could be improved perhaps by a more suitable choice of basis functions, as polynomial functions suffer inherent limitations. In the coming sections, kernel functions will be outlined as a natural way to alleviate the issue of choice of basis function, although for now, the next section explores how the current parametric regression model can be advanced into a probabilistic engine with Bayesian linear regression.

Bayesian linear regression The Bayesian framework explores the regression problem from a probabilistic viewpoint. In this case, rather than deriving point estimates of the regression parameters as in Equations (2.6) and (2.12), the aim is to instead derive a distribution of values, where the collection of values form a Gaussian probability distribution. This therefore allows uncertainty estimates to be computed on both the parameters and the predictions, from a single training set alone. Consider once again the standard supervised learning problem, but with some additional information:

$$\mathbf{y} = f(\boldsymbol{\phi}) + \boldsymbol{\varepsilon} = \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2). \tag{2.13}$$

Specifically, the random noise $\boldsymbol{\varepsilon}$ is considered to be independent and identically distributed Gaussian noise, with mean 0 and variance σ^2. Given this model, each observation point y_i is assumed to be a random variable drawn independently from a Gaussian distribution – known as the *likelihood distribution*, which can be computed as a conditional probability distribution over the outputs:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\beta}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_i - \boldsymbol{\phi}_i\boldsymbol{\beta})^2}{2\sigma^2} \right) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\boldsymbol{V}|^{1/2}} \exp\left(- \frac{1}{2}(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{V}^{-1}(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta}) \right). \end{aligned} \tag{2.14}$$

Here $\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta}$ and $\boldsymbol{V} = \sigma^2\boldsymbol{I}$ correspond to the mean and covariance of the multivariate likelihood distribution respectively. From the definition above we can see that any Gaussian distribution can be defined explicitly in terms of its mean and covariance (or variance in the case of a univariate distribution). It is also worth noting that the probabilistic equivalent of OLS regression involves the maximisation of the log of the likelihood function above – known as *maximum likelihood*, and while this *does* provide a framework in which to derive uncertainty estimates on both the parameters and the predictions, the approach still suffers from the problems of over-fitting (Bishop, 2006). The upcoming discussion however will show how a penalty constraint similar to RLS regression arises naturally in the Bayesian approach.

Next, a *prior* Gaussian probability distribution is assigned over the regression parameters, which incorporates any beliefs about how the parameters are expected to behave. Without a-priori information, the prior is typically chosen to be broad and centred on zero to reflect uncertainty in the magnitude and sign of the regression parameters respectively. For completeness however, a generalised prior with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is given here, such that $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, or equivalently:

$$p(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(- \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) \right). \tag{2.15}$$

Note that the prior is independent of the observations $\mathcal{D}$. By using Bayes' Law, the prior and likelihood distributions can be combined to give a *posterior* distribution, which describes the updated probability of the parameters, given the observations:

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi}) = \frac{p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\beta})p(\boldsymbol{\beta})}{p(\mathbf{y}|\boldsymbol{\phi})}. \tag{2.16}$$

The term in the denominator here is the *marginal likelihood*, which will be discussed in more detail in section 2.2.2. For now it can be considered as a normalisation term which integrates (marginalises) over the regression parameters, and ensures that probabilities in the posterior sum to one:

$$p(\mathbf{y}|\boldsymbol{\phi}) = \int p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\beta})p(\boldsymbol{\beta}) \;\; d\boldsymbol{\beta}. \tag{2.17}$$

The nature of conjugate priors (see Raiffa & Schlaifer 1961) means that the product of the Gaussian prior and Gaussian likelihood results in a posterior distribution which is also Gaussian. As such, the posterior must take the typical quadratic form:

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\tilde{\boldsymbol{\Sigma}}|^{1/2}} \exp\bigg(-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\mu}})^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta} - \tilde{\boldsymbol{\mu}}) \bigg). \tag{2.18}$$

The aim is to then produce a solution for the posterior mean $\tilde{\boldsymbol{\mu}}$ and covariance $\tilde{\boldsymbol{\Sigma}}$. This is achieved by first evaluating the numerator in Equation (2.16) to produce the un-normalised posterior:

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi}) \propto \exp\bigg(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{V}^{-1} (\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \bigg). \tag{2.19}$$

At which point, the matrix identity $\mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{C}\mathbf{A} \equiv \mathbf{A}^{\mathrm{T}}(\mathbf{B} + \mathbf{C})\mathbf{A}$ can be used to *complete the square* in order to give:

$$p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi}) \propto \exp\bigg(-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{\phi}\boldsymbol{V}^{-1}\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1})\boldsymbol{\beta} + \boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{\phi}\boldsymbol{V}^{-1}\mathbf{y} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}) + C \bigg), \tag{2.20}$$

where C is a constant which contains terms that are independent of $\boldsymbol{\beta}$. Finally, by recognising that the quadratic term inside the exponent of Equation (2.18), or in fact any Gaussian distribution, can be written as:

$$-\frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\mu}})^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta} - \tilde{\boldsymbol{\mu}}) = -\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\beta} + \boldsymbol{\beta}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{\mu}} + C, \tag{2.21}$$

we can notice the equivalence between this alternative form and Equation (2.20). The mean and covariance of the un-normalised posterior are then given simply as:

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \big(\boldsymbol{\phi}\boldsymbol{V}^{-1}\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1}\big)^{-1}\big(\boldsymbol{\phi}\boldsymbol{V}^{-1}\mathbf{y} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\big) \\ \tilde{\boldsymbol{\Sigma}} &= (\boldsymbol{\phi}\boldsymbol{V}^{-1}\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1})^{-1}. \end{aligned} \tag{2.22}$$

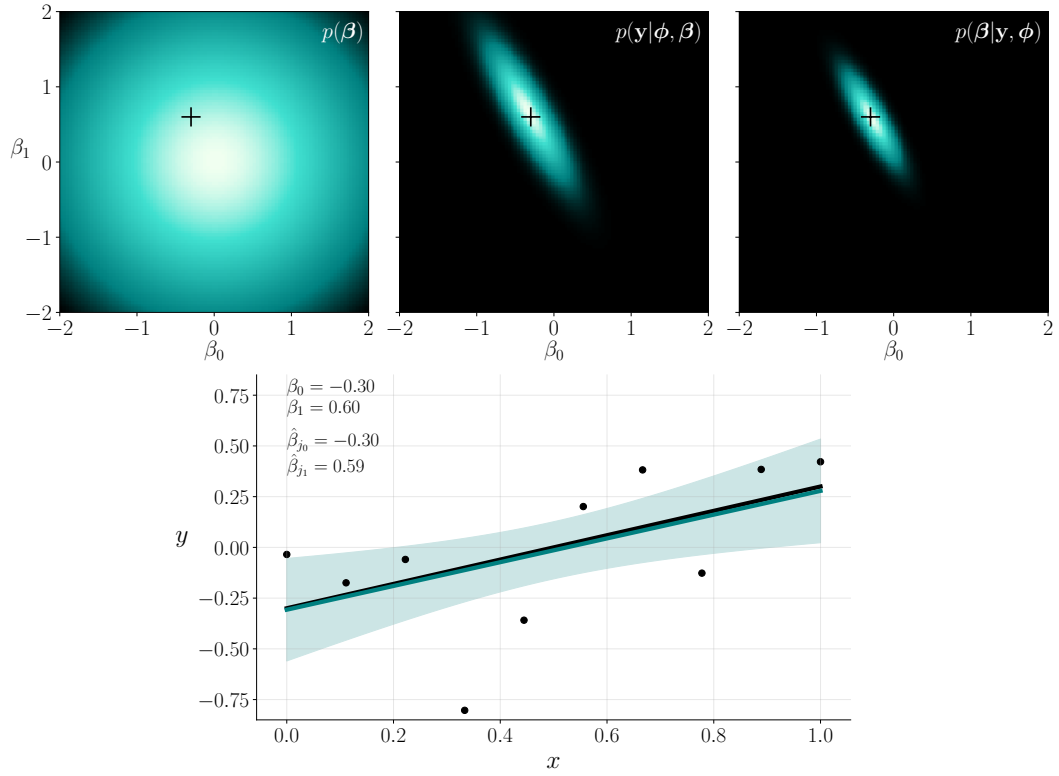


Figure 2.7: Synthetic example of 1-D Bayesian Linear Regression, using the same data as in Figure 2.2, hence $\boldsymbol{\phi} = \mathbf{x}$. The top density plots show the prior (left), likelihood (middle) and posterior (right) distributions. The black cross-hair is the location of the true parameters $\boldsymbol{\beta}$ from which the data were generated. The predicted function values and uncertainty range in the scatter plot (bottom) are computed from Equation (2.23).

The solution for $\tilde{\boldsymbol{\mu}}$ is often referred to as the *maximum a posteriori* (MAP) estimate of the parameters, as the mean of a Gaussian distribution is also its mode. The square-root of the diagonal elements in $\tilde{\boldsymbol{\Sigma}}$ then gives the standard deviation (uncertainty) on each of the regression parameters. Note that it is not necessary to normalise the posterior solution as per Equation (2.16) here, as this does not change the location of the mean, or alter the covariance. Finally, generating predictions now come in the form of a Gaussian distribution where, for a given set of test inputs $\boldsymbol{\phi}_*$, the posterior predictive distribution is given as $\boldsymbol{f}_* \sim \mathcal{N}(\bar{\boldsymbol{f}}_*, \boldsymbol{\sigma}^2_{\boldsymbol{f}_*})$, with mean $\bar{\boldsymbol{f}}_*$ and covariance $\boldsymbol{\sigma}^2_{\boldsymbol{f}_*}$:

$$\begin{aligned} \bar{\boldsymbol{f}}_* &= \boldsymbol{\phi}_*^{\mathrm{T}} \tilde{\boldsymbol{\mu}} \\ \boldsymbol{\sigma}^2_{\boldsymbol{f}_*} &= \boldsymbol{\phi}_*^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}} \boldsymbol{\phi}_*. \end{aligned} \tag{2.23}$$

Figure 2.7 revisits the linear regression problem from Figure 2.2, although now from the Bayesian perspective. The linear case is presented here for ease of visualisation, although the same principles apply to any number of dimensions. In this example, a broad zero-mean prior distribution is assigned over the parameters $p(\boldsymbol{\beta})$ before inspecting the data, at which point the likelihood function can be computed $p(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\beta})$,

which provides a distribution of $\boldsymbol{\beta}$ values which most likely explain the set of observations. Finally, combining the prior and likelihood gives rise to the posterior distribution over the parameters $p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi})$, where we can see how the uncertainty in the parameters is reduced relative to the likelihood, corresponding to an update in the probability of the MAP estimate.

At this point it is worth drawing some parallels between Bayesian linear regression and Ridge regression, in order to keep track of the principles that have been outlined so far. Specifically, consider the following zero-mean prior distribution over the parameters $\boldsymbol{\beta} \sim \mathcal{N}(0, \lambda^{-1}\boldsymbol{I})$, where λ is a constant. For this prior, one obtains a posterior mean equal to $\tilde{\boldsymbol{\mu}} = \left(\boldsymbol{\phi}\boldsymbol{V}^{-1}\boldsymbol{\phi}^{\mathrm{T}} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{\phi}\boldsymbol{V}^{-1}\mathbf{y}$. By then also considering the special case where $\boldsymbol{V} = \boldsymbol{I}$, we notice how the Bayesian solution reverts to Ridge regression: $\tilde{\boldsymbol{\mu}} = \hat{\boldsymbol{\beta}} = \left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{\phi}\mathbf{y}$. Recall how λ plays an important role in governing model complexity in Ridge regression by controlling the effectiveness of the quadratic penalty term $\lambda\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}$. In the Bayesian framework, the quadratic penalty term becomes $(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})/\sigma^2$; the effectiveness of which is controlled by any parameters which make up the prior covariance matrix (see section 2.2.2). From this it is possible to see how the seemingly complex Bayesian linear regression basis function model can be traced back to its origins in linear OLS regression (Equation 2.6). The coming sections will show how this also holds for Gaussian process regression, however before moving onto this, the next section introduces kernel functions; a key concept in the progression to a non-parametric framework for regression analysis.

Kernel functions So far this chapter has been concerned with linear parametric models, where the aim has been to estimate a vector of parameters $\hat{\boldsymbol{\beta}}$, or similarly a multivariate distribution of parameters $p(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\phi})$, and to generate predictions through linear combinations of the test inputs (e.g., Equation 2.7). Basis functions have been outlined as a tool for allowing such linear models to be flexible in terms of learning both linear and non-linear functions, however in some cases, restricting a given model to a particular set of basis functions (e.g., polynomials) may limit the class of functions the model is able to learn. An alternative approach is to generate predictions in terms of a *kernel function* (Aizerman 1964; Boser et al. 1992) whereby, rather than explicitly choosing a set of fixed basis functions which represent a projection to some higher dimensional feature space, the kernel function defines them implicitly. For example, consider the simple case where we have a set of inputs $\mathbf{x} = (x_1, x_2)$, and we wish to map them to a higher dimensional feature space using a second-order polynomial basis function, so that $\phi(\mathbf{x}) = (1, x_1, x_2, x_1^2, x_2^2, x_1x_2)^{\mathrm{T}}$, hence $\phi(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}^6$. A kernel function $k(\mathbf{x}, \mathbf{x}')$ represents an inner product in

some higher dimensional feature space, which in this example corresponds to:

$$k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^{\mathrm{T}}\phi(\mathbf{x}') = 1 + x_1x_1' + x_2x_2' + x_1^2x_1'^2 + x_2^2x_2'^2 + x_1x_1'x_2x_2'. \tag{2.24}$$

Notice here how this definition still requires explicit reference to the polynomial basis functions, however a way to circumvent this is by using what is known as the *kernel trick*, whereby $k(\mathbf{x}, \mathbf{x}')$ is computed without explicitly performing the mapping $\phi(\cdot)$. In other words, a kernel function can be defined which in principle represents an inner product in some higher dimensional feature space, but in reality is just a function of the inputs $\mathbf{x}$ and $\mathbf{x}'$:

$$k(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}')^2 = 1 + 2x_1x_1' + 2x_2x_2' + x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_1'x_2x_2'. \tag{2.25}$$

Working backwards, we notice how the kernel function above is equivalent to the inner product of the 6-dimensional feature space vector $(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)$, despite never having explicitly entered that space (note that this is just a scaled version of the solution in Equation 2.24). This is a particularly useful property of kernel functions for a few reasons. For one, it allows kernel functions to be created which represent inner products in potentially infinite-dimensional feature spaces, and secondly, observations which may contain unique individual components (e.g., a long-term increasing trend due to climate change, together with a periodic seasonal cycle), can be modelled more effectively by noticing that the sum and/or product of multiple kernel functions, is itself a kernel function. For example:

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')\big(k_2(\mathbf{x}, \mathbf{x}') + k_3(\mathbf{x}, \mathbf{x}')\big). \tag{2.26}$$

To set the stage for Gaussian process regression in the next section, the following discussion shows how the Bayesian linear regression basis function model implicitly defines a kernel function, and subsequently leads to a particular class of regression known as *linear smoothers* (Hastie & Tibshirani, 1990). For convenience, consider the case where the posterior predictive mean (Equation 2.23) is evaluated at the location of the training inputs, such that $\boldsymbol{\phi}_* = \boldsymbol{\phi}$. By then expanding $\boldsymbol{\phi}_*^{\mathrm{T}}\tilde{\boldsymbol{\mu}}$, we notice how the solution can be expressed as a linear combination of the training outputs:

$$\bar{\boldsymbol{f}}_* = \sum_{i=1}^{n}(\boldsymbol{\phi}_*^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\phi}_i^{\mathrm{T}}\mathrm{y}_i)/\sigma^2 + \boldsymbol{\phi}_*^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \mathbf{h}(\mathbf{x}_*)\mathbf{y} + \boldsymbol{\phi}_*^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \tag{2.27}$$

or for a zero-mean prior:

$$\bar{\boldsymbol{f}}_* = \mathbf{h}(\mathbf{x}_*)\mathbf{y}. \tag{2.28}$$

Here $\mathbf{h}(\mathbf{x}_*) = (\boldsymbol{\phi}_*^{\mathrm{T}}\tilde{\boldsymbol{\Sigma}}\boldsymbol{\phi})/\sigma^2$ represents an inner product in the feature space defined by $\psi(\mathbf{x}_*) = (\tilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\phi}_*)/\sigma$, hence the Bayesian linear regression basis function model

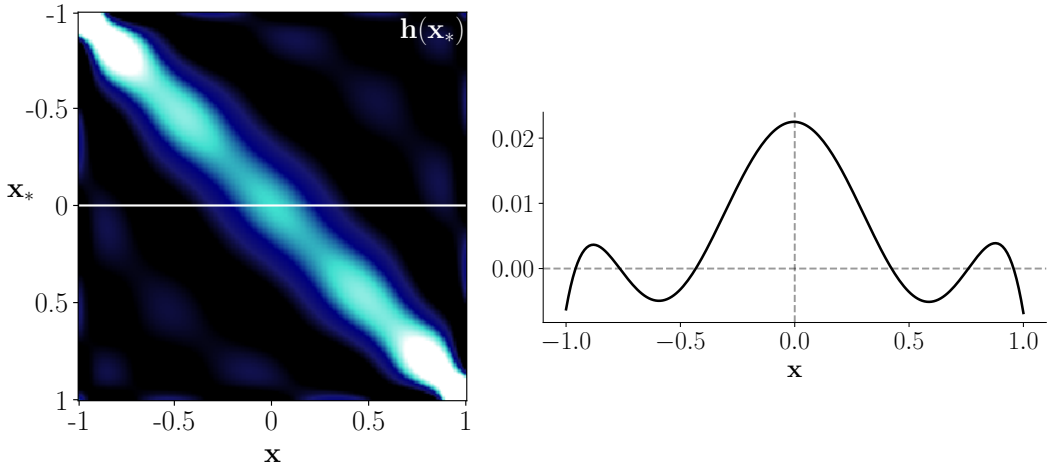


Figure 2.8: The equivalent kernel $\mathbf{h}(\mathbf{x}_*)$ (left), evaluated at the location of the training inputs ($\mathbf{x}_* = \mathbf{x}$), using polynomial basis functions, with $n = 200$ and $m = 6$. A profile (white line) is extracted along the kernel at $\mathrm{x}_* = 0$ (right), illustrating the localised nature of this weight function (training locations closer to x_* are assigned higher weights). Adapted from Bishop (2006).

implicitly defines a kernel function $\mathbf{h}(\mathbf{x}_*) = \psi(\mathbf{x}_*)^{\mathrm{T}}\psi(\mathbf{x})$, known as the *equivalent kernel*, or *idealised weight function* (Silverman, 1984). Equation (2.28) is therefore a linear smoother, whereby predictions at test input locations are given by a weighted average (smoothing) of the outputs, with weights defined by $\mathbf{h}(\mathbf{x}_*)$. We can see this more clearly in Figure 2.8, by extracting the corresponding weight vector for the test input $\mathrm{x}_* = 0$, and noticing how the equivalent kernel produces a form of localised weighting, with training points closer to x_* being assigned higher weights than those further away. It is also interesting to note that by evaluating the predictive distribution at the training locations, the equivalent kernel is equal to the posterior predictive covariance matrix, scaled by the noise precision factor (inverse variance) $\mathbf{h}(\mathbf{x}_*) = \boldsymbol{\sigma}^2_{\boldsymbol{f}_*}\sigma^{-2}$ (cf. Equation 2.23). From this it is clear to see why kernel functions are often referred to as *covariance functions*, although it should be noted that this is only the case for *positive semi-definite* kernels; which can be understood as a symmetric kernel $k(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{n\times n}$ (as in Figure 2.8), whose eigenvalues are strictly non-negative, and for any arbitrary vector $\boldsymbol{v} \in \mathbb{R}^n$, the following condition is held: $\boldsymbol{v}^{\mathrm{T}}k(\mathbf{x}, \mathbf{x}')\boldsymbol{v} \geq 0$ (see section 4.1 of Rasmussen & Williams 2006 for more details).

The next section introduces the first of the novel machine learning approaches of this thesis, Gaussian process regression. This technique incorporates all of the principles that have been outlined so far, into a powerful non-parametric probabilistic framework for learning complex functions, and forms the basis of results presented in chapters 3 and 4.

Gaussian process regression In the section on Bayesian linear regression, Gaussian distributions were outlined as useful tools for defining probability distributions over random variables. An extension to this are Gaussian processes, which define the properties of stochastic (random) functions. Rasmussen & Williams (2006) define a Gaussian Process ($\mathcal{GP}$) as a collection of random variables, any finite number of which form a joint Gaussian distribution. As an example, the random variables $\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{a}, \boldsymbol{\Sigma}_{\mathbf{aa}})$ and $\mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{bb}})$ could be considered to form a joint Gaussian distribution $\mathbf{c}$ (see Figure 2.9), whose mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ can be broken down into a set of *partitions*, which together describe the mean and covariance of (and between) $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{c} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_\mathbf{a} \\ \boldsymbol{\mu}_\mathbf{b} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{aa}} & \boldsymbol{\Sigma}_{\mathbf{ab}} \\ \boldsymbol{\Sigma}_{\mathbf{ba}} & \boldsymbol{\Sigma}_{\mathbf{bb}} \end{bmatrix} \right). \tag{2.29}$$

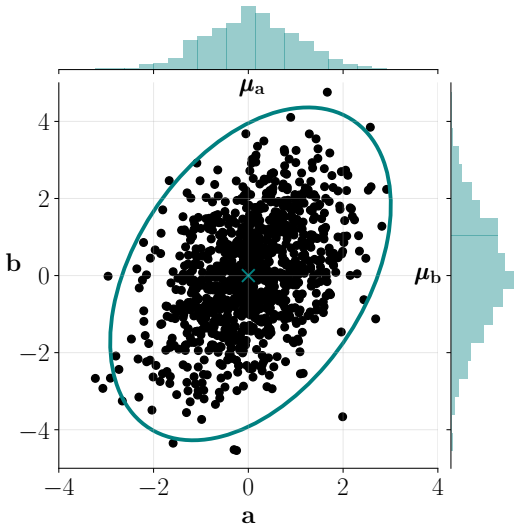


Figure 2.9: Synthetic example of two random variables $\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{a}, \boldsymbol{\Sigma}_{\mathbf{aa}})$ and $\mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{bb}})$ forming a joint Gaussian distribution $\mathbf{c}$, whose mean and 3 standard deviations are marked by the teal cross and ellipse respectively.

In Gaussian Process Regression (GPR), the functional relationship between inputs and outputs

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon \quad , \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{2.30}$$

is considered to be a $\mathcal{GP}$, where once again $\boldsymbol{\varepsilon}$ is assumed to represent independent and identically distributed random noise with mean 0 and variance σ^2. Learning the function $f(\mathbf{x})$ is then achieved through the Bayesian principles that were outlined previously, however in this case rather than placing prior distributions over parameters, prior probabilities are placed over all possible functions directly. As it happens, this results in a very useful non-parametric approach to learning functions, as only two

properties are needed to explicitly define a $\mathcal{GP}$ prior over functions:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{2.31}$$

Specifically, a mean $m(\mathbf{x})$ and covariance $k(\mathbf{x}, \mathbf{x}')$, or kernel, function:

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}\big[\big(f(\mathbf{x}) - m(\mathbf{x})\big)\big(f(\mathbf{x}') - m(\mathbf{x}')\big)\big]. \end{aligned} \tag{2.32}$$

In principle, GPR then assigns higher prior probabilities to functions which exhibit similar characteristics to those defined by the choice of $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$. The covariance function in particular bares significant importance, for it can describe a wide variety of properties, including whether the function is perhaps periodic, or orientation dependent (anisotropic), or smooth, or non-stationary (the smoothness varies for different locations of $\mathbf{x}$), or simply linear *et cetera*. A very common covariance function is the *squared exponential* function:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(- \frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2} \right), \tag{2.33}$$

also known as the Gaussian kernel, or radial basis function. The term $\|\cdot\|$ computes the Euclidean distance between the vectors $\mathbf{x}$ and $\mathbf{x}'$, and σ_f^2 and ℓ are known as *hyperparameters* of the covariance function, where each takes a real value greater than zero (see section 2.2.2). The squared exponential covariance function is a particularly smooth class of kernels, which actually corresponds to the linear combination of an infinite number of Gaussian basis functions (see e.g., MacKay 1998). From the definition above, we notice that the covariance function takes two arbitrary inputs, which in the case of GPR, correspond to either the training inputs $\mathbf{x}$, or the test inputs $\mathbf{x}_*$. Given the choice of $\mathcal{GP}$ prior, it is therefore possible to construct a series of kernel functions that reflect the prior covariance over all possible combinations of inputs:

$$\begin{aligned} \mathbf{K} &= k(\mathbf{x}, \mathbf{x}) + \boldsymbol{V} \quad , \quad \boldsymbol{V} = \sigma^2 \boldsymbol{I} \\ \boldsymbol{k}_* &= k(\mathbf{x}_*, \mathbf{x}_*) \\ \mathbf{k}_* &= k(\mathbf{x}, \mathbf{x}_*). \end{aligned} \tag{2.34}$$

In the traditional Bayesian setting, deriving the posterior predictive function values $\boldsymbol{f}_*$ is then achieved by updating the prior mean and covariance estimates. To do this, the same principles from Equation (2.29) can be used to show how $\boldsymbol{f}_*$ and $\mathbf{y}$ form a joint Gaussian distribution under the $\mathcal{GP}$ prior:

$$\begin{bmatrix} \boldsymbol{f}_* \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{x}_*) \\ m(\mathbf{x}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{k}_* & \mathbf{k}_*^{\mathrm{T}} \\ \mathbf{k}_* & \mathbf{K} \end{bmatrix} \right), \tag{2.35}$$

where it follows that the *marginal* distributions (under the prior) of both $\boldsymbol{f}_*$ and $\mathbf{y}$ are given as:

$$\boldsymbol{f}_* \sim \mathcal{N}\big(m(\mathbf{x}_*), \boldsymbol{k}_*\big) \quad , \quad \mathbf{y} \sim \mathcal{N}\big(m(\mathbf{x}), \mathbf{K}\big). \tag{2.36}$$

Next, in order to generate the conditional distribution $p(\boldsymbol{f}_*|\mathbf{y}, \mathbf{x}_*, \mathbf{x})$, Equation (2.35) needs to be *conditioned* on the outputs $\mathbf{y}$. This is achieved by computing the inverse of each of the relevant partitions of the joint prior covariance, using the following identity after Press et al. (1992):

$$\begin{bmatrix} \boldsymbol{k}_* & \mathbf{k}_*^{\mathrm{T}} \\ \mathbf{k}_* & \mathbf{K} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}_*}^{-1} & \boldsymbol{\Sigma}_{\mathbf{k}_*^{\mathrm{T}}}^{-1} \\ \boldsymbol{\Sigma}_{\mathbf{k}_*}^{-1} & \boldsymbol{\Sigma}_{\mathbf{K}}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{Z} & -\mathbf{Z}\mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1} \\ -\mathbf{K}^{-1}\mathbf{k}_*\mathbf{Z} & \mathbf{K}^{-1} + \mathbf{K}^{-1}\mathbf{k}_*\mathbf{Z}\mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1} \end{bmatrix} \tag{2.37}$$

where

$$\mathbf{Z} = \big(\boldsymbol{k}_* - \mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k}_*\big)^{-1}.$$

The quadratic form of the conditional Gaussian $p(\boldsymbol{f}_*|\mathbf{y}, \mathbf{x}_*, \mathbf{x})$ can then be broken down into its individual partitions:

$$\begin{aligned} -\frac{1}{2}(\boldsymbol{f}_* - \bar{\boldsymbol{f}}_*)^{\mathrm{T}}\boldsymbol{\sigma}_{\boldsymbol{f}_*}^{-2}(\boldsymbol{f}_* - \bar{\boldsymbol{f}}_*) = & \\ & -\frac{1}{2}\big(\boldsymbol{f}_* - m(\mathbf{x}_*)\big)^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{k}_*}^{-1}\big(\boldsymbol{f}_* - m(\mathbf{x}_*)\big) \\ & -\frac{1}{2}\big(\boldsymbol{f}_* - m(\mathbf{x}_*)\big)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{k}_*^{\mathrm{T}}}^{-1}\big(\mathbf{y} - m(\mathbf{x})\big) \\ & -\frac{1}{2}\big(\mathbf{y} - m(\mathbf{x})\big)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{k}_*}^{-1}\big(\boldsymbol{f}_* - m(\mathbf{x}_*)\big) \\ & -\frac{1}{2}\big(\mathbf{y} - m(\mathbf{x})\big)^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{K}}^{-1}\big(\mathbf{y} - m(\mathbf{x})\big). \end{aligned} \tag{2.38}$$

Using the same principles as in Equation (2.20), the mean $\bar{\boldsymbol{f}}_*$ and covariance $\boldsymbol{\sigma}_{\boldsymbol{f}_*}^2$ of the posterior predictive distribution for GPR can be computed by *completing the square* and utilising the equivalent expression for the Gaussian quadratic from Equation (2.21):

$$\begin{aligned} \bar{\boldsymbol{f}}_* &= m(\mathbf{x}_*) + \mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1}\big(\mathbf{y} - m(\mathbf{x})\big) \\ \boldsymbol{\sigma}_{\boldsymbol{f}_*}^2 &= \boldsymbol{k}_* - \mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k}_*. \end{aligned} \tag{2.39}$$

Figure 2.10 presents a synthetic example of GPR, using the same observations as in the non-linear OLS example from Figure 2.3. Here the $\mathcal{GP}$ prior over functions is given as zero-mean, with covariance defined by the squared exponential kernel (Equation 2.33). From Figure 2.10a we can see more clearly how the $\mathcal{GP}$ prior represents a prior over functions, as there are in principle an infinite number of possible functions which could be drawn from the shaded region. In this case, three different possibilities are shown. The predictions using GPR (Figure 2.10b) appear to be an improvement over the OLS basis function model from Figure 2.3, with significantly less over-fitting in areas with larger noise content. Furthermore we can see how GPR provides

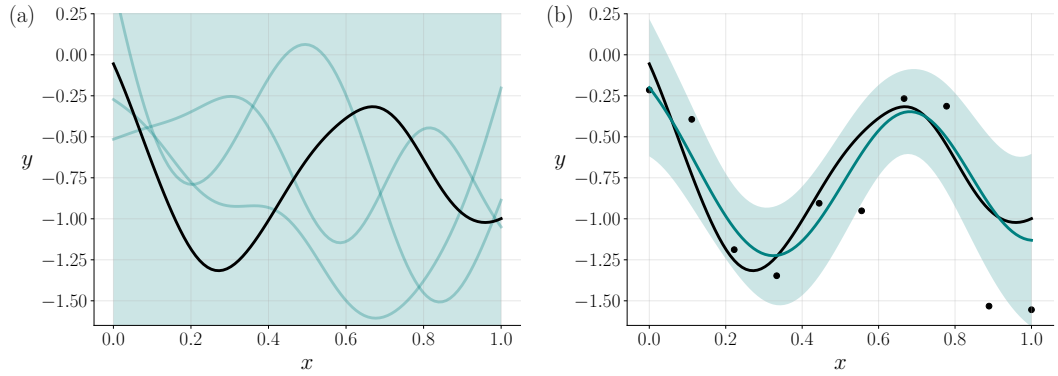


Figure 2.10: Synthetic example of 1-D GPR, using a zero-mean squared exponential covariance function ($\sigma_f^2 = 1$, $\ell = 0.2$). (a) The shaded area represents the prior space of functions, from which 3 random samples have been drawn (teal curves). The black curve is the true function f we aim to estimate. (b) The posterior predictive mean (teal curve) and uncertainty (1 standard deviation), computed from Equation (2.39).

the added benefit of uncertainty estimates, where in this case the true function lies within 1 standard deviation (shaded region) of the posterior predictive mean estimate.

Once again, it is worth drawing parallels between GPR and some of the concepts which were outlined in the previous sections. First, consider the Bayesian linear regression problem from Equation (2.13), where the outputs were assumed to be linear functions of the inputs $f(\boldsymbol{\phi}) = \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\beta}$. This means that any given value of $\boldsymbol{\beta}$ produces a corresponding value of the function $f(\boldsymbol{\phi})$. Therefore, placing a prior over the regression parameters automatically places a prior over functions, and as such, Bayesian linear regression corresponds to a specific example of GPR which assigns higher probabilities to linear functions. For a generic prior over the parameters $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the corresponding prior mean and covariance function for GPR follows:

$$\begin{aligned} m(\mathbf{x}) &= \boldsymbol{\phi}^{\mathrm{T}}\mathbb{E}[\boldsymbol{\beta}] = \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\mu} \\ k(\mathbf{x}, \mathbf{x}') &= \boldsymbol{\phi}^{\mathrm{T}}\mathbb{E}[\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}]\boldsymbol{\phi} = \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}, \end{aligned} \tag{2.40}$$

which leads to the posterior predictive GPR equations:

$$\begin{aligned} \bar{\boldsymbol{f}}_* &= \boldsymbol{\phi}_*^{\mathrm{T}}\boldsymbol{\mu} + \boldsymbol{\phi}_*^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}\big(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi} + \boldsymbol{V}\big)^{-1}\big(\mathbf{y} - \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\mu}\big) \\ \boldsymbol{\sigma}_{\boldsymbol{f}_*}^2 &= \boldsymbol{\phi}_*^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}_* - \boldsymbol{\phi}_*^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}\big(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi} + \boldsymbol{V}\big)^{-1}\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}_*. \end{aligned} \tag{2.41}$$

Although expressed here in a different form, this produces an identical solution to that of Equation (2.23). The only difference is that in the Bayesian linear regression problem, a $N \times N$ parameter matrix $(\boldsymbol{\phi}\boldsymbol{V}^{-1}\boldsymbol{\phi}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1})^{-1}$ must be inverted, while in the GPR case it is a $n \times n$ covariance matrix $(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi} + \boldsymbol{V})^{-1}$. From these equations, we can also see how GPR is a linear smoother, with a weight function defined by $\boldsymbol{\phi}_*^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}\big(\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi} + \boldsymbol{V}\big)^{-1}$ – or equally $\mathbf{k}_*^{\mathrm{T}}\mathbf{K}^{-1}$ from Equation (2.39).

2.2.2 Bayesian model selection

At some points throughout the GPR workflow, various decisions were made in some form or another, whether related to the choice of prior covariance function, or by assigning values to each of the possible hyperparameters. The concluding discussion on Bayesian linear regression briefly mentioned how the hyperparameters play a key role in reducing over-fitting, therefore conceivably optimal values should be assigned to each of these in order to generate reliable predictions. Generally speaking, any unique choice of $\mathcal{GP}$ prior (i.e., choice of $m(\mathbf{x})$, $k(\mathbf{x}, \mathbf{x}')$ and hyperparameters) produces a unique model with which to undertake the supervised learning problem, however deciding on which model is most suitable for a given application may not be immediately obvious. Model selection is therefore concerned with finding the model that is most *preferred* by the data; which as it turns out, can be achieved rather effectively through the Bayesian principles that have already been outlined. Consider for a moment a generic $\mathcal{GP}$ prior over functions with mean and covariance function given by $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ respectively. This prior can be referred to as a specific model Λ_i, out of any number of different models $i = \{1, 2, \ldots, I\}$ which may perhaps assign different classes of prior covariance functions, or different values for their hyperparameters. Through the Bayesian formalism, the posterior probability of any given model can be expressed as:

$$p(\Lambda_i|\mathbf{y}, \mathbf{x}) \propto p(\mathbf{y}|\mathbf{x}, \Lambda_i)p(\Lambda_i).$$

By then assuming that each model has equal prior probability $p(\Lambda_1) = p(\Lambda_2) = \ldots$ *et cetera*, the distribution $p(\mathbf{y}|\mathbf{x}, \Lambda_i)$ provides a direct indication as to which model is more likely to have generated the data. Notice that $p(\mathbf{y}|\mathbf{x}, \Lambda_i)$ was already introduced as the marginal likelihood (also known as the *evidence*) in Equation (2.17), which corresponded to integrating over the regression parameters in Bayesian linear regression. In the non-parametric GPR approach, the marginal likelihood corresponds to integrating over the function values – which was also already done in Equation (2.36). Writing it in full here, we have:

$$p(\mathbf{y}|\mathbf{x}, \Lambda_i) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\mathbf{K}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - m(\mathbf{x}))^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - m(\mathbf{x})) \right). \tag{2.42}$$

Finding the model which then maximises the log of the marginal likelihood:

$$\ln p(\mathbf{y}|\mathbf{x}, \Lambda_i) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{K}| - \frac{1}{2}(\mathbf{y} - m(\mathbf{x}))^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - m(\mathbf{x})), \tag{2.43}$$

leads to a very common model selection technique in Bayesian statistics known as *empirical Bayes* (Bernardo & Smith, 2009), or *type-II maximum likelihood* (Berger, 2013). This technique corresponds to maximising the probability of the observations

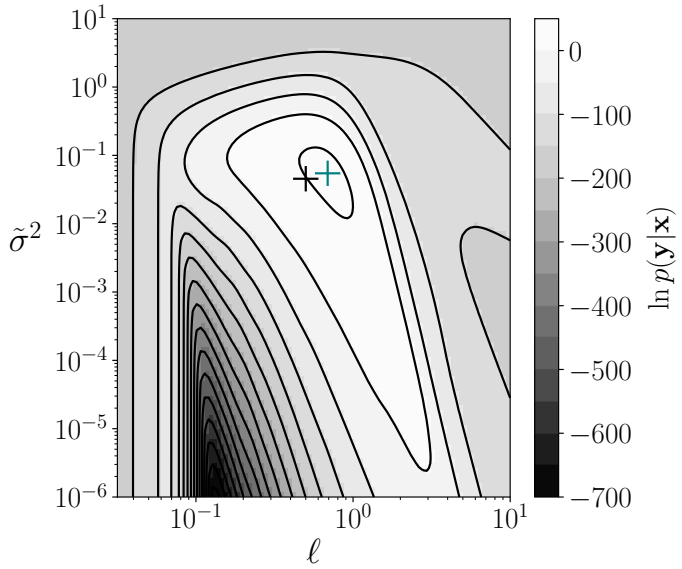


Figure 2.11: Bayesian model selection through the empirical Bayes approach. The contour map shows values of the log marginal likelihood for different combinations of hyperparameters $\tilde{\sigma}^2$ and ℓ. The true values from which the data were generated are given by the black cross-hair $\tilde{\sigma}^2 = 0.045$, $\ell = 0.5$. The teal cross-hair is then the values which maximise the log marginal likelihood function $\tilde{\sigma}^2 = 0.054$, $\ell = 0.68$.

$\mathbf{y}$ under the choice of model Λ_i, and provides a very useful solution to the model selection problem. For one, it allows all of the available training data to be used for determining the most optimal model (as opposed to alternative approaches such as cross-validation, where the training data must be divided into disjoint training and validation sets). Secondly, through inspection of Equation (2.43), we can see how this technique not only selects the model which best fits the data (given by $-(\mathbf{y} - m(\mathbf{x}))^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{y} - m(\mathbf{x}))/2$), but also penalises complex models (via $-\ln|\mathbf{K}|/2$), and as such directly embodies the principles of Occam's razor whereby, given two models which fit the data equally well, the log marginal likelihood will show preference towards the model with lower complexity (see chapter 5 of Rasmussen & Williams 2006 for an illuminating discussion).

As a final segment to this discussion on supervised learning, Figure 2.11 presents a synthetic example of the empirical Bayes approach, which attempts to learn a set of hyperparameters $\boldsymbol{\theta} = (\ell, \sigma_f^2, \sigma^2)^{\mathrm{T}}$ for the squared exponential covariance function from Equation (2.33) . Here ℓ is often referred to as the *correlation length scale*, which controls the smoothness of the function values, and can be equated to a distance in the input space for which the observations are expected to to remain correlated. σ_f^2 is then a scaling factor which governs the variance of the function values, and σ^2 is the noise variance which is added to the diagonal entries of the covariance matrix (e.g., Equation 2.34). In this synthetic example, the set of observations are generated

by drawing $n = 100$ samples from a multivariate Gaussian distribution with mean 0 and covariance given by the squared exponential function, with hyperparameters $\ell = 0.5$, $\sigma_f^2 = 2.2$, $\sigma^2 = 0.1$. After Sollich (2005), it is possible to then define $\tilde{\sigma}^2 = \sigma^2/\sigma_f^2 = 0.045$ and work with the set of hyperparameters $\boldsymbol{\theta} = (\ell, \sigma_f^2, \tilde{\sigma}^2)^{\mathrm{T}}$, such that σ_f^2 can be optimised analytically to obtain $\sigma_f^2 = n^{-1}(\mathbf{y} - m(\mathbf{x}))^{\mathrm{T}}\tilde{\mathbf{K}}^{-1}(\mathbf{y} - m(\mathbf{x}))$, where $\tilde{\mathbf{K}} = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2/2\ell^2) + \tilde{\sigma}^2\boldsymbol{I}$. This way, only ℓ and $\tilde{\sigma}^2$ need to be optimised numerically. Maximising the log marginal likelihood in Figure 2.11 leads to the recovered hyperparameters $\tilde{\sigma}^2 = 0.054$ and $\ell = 0.68$, which are relatively close to the true values. Note however that this approach generates point-estimates of the hyperparameters, whereas in the *fully* Bayesian approach, distributions of values are typically generated through Bayes' Law. Assigning prior distributions over the hyperparameters however, requires the evaluation of numerous integrals (see section 3.5 of Bishop 2006), which is no longer analytically tractable. In this situation approximation techniques such a Markov Chain Monte Carlo sampling are required (e.g., Neal 1993), which in many cases can be computationally prohibitive.

2.3 Unsupervised learning

The previous section was concerned with learning the function which mapped a set of one-dimensional inputs $\mathbf{x}$ to a set of one-dimensional outputs $\mathbf{y}$. This can of course be extended to higher dimensions by incorporating multiple predictor variables as additional columns of $\mathbf{x}^{\mathrm{T}}$, or similarly modelling multiple outputs, which then formulates a multi-task learning problem (see section 6.1). In contrast, the unsupervised learning framework dispenses with the outputs and instead aims to explore the hidden properties which are unique to the set of input observations only. In most practical applications, unsupervised learning deals with high-dimensional data sets (i.e., where the number of columns of $\mathbf{x}^{\mathrm{T}}$ is large), and the goal is to project the key features within those data onto a lower-dimensional subspace, or *manifold*, through some form of *dimensionality reduction*. In the case of climate data, the set of observations $\mathbf{x}$ often corresponds to time series information measuring a particular component of the Earth's climate at n regularly sampled points in time and at P fixed geographical locations, such that $\mathbf{x}^{\mathrm{T}}$ is a $n \times P$ matrix, where $P \gg n$. Section 1.1.3 outlined how teleconnections are often associated with large-scale patterns of variability of a particular climate field, such as sea-level pressure variability over the Arctic Ocean in the case of the Arctic Oscillation (AO). It therefore seems reasonable to expect that if the set of inputs were, for example, a reanalysis data set containing estimates of sea-level pressure north of 60°, then many of the columns of $\mathbf{x}^{\mathrm{T}}$ would be correlated in some way due to the influence of the AO. This means that many of the columns of $\mathbf{x}^{\mathrm{T}}$ provide redundant information and superfluous dimensionality to our data, and what would perhaps be more useful is if we could extract the dominant

patterns, or *modes*, of variability related to the AO itself. In climate science we are often interested in these dominant (lower-dimensional) modes of variability as, for one, high-dimensional spaces are typically undesired due to observations becoming increasingly sparse with increasing dimensionality (Bellman, 2015), but also because these modes more generally contain key predictive information about the state of the climate system. It is therefore not uncommon for unsupervised learning to be used as a precursor for prediction (forecasting) purposes in a supervised learning framework (see chapter 3). In the following sections, two approaches towards unsupervised learning are outlined. First, a brief overview of principal component analysis is provided, which can be thought of as the industry standard for analysing spatio-temporal patterns in climate time series data sets. Following which, the second novel machine learning technique of this thesis is introduced, which is based on a combination of cluster analysis and complex network theory.

2.3.1 Principal component analysis

Principal Component Analysis (PCA) – also known as Empirical Orthogonal Function (EOF) analysis – is a method for dimensionality reduction which has a long history in climate data science, and has been used extensively in applications ranging from analysis of climatological teleconnections (Kutzbach 1967; North et al. 1982; Zhang et al. 1997; Thompson & Wallace 1998; Ambaum et al. 2001) to time series forecasting (Lorenz 1956; Walsh 1980; Tangang et al. 1998). PCA aims to decompose a given data set into dominant spatial and temporal modes of variability, known as the EOFs and Principal Components (PCs) of the data respectively. The derived modes correspond to a set of orthogonal vectors which retain specific fractions of the variance of the input signal, where each consecutive mode describes a decreasing percentage of explained variance relative to the previous modes. Commonly, the mode which explains the highest percentage of variance of the input signal is referred to as the *leading* EOF or PC of that particular data field, and indeed many climatological teleconnections can be defined as such, including e.g., the North Atlantic Oscillation (Ambaum et al., 2001), or the El Niño Southern Oscillation (Zhang et al., 1997).

Consider the case where we have a de-trended (zero-mean) time series data set $\mathbf{x} = \{\boldsymbol{x}_p\}_{p=1}^{P}$ representing n regularly sampled observations in time $\boldsymbol{x}_p = (x_{1p}, x_{2p}, \ldots, x_{np})$, at P fixed geographical locations, such that $\mathbf{x} \in \mathbb{R}^{P \times n}$. The data are first weighted by an appropriate grid cell weighting $\mathbf{X} = \{\boldsymbol{x}_p \sqrt{\mathrm{w}_p}\}_{p=1}^{P}$, which may correspond to $\mathrm{w}_p = \cos(\omega_p)$ for a regular latitude-longitude grid (ω_p is the latitude of grid cell p), or simply $\mathrm{w}_p = d_p$ for an area grid (d_p is the area in km^2 of grid cell p). The linear temporal covariance of the weighted data matrix is then

given as:

$$\mathrm{C}_{ij} = \frac{1}{P-1}\sum_{p=1}^{P}(\mathrm{X}_{pi} - \mathbb{E}[\mathbf{X}_i])(\mathrm{X}_{pj} - \mathbb{E}[\mathbf{X}_j]), \tag{2.44}$$

where $\mathbb{E}[\mathbf{X}_i]$ is the average of all grid points at time index i (this is not generally zero, as the grid points are de-trended independently). Note here how the definition of the covariance matrix involves an inner product, therefore in theory the *kernel trick* (discussed in section 2.2.1) could be used to represent this inner product in some higher-dimensional feature space, leading to an alternative non-linear form of PCA known as kernel PCA (Schölkopf et al., 1998) – see also other non-linear extensions such as Non-linear Laplacian Spectral Analysis (Giannakis & Majda, 2012). For simplicity, the linear definition above is used here. Next, solving the eigendecomposition problem leads to:

$$\mathbf{CV} = \mathbf{VS}, \tag{2.45}$$

where $\mathbf{V}$ is a $n \times n$ matrix whose columns are the set of orthogonal[2] eigenvectors of $\mathbf{C}$, which are, in fact, the temporal modes of variability, or PCs, of the input signal. The diagonal matrix $\mathbf{S}$ then contains the eigenvalues of $\mathbf{C}$, where generally speaking, the eigenvector which corresponds to the largest eigenvalue, is the direction along which the variance of the input data is largest. The fraction of variance explained by mode i, for example, is given by $\mathrm{S}_{ii}/\sum_{j=1}^{n}\mathrm{S}_{jj}$. Each subsequent column of $\mathbf{V}$ then explains, in a statistical sense, a decreasing percentage of variance of the input signal, in a direction that is orthogonal to the previous leading vectors. Deriving the spatial modes of variability is achieved by first computing:

$$\mathbf{U} = \mathbf{XVS}^{-1/2}, \tag{2.46}$$

and subsequently dividing by the grid cell weights. Here $\mathbf{U}$ is a $P \times n$ matrix whose columns are ordered in terms of decreasing explained variance of the input signal, and $\mathbf{S}^{-1/2}$ is a diagonal matrix whose entries are defined by $(\sqrt{\mathrm{S}_{11}}^{-1}, \sqrt{\mathrm{S}_{22}}^{-1}, \ldots, \sqrt{\mathrm{S}_{nn}}^{-1})$.

Figure 2.12 shows an example of the 3 leading spatial modes of variability in winter (December, January, February) sea ice concentration anomalies in the Arctic, generated via PCA. The input data here $\mathbf{X} \in \mathbb{R}^{48879\times 43}$ were taken from the National Snow and Ice Data Center (NSIDC; Cavalieri et al. 1996 – see section 1.1). Concentration anomalies then correspond to de-trending each grid cell between 1979 and 2021. The leading mode of variability here explains approximately 22% of the variance of the input signal, where we can see how each subsequent mode explains a lower percentage of variance than the last. The darker colours in each

[2]A way to check that the columns of $\mathbf{V}$ are orthogonal is by confirming that $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \boldsymbol{I}$.

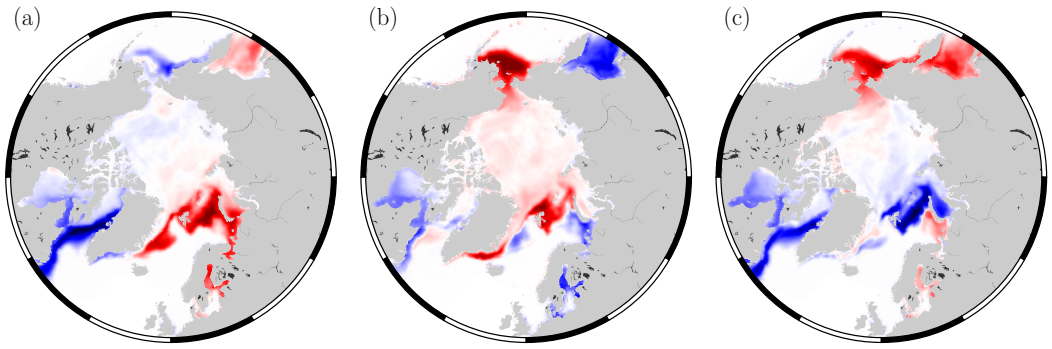


Figure 2.12: Modes of winter (DJF) sea ice concentration variability (EOFs) between 1979 and 2021, from PCA. (a) Leading mode explaining $\sim 22\%$ of the variability of the input signal. (b) Second leading mode explaining $\sim 13\%$ variability. (c) Third leading mode explaining $\sim 10\%$ variability.

EOF map show the dominant regions of winter sea ice variability, where colours provide an indication as to the phase nature of the variability. For example, regions which are blue are out of phase with areas which are red, meaning that when one area experiences higher than average concentration, the other experiences lower than average concentration. In the next two sections, clustering is outlined as an alternative framework for deriving similar spatio-temporal patterns of variability, and also information about higher-order structures of the climate system by adopting a complex network approach.

2.3.2 Cluster analysis

Cluster analysis is an alternative method for dimensionality reduction which has been utilised across a wide variety of disciplines ranging from ecology (James & McCulloch 1990; McKenna Jr 2003), through to social education (Myers III & Fouts, 1992), and forensic psychology (Spaans et al., 2009). Its applications in polar climate science have historically been for the purpose of detecting sea ice properties from satellite radiometers (Comiso, 1983, 1986) and altimeters (Kwok et al., 1992), and more recently for investigating high-latitude atmospheric aerosol and particle patterns (Dall et al. 2017, 2018; Lange et al. 2018). Generally, cluster analysis can be considered as an approach which seeks to group data points together into a number of distinct partitions, or *clusters*, based on some metric of similarity. These definitions of 'similarity' and 'clusters' are somewhat heuristic, which has resulted in a plethora of clustering algorithms being developed over recent decades, and subsequently choosing the most suitable one for a particular problem may not be immediately obvious – the reader is referred to Jain et al. (1999) and Saxena et al. (2017) for excellent reviews on the many types of algorithms available. In this section, a *grid-based* clustering algorithm is introduced for identifying modes of variability within climate

time series data products; an approach which has previously been explored in the context of mid-latitude sea-surface temperature patterns (Fountalis et al., 2014). In the application here, this approach is extended to polar climate data sets.

Consider once again the zero-mean, grid-cell-weighted $P \times n$ data matrix of observations $\mathbf{X}$ from section 2.3.1. From here, the pairwise correlations are computed between all grid points:

$$\mathrm{R}_{pq} = \frac{\sum_{i=1}^{n}(\mathrm{X}_{pi} - \mathbb{E}[\mathbf{X}_p])(\mathrm{X}_{qi} - \mathbb{E}[\mathbf{X}_q])}{\sqrt{\sum_{i=1}^{n}(\mathrm{X}_{pi} - \mathbb{E}[\mathbf{X}_p])^2 \sum_{i=1}^{n}(\mathrm{X}_{qi} - \mathbb{E}[\mathbf{X}_q])^2}}. \tag{2.47}$$

Notice the difference between this definition and Equation (2.44). In the PCA application, $\mathbf{C}$ was a $n \times n$ covariance matrix, whereas here $\mathbf{R}$ is a $P \times P$ correlation matrix. Furthermore, $\mathbb{E}[\mathbf{X}_p]$ equates here to the average across the temporal dimension at grid point p, which in this example is in fact zero. In any case, the correlation matrix above is used to define the similarity between grid cells, which will establish whether cells should be grouped together to form a cluster. This grid-based approach follows a *greedy* algorithm, whereby individual grid cells continuously search for neighbours to group with, providing the correlation between them exceeds a certain threshold. In the fashion of a greedy algorithm, this form of local clustering aims to be heuristically optimal in the sense that small-scale homogeneities identified with local clustering will approximate the large-scale homogeneities across the whole data set (Cormen et al., 2009). Subsequently, for grid cells to form a cluster $\mathcal{C}_k$, 3 criteria must be adhered to. The first states that the minimum number of grid cells in any cluster must be at least 2:

- $|\mathcal{C}_k| \geq 2$.

The second states that, for any two grid cells p and q that are both part of a cluster $\mathcal{C}_k$, there must be a connected path (or continuous function h) that joins p to q:

- $\forall\ (p, q) \in\ \mathcal{C}_k, \quad h : [0, 1] \to \mathcal{C}_k$ such that $h(0) = p$ and $h(1) = q$.

Finally, the third condition states that the average correlation of all grid cells in each cluster must be greater than a pre-defined threshold factor τ:

- $\dfrac{\sum\limits_{p \neq q \in \mathcal{C}_k} \mathrm{R}_{pq}}{|\mathcal{C}_k|(|\mathcal{C}_k| - 1)} > \tau,$

which here is defined as the mean of all positive correlation values (from $\mathbf{R}$) whose associated p-values are less than some pre-defined significance level η. The p-values can be computed from a 1-sided t-test:

$$t\text{-}score = \mathrm{R}_{pq}\sqrt{\frac{n-2}{1-\mathrm{R}_{pq}^2}}. \tag{2.48}$$

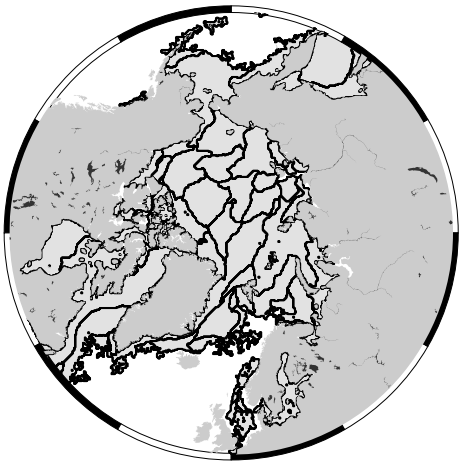


Figure 2.13: 80 spatially contiguous clusters ($\eta = 0.01$, see main text) generated from de-trended winter sea ice concentration time series data between 1979 and 2021. Each cluster represents a region of sea ice concentration which has behaved homogeneously over the length of the observational period.

Clustering then proceeds based on the above criteria until no further grid cells can be added to any cluster. After which, a second round of clustering is performed, whereby clusters are merged together providing the three outlined criteria are still upheld. This is achieved by a similar process to the generation of the initial clusters, except rather than grid cells searching for neighbours, clusters instead search for neighbouring clusters to determine whether, if merged, the average correlation between all the grid cells in a newly merged cluster exceeds τ. The neighbouring cluster with the greatest average correlation above τ after merging is then chosen. This continues until no further clusters can be merged.

Figure 2.13 shows the resultant clusters after performing the grid-based algorithm to the winter sea ice concentration data set described in section 2.3.1, with a significance level of $\eta = 0.01$ (see also Figure A.3 for comparisons with $\eta = 0.05$ and $\eta = 0.10$). The clusters are defined by a set of spatially contiguous areas, where each area represents a region of sea ice concentration that has behaved in a homogeneous way over the length of the time series record, and in this case corresponds to a particular spatial mode of sea ice variability (analogous to EOFs). We can immediately see here how clustering acts as a dimensionality reduction method, as the input data $\mathbf{X}$ have been decomposed from the original space of $P = 48879$ grid cells to $N = 80$ clusters. By adopting a complex network approach next, further information can then be extracted from these clusters, such as which represent the dominant, or *leading*, modes of sea ice variability, and subsequently how each are connected via a set of statistical interdependencies.

2.3.3 Complex networks

Networks provide a relatively simple and visual framework with which to analyse and display large volumes of data that typically represent complex physical systems. Their use across multiple disciplines has grown considerably over recent decades, with intuitive applications in computer science and social networks (Albert & Barabási 2002; Newman 2003; Boccaletti et al. 2006; Cohen & Havlin 2010) – known as *structural networks*, to more abstract applications in e.g., neuroscience (Zhou et al. 2007; Morabito et al. 2015; delEtoile & Adeli 2017), seismology (Abe & Suzuki, 2006), and climate science (Tsonis & Roebber 2004; Tsonis et al. 2006; Donges et al. 2009, 2015; Fountalis et al. 2014, 2018; Dijkstra et al. 2019) – known as *functional networks*. Climate network analysis was first introduced by Tsonis & Roebber (2004), and has subsequently proven to be a powerful tool kit for extracting significant statistical relationships from the multitude of entangled interactions within the climate system, and is a useful complement to standard methods of analysing patterns of climate variability, such as PCA (Donges et al., 2015). In general, a network can be seen as a group of vertices, or *nodes*, whereby each node k may be connected to any other node in the network l via a weighted edge, or *link*. In the common example of social networks, people are represented by nodes and their relationships by links. The weight of the link then depends on the strength of the relationship existing between the two people (nodes). In the case of the climate network, a network, or *graph*, of N nodes may correspond to time series data $\boldsymbol{G} = \{\boldsymbol{g}_k\}_{k=1}^{N}$ representing n regularly sampled observations in time $\boldsymbol{g}_k = (g_{1k}, g_{2k}, \ldots, g_{nk})$, at N fixed geographical locations, and the links may represent statistical interdependencies between any pair of node time series $\boldsymbol{g}_k$ and $\boldsymbol{g}_l$. In the application outlined here, the N clusters identified from the grid-based algorithm in the previous section (e.g., Figure 2.13) are used to represent the nodes of a complex network, and they subsequently form a connected graph by generating weighted links between them. To do this, the cumulative anomaly time series of each network node is computed, which is taken as the sum of the grid-cell-weighted de-trended time series of all cells within each respective network node:

$$\boldsymbol{g}_k = \sum_{p \in \mathcal{C}_k} \mathbf{X}_p, \tag{2.49}$$

where $\boldsymbol{g}_k$ represents the temporal mode of variability of each network node, and as such can be considered analogous to the PCs derived from PCA. Next, the link weights $\boldsymbol{w}$ are calculated as the temporal covariance between two network node anomaly time series:

$$w_{kl} = \frac{1}{n-1} \sum_{i=1}^{n} (g_{ik} - \mathbb{E}[\boldsymbol{g}_k])(g_{il} - \mathbb{E}[\boldsymbol{g}_l]). \tag{2.50}$$

Note that the weights are not normalised, such that the weight of the link is proportional to the power of the anomaly time series $\boldsymbol{g}_k$ and $\boldsymbol{g}_l$ in nodes $\mathcal{C}_k$ and $\mathcal{C}_l$, rather than just their correlation value (Fountalis et al., 2014). Furthermore, although the link weights have been defined based on linear covariance here, this could in theory be altered to incorporate non-linear metrics such as, e.g., mutual information (Donges et al., 2009), event synchronisation (Malik et al. 2012; Boers et al. 2014), or transfer entropy (Runge et al., 2012). The structure of the network can then be represented in an *adjacency matrix* $\mathbf{M} \in \mathbb{R}^{N \times N}$, that describes how all of the nodes are connected across the network:

$$\mathrm{M}_{kl} = \begin{cases} w_{kl} & \text{if } k \neq l \\ 0 & \text{otherwise} \end{cases}, \tag{2.51}$$

where the fact that the diagonal entries are zero implies that the network contains no self-loops (i.e., a node does not connect to itself). Chapter 3 will show how the adjacency matrix is a particularly useful concept, as it easily allows the connected structure of the network to be incorporated into other models, such as *graph kernels* (Vishwanathan et al., 2010) for regression analysis. Finally, the weighted degree, or *strength*, of a given network node $\mathcal{S}_k$ is defined here as the sum of the absolute value of all its associated link weights:

$$\mathcal{S}_k = \sum_{l=1}^{N} |\mathrm{M}_{kl}|. \tag{2.52}$$

The node with the highest strength is commonly referred to as the *hub* of the network (Tsonis & Roebber, 2004), and represents the leading mode of variability of the input data set $\mathbf{X}$. The links between nodes can then be seen as a proxy for teleconnections, which in this example are inherently instantaneous, as the covariance-based link weights are computed at zero-lag. The network framework can however be easily adapted to incorporate lagged correlations, and also additional components of the climate system (e.g., atmospheric and/or oceanic variables), at which point the network becomes a *multi-layer network* (see chapter 5).

Figure 2.14 shows the complex network representation of winter sea ice concentration anomalies, based on the clusters identified in Figure 2.13 (see also Figure A.4 for a comparison with clusters based on $\eta = 0.05$ and $\eta = 0.10$). Here we can see how the dominant mode of winter sea ice concentration variability (i.e., the node with the highest strength) extends from the Labrador Sea, through Baffin and Hudson bays, and up to the Canadian Archipelago. We also notice how the node links highlight a similar out-of-phase relationship between the east and west Atlantic (and Pacific) sectors of the Arctic (similar to the leading EOF in Figure

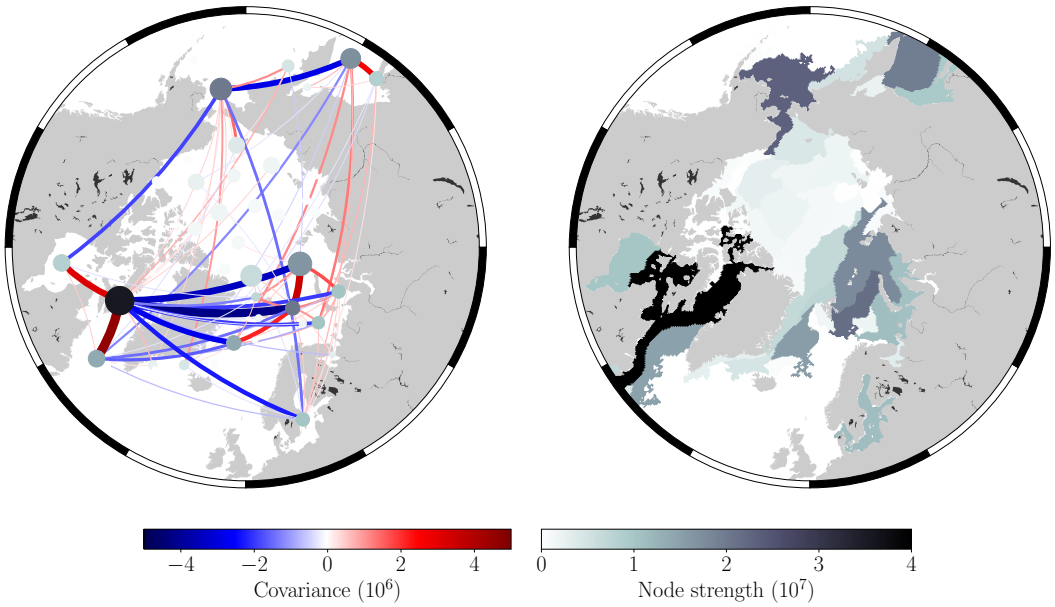


Figure 2.14: Complex network of winter sea ice concentration. The network can be visualised as node points (left) which take the centre position of each cluster from Figure 2.13, where the size of the node is proportional to the area of the cluster. Links signify the temporal covariance between network nodes, where the width of each link is proportional to the covariance. Only links which have a p-value < 0.1 are shown here, to aid visualisation. The network can also be visualised by weighting each grid cell by the strength of the node in which that cell belongs (right), producing a strength map of the network.

2.12), represented here by the strong negative network links. Each individual sector then shows the same phase of variability, given by the positive links. From this it is possible to see how complex networks and PCA are complementary in their ability to exploit spatio-temporal patterns of variability, however it could be argued that the network framework provides some advantages. Specifically, the network approach does not impose orthogonality constraints between modes, and it also does not mask patterns of lower variance. Furthermore, the network framework can in principle be used to derive higher-order information about the input data, such as the rate of information flow between nodes. This section does not go into further detail as these metrics do not form the basis of results presented in subsequent chapters, however the interested reader is referred to Donges et al. (2009, 2015) for more information.

2.4 Thesis aims

This thesis explores the application of the two previously outlined machine learning techniques, Gaussian process regression and complex networks, to a variety of problems facing the polar climate science community at present. Although these problems may seem broad in scope, they can each be considered as components of the larger sea ice prediction problem, and furthermore they provide individual frameworks

to showcase the flexibility of each of the novel machine learning methodologies. The results which will be presented are:

- In chapter 3 the problem of providing reliable seasonal predictions of the September sea ice cover will be addressed. September is the point of maximum sea ice melt in the Arctic, and as such it is a crucial time of the year for a wide variety of communities who rely on sea ice forecasts for safe and cost-effective passage through the Arctic. This study will look to exploit the predictability mechanism of sea ice persistence by computing complex networks of sea ice concentration data from the preceding June, July and August, and subsequently pass the connected structure of each of these networks into a Gaussian process regression forecast model in the form of a random walk graph kernel, in order to ultimately predict regional and pan-Arctic September extents for all years between 1985 and 2019.

- In chapter 4 the problem of merging multiple observation sets and improving data coverage will be considered. Satellite altimeters have been groundbreaking in our ability to remotely monitor the changes in Arctic sea ice thickness that have occurred over the past decade, however generating pan-Arctic observations typically takes up to one month. This has implications for our ability to understand how various chaotic events, such as weather, affect sea ice thickness on shorter time-scales, and also our ability to provide accurate initial conditions for seasonal sea ice forecasts in dynamical models. In this work, the Gaussian process regression methodology will be explored as a way to combine radar freeboard observations from the European Space Agency's altimeters: CryoSat-2, Sentinel-3A and Sentinel-3B, in order to produce daily pan-Arctic estimates of radar freeboard, along with uncertainty.

- Chapter 5 will assess the ability of the latest generation of coupled climate models to reflect one of the key atmospheric drivers of summer sea ice variability, and subsequently discuss the consequences for seasonal sea ice predictions in these dynamical models. Specifically, the complex networks methodology will be used to compare the dominant spatio-temporal patterns of variability in winter sea-level pressure, summer sea ice concentration, and subsequently compare the degree of connectivity between the winter Arctic Oscillation and summer sea ice concentration in both observations and CMIP6 models by using two metrics for comparing sets of networks. An assessment of how this teleconnection may be changing over time will also be outlined for both observations and models, as ice conditions change due to the effects of anthropogenic warming.

Chapter 3

Skilful sea ice forecasting

The following chapter is based on the article "Gregory, W., Tsamados, M., Stroeve, J. and Sollich, P., (2020). Regional September Sea Ice Forecasting with Complex Networks and Gaussian Processes. *Weather and Forecasting*, 35(3), pp.793-806." A novel statistical forecast model is presented which is able to generate reliable predictions of both pan-Arctic and regional September sea ice extents, with up to 3 months lead time. Here the complex network framework is used in order to exploit spatio-temporal patterns of sea ice concentration variability in the preceding 3 months of summer (June, July and August), which are then incorporated into a supervised learning linear Gaussian process regression forecast model to predict September extents. The regional forecasts are performed over 9 Arctic regions which encompass the most likely maritime navigation routes at the point of the summer minimum. Comparisons of the forecast skill achieved with this model are made with statistical and dynamical model forecasts from the existing literature, using either the metric of 'skill score' or 'anomaly correlation coefficient'. The score of pan-Arctic and various regional forecasts are also broken down across different portions of the time series record to understand whether the predictability of various regions may be changing over time.

3.1 Introduction

Across all months, September has exhibited one of the fastest rates of decline in Sea Ice Extent (SIE) since 1979 (Overland & Wang 2013; Serreze & Stroeve 2015), as well as the largest inter-annual variability (Stroeve & Notz, 2018). Significant anomalies such as the 1996 maximum and the 2007 and 2012 minima contribute largely to this variability and make seasonal predictions difficult. The extreme 2007 minimum in particular led to the initiation of the Sea Ice Outlook (SIO) project

by the Study of Environmental Arctic Change (SEARCH), with the initial aim of providing annual summaries of the expected pan-Arctic September SIE. In 2014 this was formally turned in to the Sea Ice Prediction Network through U.S. interagency funding, and expanded to include regional forecasts of September Sea Ice Concentration (SIC), sea ice probability, first ice-free date and first ice-advance date. Starting each May, the SIO solicits predictions of the September sea ice cover from the sea ice community based on forecasts made on the 1st day of June, July and August. Approaches vary from fully coupled ice-ocean-atmosphere or ice-ocean dynamical models, to statistical models and heuristic and/or best guesses. The ability of these forecast models to accurately capture the variability of Arctic sea ice, with lead times of days to months, is becoming increasingly important from an ecological, economical and societal standpoint. The opening of the Northern sea route for example has the potential to reduce the shipping route between Europe and Asia by 5000 nautical miles in comparison to the Suez Canal (Lee & Song, 2014), hence advanced timing of this opening, on the order of months, can make for cost-effective planning (Eguíluz et al., 2016).

While fully coupled climate models are intuitively the model of choice for sea ice forecasting, with a number of studies reporting high de-trended prediction skill (Chevallier et al. 2013; Sigmond et al. 2013; Wang et al. 2013; Msadek et al. 2014; Peterson et al. 2015; Bushuk et al. 2017), on average past summaries of SIO submissions (Stroeve et al. 2014c; Hamilton & Stroeve 2016) have shown that generally dynamical models do not significantly out-perform their statistical model counterparts. In fact the majority of models participating in the SIO (both statistical and dynamical) only perform relatively successfully when sea ice anomalies occur close to the long-term trend. This leaves a window open to explore statistical models, which have in some cases shown promise in exploiting sources of predictability with up to 4 months lead time (Lindsay et al. 2008; Schröder et al. 2014; Yuan et al. 2016; Stroeve et al. 2016; Petty et al. 2017; Ionita et al. 2019).

This chapter presents a statistical approach that exploits spatio-temporal variability within the sea ice record via construction of complex networks based on monthly mean fields of SIC from June, July and August. These networks are then utilised within a Gaussian Process Regression (GPR) model in order to predict both pan-Arctic and regional September monthly mean SIE anomalies, for all years between 1985 and 2019. In the context of real-time forecasts, this would correspond to forecasts being made on July 1st (June data), August 1st (July data), and September 1st (August data), with forecast validation made on October 1st. Hence forecast lead times extend from 1 to 3 months.

Section 2.3.3 has shown how complex networks provide a relatively new perspective with which to extract statistical relationships from the multitude of entangled interactions within the climate system, which were referred to as climatology tele-

connections. Such teleconnections have been studied for much of the last century (Walker & Bliss 1932; Wallace & Gutzler 1981; Glantz et al. 1991), and also in recent years for the similar purpose of Arctic sea ice prediction (Yuan et al. 2016; Comeau et al. 2019). GPR is then a natural complement to complex networks, whereby the construction of kernel functions for random graphs (networks) has been studied extensively (Kondor & Lafferty 2002; Smola & Kondor 2003; Vishwanathan et al. 2010; Spielman 2010; Urry & Sollich 2013), and provides the ability to integrate the network information directly into the regression model in the form of Gaussian prior and likelihood distributions (see section 3.3.2). To our knowledge, this is the first time this methodology has been implemented for the purpose of sea ice forecasting. This chapter is structured as follows: section 3.2 provides an overview of the data used for this study, section 3.3 describes the implementation of the complex networks and GPR methodology used for this study, results are presented in section 3.4, followed by a discussion (section 3.5) and conclusions (section 3.6).

3.2 Data

For the generation of the complex networks, monthly mean SIC fields between 1979 and 2018 were extracted from the National Snow and Ice Data Center (NSIDC) based on the NASA Team sea ice algorithm applied to passive microwave brightness temperatures (Cavalieri et al., 1996). Multiple satellites comprise this data record: Nimbus-7 SMMR (1979 – 1987), the DMSP F-8, F-11 and F-13 SSM/Is (1987 – 2007), and finally the DMSP F-18 SSM/I (2007 – 2017). To extend beyond 2018, these data are combined with the 2019 near-real-time daily SIC fields from NSIDC (Maslanik & Stroeve, 1999), which are averaged to generate the corresponding monthly fields. Data sets are provided on a polar stereographic 25×25 km grid, which are re-gridded here to a polar stereographic 100×100 km grid prior to the calculation of the grid-based pairwise correlations (Equation 2.47), for computational reasons. Grid cell area information (used to generate area-weighted time series, see section 3.3.1) was also extracted from NSIDC's pixel area tools library. Finally, it should be noted that the time varying polar hole in the SIC data is filled for the purpose of this study. This is achieved by simply filling the hole with the mean SIC value at 0.5° below the hole latitude, for each respective year.

To predict SIE on a regional level, the Arctic is separated into 9 geographic areas which surround the Arctic coastlines and encompass the most likely maritime navigation routes (Figure 3.1) – the data mask for these areas is also available from NSIDC (Fetterer et al., 2010). Here SIE is defined as the sum of grid cell areas for which SIC values are greater than or equal to 15%. A monthly mean value of extent is produced for each year by calculating the SIE for each day in September and then taking an average over all days.

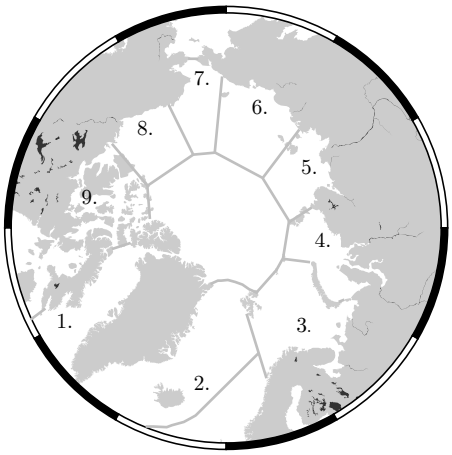


Figure 3.1: Regional Arctic areas used for September SIE forecasting (from Fetterer et al. 2010). 1: Baffin Bay, 2: Greenland Sea, 3: Barents Sea, 4: Kara Sea, 5: Laptev Sea, 6: East Siberian Sea, 7: Chukchi Sea, 8: Beaufort Sea, 9: Canadian Archipelago.

3.3 Methods

3.3.1 Complex networks

The process of generating complex networks from monthly mean SIC data (for June, July and August) follows the steps outlined in sections 2.3.2 and 2.3.3. The key points are briefly summarised again here. First, $\mathbf{x} = \{\boldsymbol{x}_p\}_{p=1}^{P}$ is defined as a linearly de-trended (zero-mean) time series data set representing n regularly sampled observations of sea ice concentration anomalies in time $\boldsymbol{x}_p = (x_{1p}, x_{2p}, \ldots, x_{np})$, at P fixed geographical locations, such that $\mathbf{x} \in \mathbb{R}^{P \times n}$. The data are then area-weighted $\mathbf{X} = \{\boldsymbol{x}_p \sqrt{\mathrm{w}_p}\}_{p=1}^{P}$, with w_p being the area in km^2 of grid cell p (data are on a polar stereographic grid), before subsequently computing pairwise correlations between all grid cells (Equation 2.47). Following this, the grid-based clustering algorithm outlined in section 2.3.2 is used to generate SIC network nodes, with a significance level of $\eta = 0.01$. The cumulative anomaly time series of each network node $\boldsymbol{g}_k$ then corresponds to the sum of all grid cells belonging to each cluster (Equation 2.49), and the links between nodes $\boldsymbol{w}$ are generated as the linear temporal covariance between the anomaly time series of each node (Equation 2.50). For a given network, the collection of all node time series is then referred to as $\boldsymbol{G} = \{\boldsymbol{g}_k\}_{k=1}^{N}$, such that $\boldsymbol{G}^{\mathrm{T}}$ is a $n \times N$ matrix.

As the SIC networks are used to make predictions of September SIE for all years between 1985 and 2019, it is worth commenting on how the network structure varies over time, as new networks are created for each forecast year. Figure 3.2 provides a visual illustration of this for August SIC data as an example. The network

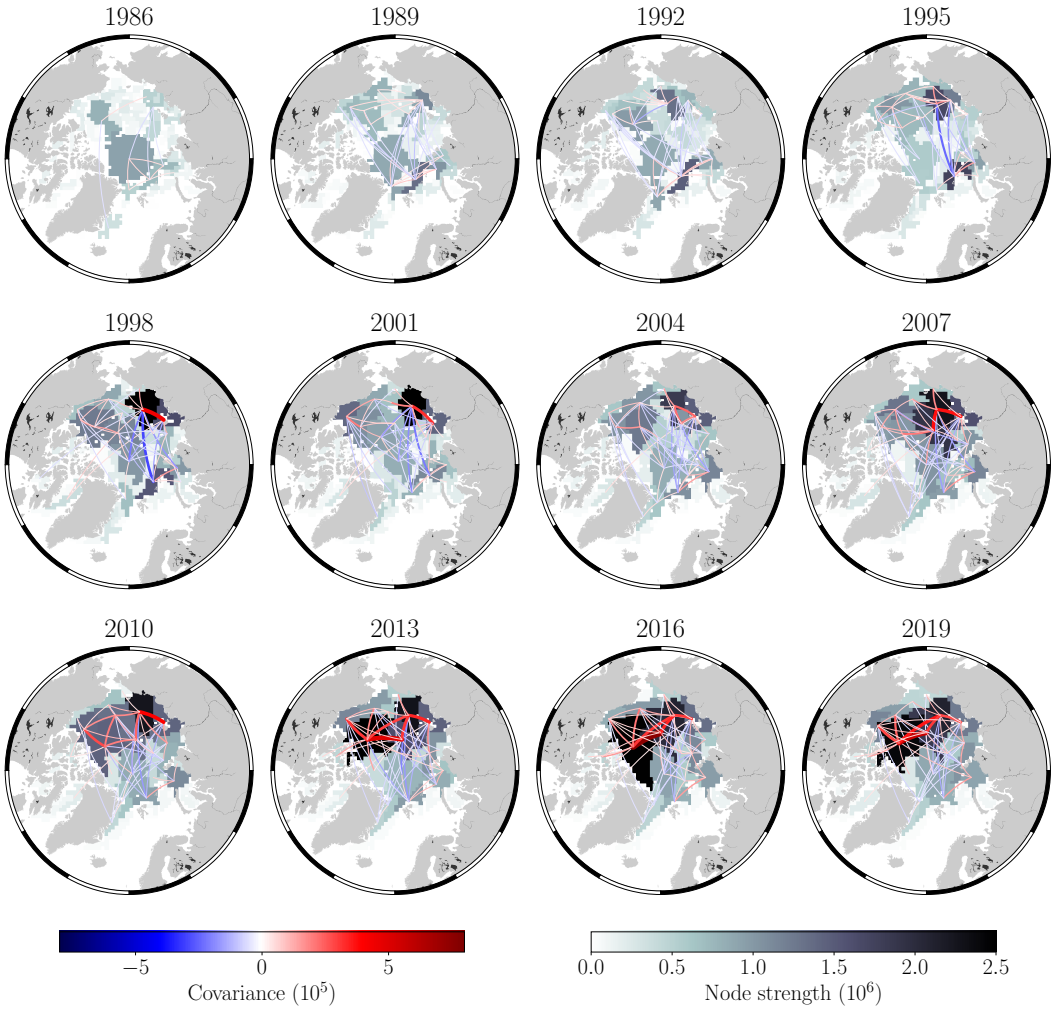


Figure 3.2: Network nodes and teleconnection links derived from linearly de-trended monthly mean August SIC data. The length of the time series used to create each network extends from 1979 to the years marked above each image. Each network area node $\mathcal{C}_k$ represents a clustering of grid cells. The thickness of each of the links here is proportional to the link weight w_{ij} (temporal covariance) between each pair of nodes. Only links which have a p-value < 0.1 are shown here to aid visualisation. See Equation (2.52) for the definition of node strength.

structure appears to vary significantly across the 34-year period shown here. Further investigation is required as to whether this variation is due to actual changes in sea ice behaviour or the fact that the length of the time series n used to construct each network is not constant. For example, the network labelled 1986 is constructed with $n = 8$, whereas the network labelled 2019 is constructed with $n = 41$ (time series begin at 1979). This does coincide with significant differences in the threshold factor τ, and subsequently the number of network nodes, as shown in Figure 3.3. Networks generated with shorter time series show a much larger value of τ which means that SIC grid cells will only cluster with neighbouring cells if the correlation between them is significantly large. As this is only likely to occur for perhaps the first few

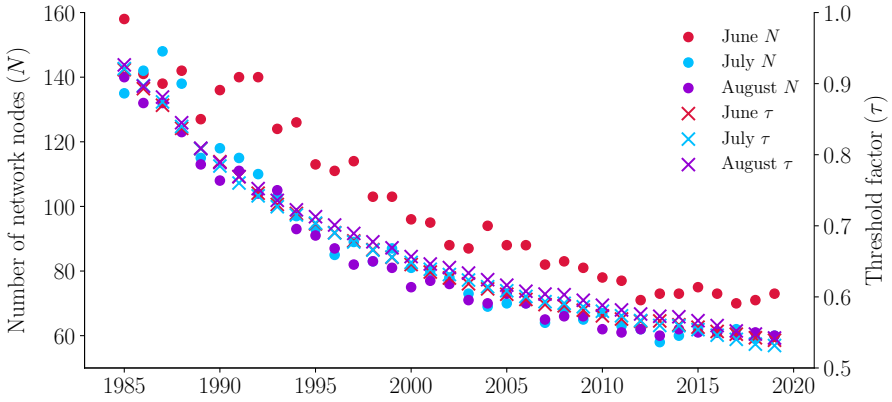


Figure 3.3: For networks based on linearly de-trended monthly mean SIC data from June, July and August, the variation in the number of network nodes N and also the threshold factor τ are shown here for all years between 1985 and 2019.

neighbouring cells, the resultant network hosts a very large number of nodes which contain a very small number of grid cells. Indeed it may be necessary to impose a pre-defined threshold of τ for years with shorter time series, although this goes beyond the scope of this study.

3.3.2 Gaussian process regression

Consider once again the supervised learning problem from section 2.2, which describes the functional relationship between inputs and outputs:

$$\mathbf{y} = f(\boldsymbol{G}) + \boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where $\boldsymbol{\varepsilon}$ represents independent and identically distributed Gaussian noise with mean 0 and variance σ^2. In the application here, $\mathbf{y}$ corresponds to n observations of September SIE anomalies and $\boldsymbol{G}$ the n observations of the cumulative SIC anomalies (Equation 2.49) from N network nodes. Note that the n observations used for training extend from (and include) 1979 to the year preceding the forecast year. For example if the forecast year is 2019 then $n = 40$, hence a new SIC network and GPR model are constructed for each forecast year. The aim of the model training is therefore to investigate the unknown relationship between the nodes of a given SIC network and the target September SIE. Here, this relationship is assumed to be linear, such that $f(\boldsymbol{G}) = \boldsymbol{G}^{\mathrm{T}}\boldsymbol{\beta}$, and subsequently the implementation becomes a Bayesian linear regression. As $f(\boldsymbol{G})$ is a Gaussian process (see section 2.2.1), it can therefore be completely defined by its mean (which is taken to be zero) and covariance function:

$$f(\boldsymbol{\Psi}) \sim \mathcal{GP}(0, k(\boldsymbol{\Psi}, \boldsymbol{\Psi})). \tag{3.2}$$

Here $\boldsymbol{\Psi}$ represents arbitrary function inputs, which correspond to either the training inputs $\boldsymbol{G}$, test inputs $\boldsymbol{G}_*$, or a combination of both. Test inputs are used to make predictions after the model training and are given here as the $n^{\text{th}} + 1$ entries that $\boldsymbol{G}$ would have if it were continued by another year, i.e., SIC anomalies of the year being forecast. A series of covariance functions can therefore be defined for all training and test inputs as:

$$\begin{aligned} \mathbf{K} &= k(\boldsymbol{G}, \boldsymbol{G}) = \boldsymbol{G}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{G} + \sigma^2 \boldsymbol{I} \\ \boldsymbol{k}_* &= k(\boldsymbol{G}_*, \boldsymbol{G}_*) = \boldsymbol{G}_*^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{G}_* + \sigma^2 \\ \mathbf{k}_* &= k(\boldsymbol{G}, \boldsymbol{G}_*) = \boldsymbol{G}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{G}_*, \end{aligned} \tag{3.3}$$

where $\boldsymbol{\Sigma}$ is the prior covariance matrix over the regression parameters $\boldsymbol{\beta}$ (see Equation 2.15), and is given here as a random walk matrix (Urry & Sollich 2013; see references therein) which takes the form $\boldsymbol{\Sigma} = \sigma_f^2 \exp(\ell \mathbf{M})$. $\mathbf{M}$ is then an adaptation of the adjacency matrix from Equation (2.51), which here corresponds to a symmetric transition matrix with off-diagonal elements containing the absolute teleconnection weights w_{kl} between nodes of the SIC network (Equation 2.50), and diagonal elements containing the negative strength of a given node:

$$\mathrm{M}_{kl} = |w_{kl}| \quad \text{for } k \neq l, \quad \mathrm{M}_{kk} = -\mathcal{S}_k = -\sum_{j \neq k}^{N} |w_{kl}|. \tag{3.4}$$

The remaining undefined variables $(\ell, \sigma_f^2, \sigma^2)$ are the hyperparameters of the prior covariance function. In this case, ℓ is known as the random walk 'time' parameter which controls the correlation between the regression parameters $\boldsymbol{\beta}$. As ℓ becomes large, the correlations between regression parameters (off-diagonal elements of $\boldsymbol{\Sigma}$) become large, and more long-ranged. At $\ell = 0$, the parameters are un-correlated, each having variance σ_f^2. These hyperparameters are optimised here using the empirical Bayes approach, which maximises the log marginal likelihood function from Equation (2.43), and is re-stated here for convenience:

$$\ln p(\mathbf{y}|\boldsymbol{G}) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{K}| - \frac{1}{2} \mathbf{y}^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{y}. \tag{3.5}$$

Following the example in section 2.2.2, it is possible to define $\tilde{\sigma}^2 = \sigma^2 / \sigma_f^2$ and work with the set of hyperparameters $\boldsymbol{\theta} = (\ell, \sigma_f^2, \tilde{\sigma}^2)^{\mathrm{T}}$. This way σ_f^2 can be optimised analytically to obtain $\sigma_f^2 = n^{-1} \mathbf{y}^{\mathrm{T}} \tilde{\mathbf{K}}^{-1} \mathbf{y}$, where $\tilde{\mathbf{K}} = \boldsymbol{G}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}} \boldsymbol{G} + \tilde{\sigma}^2 \boldsymbol{I}$, and $\tilde{\boldsymbol{\Sigma}} = \exp(\ell \mathbf{M})$ (Sollich, 2005), such that only ℓ and $\tilde{\sigma}^2$ need to be optimised numerically.
With the optimal hyperparameters found, the GPR model is then fully determined and forecasts of September SIE anomalies $\bar{f}_*$ and their corresponding variance $\sigma_{f_*}^2$ are given by:

$$\begin{aligned} \bar{f}_* &= \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{y} \\ \sigma_{f_*}^2 &= \boldsymbol{k}_* - \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{k}_*. \end{aligned} \tag{3.6}$$

3.4 Results

3.4.1 Pan-Arctic September sea ice extent forecasts

Forecasts of pan-Arctic September SIE anomalies are performed for all years between 1985 and 2019, based on monthly mean SIC fields from June, July and August. Each forecast is performed by GPR as explained above, with $\boldsymbol{G}$ representing linearly de-trended SIC data (for all network nodes) taken in the relevant month (June, July or August), and $\mathbf{y}$ the corresponding pan-Arctic September SIE. As explained in the previous section, for each forecast a new network of area nodes and hence a GPR model is constructed in order to ensure predictors capture the variability of the sea ice record over the period leading up to the year being forecast. Accordingly, optimal hyperparameters for the GPR are also determined separately for each forecast.

Before presenting the results, the definition of de-trended forecast skill $S_{\boldsymbol{f}_*}$ is given as the explained variance in the target variable $\mathbf{y}$ (e.g., observed de-trended pan-Arctic September SIE) relative to the de-trended forecast result $\boldsymbol{f}_* = (\bar{f}_{*1}, \bar{f}_{*2}, \ldots, \bar{f}_{*n})$ from Equation (3.6). Explicitly, $1 - S_{\boldsymbol{f}_*}$ is the ratio of the mean squared prediction error to the variance of these targets around their mean $\mathbb{E}[\mathbf{y}]$:

$$S_{\boldsymbol{f}_*} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{f}_{*i})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbb{E}[\mathbf{y}])^2}. \tag{3.7}$$

Furthermore, as comparisons will be made to other studies, the definition of the anomaly correlation coefficient (ACC) is also given as the linear cross-correlation coefficient between $\boldsymbol{f}$ and $\mathbf{y}$ from Equation (2.47), which is re-stated here:

$$\mathrm{ACC}_{\boldsymbol{f}_*\mathbf{y}} = \frac{\sum_{i=1}^{n}(\bar{f}_{*i} - \mathbb{E}[\boldsymbol{f}_*])(y_i - \mathbb{E}[\mathbf{y}])}{\sqrt{\sum_{i=1}^{n}(\bar{f}_{*i} - \mathbb{E}[\boldsymbol{f}_*])^2 \sum_{i=1}^{n}(y_i - \mathbb{E}[\mathbf{y}])^2}}. \tag{3.8}$$

Figure 3.4 displays the results of the pan-Arctic September SIE forecasts. The de-trended result shows remarkable skill of 0.53 at 3-month lead time using a single layer predictor (in the sense that GPR can be viewed as a single-layer SIC network), and increasing skill thereafter of 0.62 and 0.81 at 2 and 1 months respectively. Furthermore it is apparent from Figure 3.4 that each of the three forecasts appear to have performed better in the latter section of the time series (e.g., after approximately 1992). Therefore if the de-trended skill is calculated between 1993 and 2019, we notice an increase in skill for all lead times; from 0.53, 0.62, 0.81 to 0.58, 0.66, 0.88 respectively. The skill value of anomaly persistence between 1985 and 2019 is given here as 0.18, 0.52 and 0.88 at 3-, 2-, and 1-month lead times respectively, hence the GPR model out-performs persistence at 3 and 2 months, although at 1 month the skill

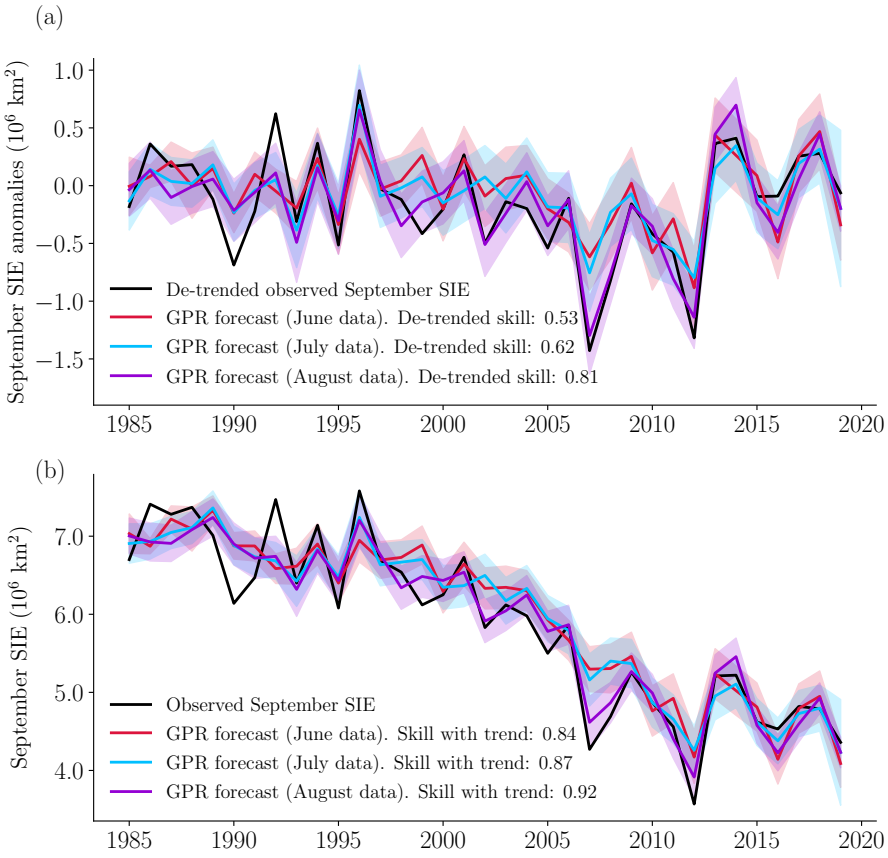


Figure 3.4: Predictions of pan-Arctic September SIE based on networks of monthly mean SIC data from June, July and August (3, 2 and 1 months lead time respectively). (a) De-trended forecast $\boldsymbol{f}_*$ showing predictive skills (Equation 3.7) of 0.53, 0.62, and 0.81 for forecasts based on June, July and August SIC networks respectively. (b) Forecasts with trend, with respective skills of 0.84, 0.87 and 0.92. The shaded areas in both plots represents the forecast standard deviation $\sigma_{\boldsymbol{f}_*}$.

of persistence is greater. Comparing these forecast results to recent statistical studies, Lindsay et al. (2008), Schröder et al. (2014), Petty et al. (2017) and Ionita et al. (2019) each presented skill scores for statistical forecasts of pan-Arctic September SIE. Skill values for each respective study were given as: [0.63, 0.81, 0.96 (with trend) at 3, 2, 1 months, respectively], [0.41 (de-trended) at 3-4 months], [0.45, 0.49, 0.59 (de-trended) at 3, 2, 1 months, respectively], and [0.78, 0.81 (de-trended) at 3, 2 months respectively]. Similarly, Yuan et al. (2016) reported de-trended ACC scores of [0.75, 0.82, 0.90 at 3, 2, 1 months respectively]; the de-trended ACC scores for the forecasts in this study are similar at 0.77, 0.83, 0.90 for the same respective lead times.

Looking to dynamical forecast comparisons, Sigmond et al. (2013), Wang et al. (2013), Msadek et al. (2014) and Bushuk et al. (2017) each presented de-trended ACC scores for various coupled model forecasts of pan-Arctic sea ice. Approximate results for each respective study were given as [0.55, 0.75, 0.90], [0.55, 0.65, 0.75], [0.60, 0.70,

0.81], and [0.50, 0.60, 0.70] for lead times of 3, 2, and 1 months respectively.
It should be noted that forecast skills presented above are not directly comparable as the period of analysis is not consistent between all studies. However given that the skill scores in this study out-perform the majority of the studies above, and that the analysis here is conducted over a longer time period should give an indication as to the exceptional performance of the combined complex network–GPR approach, for the presented lead times.

3.4.2 Regional September sea ice extent forecasts

Regional forecasts of September SIE are also performed to assess the model's performance beyond the single pan-Arctic metric and to be more in-line with the needs of a wider range of Arctic communities. As stated previously, 9 geographic regions have been chosen (Figure 3.1). These exclude the Bering Sea, Sea of Okhotsk, and Hudson Bay, as these regions are ice-free in September. In a similar vein, the Central Arctic has also been excluded as this is generally the location of the core ice pack, such that predictions there would be of limited use for many groups.
Similar to section 3.4.1, a new SIC network is constructed for each forecast year and the hyperparameters for the GPR are optimised separately for each forecast year and region.

Figure 3.5 displays the results of the regional September SIE forecasts. Across all lead times, predictions are generally most successful in regions towards the Pacific sector, with highest de-trended skill achieved in the Canadian Archipelago, Beaufort, Chukchi, East Siberian, and Laptev seas. Beyond 1-month lead time the GPR model consistently achieves poor prediction skill in Baffin Bay, and both Greenland and Barents seas, with Baffin Bay showing poor predictability for all lead times. Predictability appears to suffer in regions which have undergone significant changes in behaviour over the satellite record. For example, in the Kara Sea, the GPR model achieves very high de-trended skill between 1985 and 2006 (0.52, 0.59, 0.68 at 3-, 2-, and 1-month lead time respectively), however between 2007 and 2019 this skill drops significantly (-1.25, -0.44, and 0.35 for the same respective lead times), ultimately having a negative impact on the overall score. This change in predictability is likely due to a sudden change in the inter-annual variability of SIE in the Kara Sea after 2007 (see Figure 3.5b), such a change which would be extremely difficult for a statistical model to adapt to. The opposite can be said for other regions, in that the dominant window of predictability lies in the latter half of the time series record (i.e., after 2007). Regions such as the Canadian Archipelago, Beaufort, Chukchi, and East Siberian seas show significantly higher de-trended prediction skill between 2007 and 2019 than between 1985 and 2006 (Figure 3.6). See also Table 3.1 for a summary of

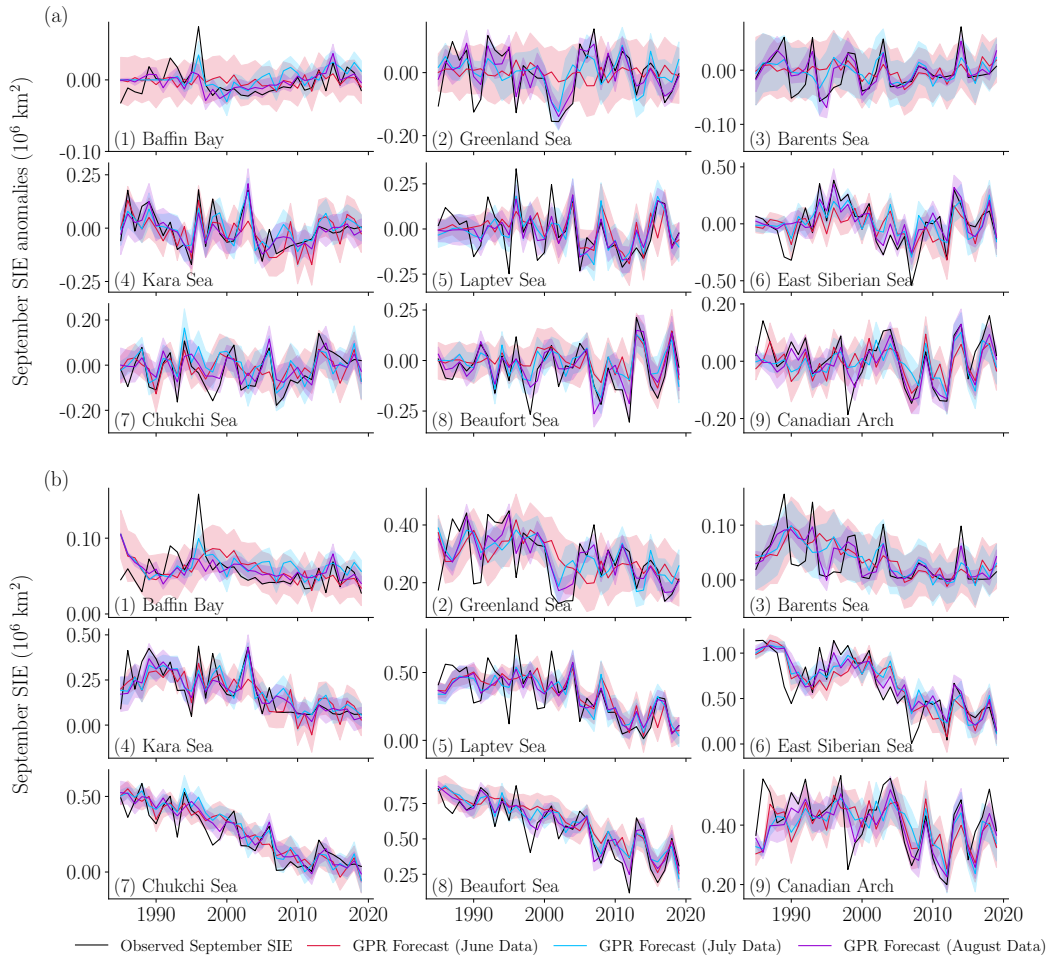


Figure 3.5: Predictions of regional September SIE based on networks of monthly mean SIC data from June, July and August. (a) De-trended forecast. (b) Forecasts with trend. The shaded areas in all plots represent the forecast standard deviation $\sigma_{\boldsymbol{f}_*}$.

the de-trended prediction skills for each region, and between different time windows. Looking again to comparison studies of regional sea ice forecasting using statistical models, Lindsay et al. (2008) made regional predictions of the September mean SIE at 3-month lead time. Reported skill values (with trend) were given as: 0.77, 0.74, 0.18, 0.08, and -0.67 for the Barents, Kara, Laptev, East Siberian and Beaufort seas respectively. In this study, skill values (also with trend) for the same respective regions at 3-month lead time are given as: 0.04, 0.54, 0.45, 0.66, and 0.65. Ionita et al. (2019) made similar predictions in the East Siberian Sea, with de-trended skill of 0.69, 0.78 at 3- and 2-month lead time respectively. Skills here are given as 0.48 and 0.47 at 3- and 2-month lead time respectively.

In terms of regional dynamical forecasts, Bushuk et al. (2017) reported de-trended ACC scores for all Arctic regions in a coupled model forecast. Here, other than the Greenland Sea at 3-month lead time, the equivalent de-trended ACC scores exceed theirs in all regions for each of the three respective lead times. Again, these skill

Skill 1985-2019 / 1985-2006 / 2007-2019 | **ACC** 1985 - 2019	**June data**	**July data**	**August data**
Pan-Arctic	0.53 / 0.32 / 0.65 | 0.77	0.62 / 0.43 / 0.73 | 0.83	0.81 / 0.63 / 0.91 | 0.90
Baffin Bay (1)	0.11 / 0.07 / 0.30 | 0.35	0.15 / 0.38 / -0.95 | 0.51	0.25 / 0.24 / 0.30 | 0.51
Greenland Sea (2)	-0.17 / -0.09 / -0.51 | -0.23	0.30 / 0.29 / 0.30 | 0.56	0.74 / 0.72 / 0.82 | 0.88
Barents Sea (3)	0.10 / 0.10 / 0.07 | 0.34	0.29 / 0.27 / 0.31 | 0.54	0.69 / 0.67 / 0.77 | 0.83
Kara Sea (4)	0.40 / 0.52 / -1.25 | 0.67	0.53 / 0.59 / -0.44 | 0.74	0.67 / 0.68 / 0.35 | 0.83
Laptev Sea (5)	0.25 / 0.22 / 0.16 | 0.52	0.53 / 0.46 / 0.62 | 0.73	0.62 / 0.52 / 0.84 | 0.80
East Siberian Sea (6)	0.48 / 0.34 / 0.64 | 0.70	0.47 / 0.48 / 0.44 | 0.70	0.63 / 0.62 / 0.63 | 0.82
Chukchi Sea (7)	0.26 / 0.12 / 0.46 | 0.54	0.31 / 0.18 / 0.48 | 0.65	0.56 / 0.53 / 0.59 | 0.79
Beaufort Sea (8)	0.34 / -0.07 / 0.57 | 0.71	0.51 / 0.36 / 0.59 | 0.77	0.71 / 0.63 / 0.76 | 0.86
Canadian Archipelago (9)	0.33 / 0.03 / 0.57 | 0.59	0.40 / 0.12 / 0.62 | 0.65	0.66 / 0.47 / 0.82 | 0.84

Table 3.1: De-trended regional forecast skill values for predictions of September SIE based on monthly mean SIC data from: June, July and August. Skill values (Equation 3.7) are calculated between three periods: 1985–2019, 1985–2006 and 2007–2019. De-trended ACC scores (Equation 3.8) are also given for the period 1985–2019. Labels 1 - 9 correspond to area labels in Figure 3.1.

metrics are not completely comparable due to different analysis periods.

3.5 Discussion

Several studies have linked sea ice persistence to high forecast skill of the pan-Arctic SIE metric with lead times ranging from 1 to 3 months (Drobot et al. 2006; Lindsay et al. 2008; Petty et al. 2017; Ionita et al. 2019). In this study, forecasts of pan-Arctic SIE out-perform the majority of both statistical and dynamical forecast models presented in the previous section despite the poor performance between 1985 and 1993, which is likely due to lack of observations in the GPR model training. Indeed heuristic 'rules-of-thumb' are available for deciding the minimum number of observations required to effectively model a given statistical problem (Bishop, 2006), although here we could pragmatically say that this minimum number should equate to the point after which the model is able to make accurate and precise future predictions. Nevertheless, with increasing demand for reliable forecasting of useful sea ice components, which is the focal point of the phase 2 of the Sea Ice Prediction Network (SIPN2), arguably regional forecasts of SIE and alternative metrics hold more importance than those of a single pan-Arctic extent. Therefore while skilful forecasts have been achieved for pan-Arctic SIE, this must be transferable to other regions to be useful for the various groups that are dependent on sea ice forecasts. Here, regional forecasts of SIE have proven to be relatively successful at 3-month lead time. Regions with high de-trended forecast skill (Canadian Archipelago, Beaufort, Chukchi, East Siberian, Laptev and Kara seas) coincide with areas which have been shown to contribute most to the decline in September SIE since 1979 (Onarheim et al., 2018). Similarly, regions with relatively poor de-trended forecast skill (Barents Sea, Greenland Sea and Baffin Bay) are those which show little variability in their September extents and in fact have been shown to contribute most to the decline

in March SIE since 1979 (Onarheim et al., 2018). In any case, future work must move to accurate predictions of spatio-temporal trends and derivation of metrics such as the location of the sea ice edge, and sea ice probability – in line with the SIPN2 mission statement. On this note, Increasing the present methodology to 4-month lead time (May SIC data) has been tested, however the resultant de-trended prediction skill is very poor for both the pan-Arctic and regional cases. This may be in part related to the spring predictability barrier which has been observed in model studies (Bonan et al. 2019; Bushuk et al. 2020), whereby May sea ice anomalies are unable to persist through to September (see also section 1.1.3). Irrespective of this, the methodology presented here can be advanced in order to achieve improved sea ice predictability. In terms of the GPR model the prior covariance function can be adapted such that we move away from a strictly linear system and instead allow for

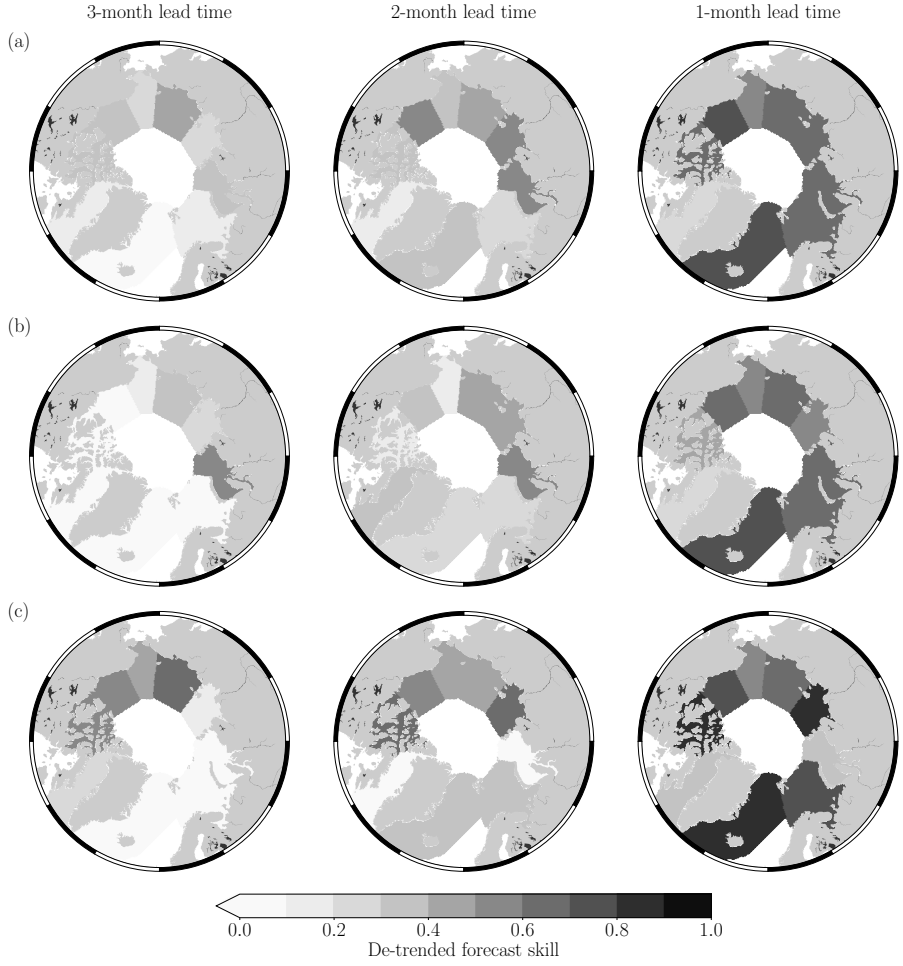


Figure 3.6: Skill by Arctic region for predictions of de-trended September SIE based on networks of monthly mean SIC data from June, July and August (3-, 2- and 1-month lead time respectively). (a) De-trended skill calculated between 1985 and 2019. (b) De-trended skill calculated between 1985 and 2006. (c) De-trended skill calculated between 2007 and 2019. See Table 3.1 for values.

fully non-parametric and non-linear fits (Sollich & Halees 2002; Girard et al. 2003; Rasmussen & Williams 2006; Sarkar et al. 2019). Furthermore one could provide an alternative Gaussian prior than the typical uniform zero-mean which is currently given to all network nodes. In the alternative case each node is assigned a different prior as necessary. In terms of the network framework, this also has the ability to grow in terms of complexity. Currently the implementation is rather simple with a single-layer network of SIC being used as a predictor. With the incorporation of multi-layer networks (Boccaletti et al., 2014) comes the ability to incorporate multiple climate variables within the regression forecasts. This may in turn lead to increasing the forecast lead time and skill (see section 6.1 for further discussion).

3.6 Conclusions

Forecast submissions to the SIO project have shown that statistical models still have the ability to be competitive with their dynamical model counterparts, within the inter-annual time frame (Stroeve et al. 2014c; Hamilton & Stroeve 2016). Although with apparent increasing temporal variability in the sea ice record (Holland et al., 2010) these statistical models need to be sufficiently complex in order to remain competitive. Not only this, the sea ice record shows significant spatial variability that can be seen across the observed SIE records (Figure 3.5). Complex networks are a method that exploits such spatio-temporal variability for purposes ranging from improving understanding of climatological teleconnections, to time series forecasting; as presented here. Forecasts of pan-Arctic September SIE have shown to be successful within a GPR forecast model and result in competitive de-trended skill values with the literature to date: 0.53, 0.62, and 0.81 at 3-, 2-, and 1-month lead time respectively. Regional forecasts made across the same time frame also show competitive de-trended skill in the Canadian Archipelago, Beaufort, Chukchi, East Siberian, Laptev and Kara seas, with skill values typically greater than 0.3. Improving forecast skill beyond 3 months requires development of the complex network and GPR methodology. The predictability of summer sea ice conditions using only concentration has been shown to diminish beyond 3 months (Drobot et al. 2006; Lindsay et al. 2008; Petty et al. 2017; Ionita et al. 2019). Extending the forecast window to 6 months or even to 1 year would require the incorporation of multiple climate variables (hence multi-layered networks) to capture sources of predictability such as ocean heat advection effects on summer sea ice conditions (Serreze & Stroeve, 2015) and possibly sea ice re-emergence patterns (Blanchard-Wrigglesworth et al. 2011; Bushuk et al. 2014), among many others.

Chapter 4

Data fusion and optimal interpolation of radar freeboards

The following chapter is based on the article "Gregory, W., Lawrence, I.R. and Tsamados, M., (2021). A Bayesian approach towards daily pan-Arctic sea ice freeboard estimates from combined CryoSat-2 and Sentinel-3 satellite observations. *The Cryosphere*, 15, 2857–2871." A supervised learning approach for deriving daily pan-Arctic estimates of radar freeboard is outlined, one which can be applied to any problem which seeks to enhance data coverage in the presence of sparse observations. Here, data sets from three separate satellite altimeters: CryoSat-2, Sentinel-3A, and Sentinel-3B, are combined in a Gaussian process regression framework to produce daily pan-Arctic estimates of radar freeboard, and uncertainty, between December 2018 and April 2019. The empirical Bayes approach is also used to learn the free hyperparameters of the model, which allows daily estimates of auxiliary products related to radar freeboard spatial and temporal correlation length scales to be derived. An analysis of the regional uplift in temporal resolution which is achieved by moving from the more generally reported monthly-averaged running means of radar freeboard, to that of a daily product is also shown. The chapter then ends by making time series comparisons of the daily radar freeboard anomalies with ERA5 reanalysis snowfall data in order to assess whether the estimated daily fields are able to capture real physical radar freeboard variability at sub-weekly time-scales.

4.1 Introduction

Estimates of Arctic sea ice thickness are crucial for a variety of purposes ranging from monitoring the volumetric response of sea ice to long-term climatic change (Slater et al., 2021), to understanding the consequences of climate change for Arctic

ecosystems, whose productivity rates are driven by the sea ice thickness distribution (Sakshaug et al. 1994; Stirling 1997; Stroeve et al. 2021). Furthermore, our ability to provide reliable forecasts of future sea ice conditions is inherently dependent on the availability of observations which allow us to exploit sources of sea ice predictability (Guemas et al., 2016). Initialising climate models with observations of sea ice thickness for example has been shown to considerably improve seasonal sea ice forecasts compared to those initialised with sea ice concentration (Chevallier & Salas-Mélia 2012; Doblas-Reyes et al. 2013; Day et al. 2014b; Collow et al. 2015; Bushuk et al. 2017; Allard et al. 2018; Blockley & Peterson 2018; Schröder et al. 2019; Ono et al. 2020; Balan-Sarojini et al. 2021).

In recent decades, advancements in satellite altimetry have enabled sea ice thickness to be estimated from space. This is achieved by measuring the sea ice freeboard; that is, the height of the sea ice surface relative to the adjacent ocean, and converting it to thickness by assuming hydrostatic equilibrium and bulk values of the ice, ocean, and overlying snow densities – and also snow depth (Laxon et al. 2003; Quartly et al. 2019). CryoSat-2 was the first radar altimeter launched with a specific focus on polar monitoring, and while it has been pivotal in improving our understanding of polar climate, its long repeat sub-cycle means that 30 days are required in order to generate pan-Arctic coverage (up to 88° N). The same is true for the ICESat-2 laser altimeter, launched in 2018, whose capability to estimate total (snow) freeboard, at monthly timescales, has been recently demonstrated (Kwok et al., 2019). Other radar altimeters in operation such as Sentinel-3A and Sentinel-3B have slightly shorter repeat cycles (27 days) however only extend to 81.5° latitude. With this in mind, the maximum temporal resolution that can be achieved for pan-Arctic freeboard from one satellite alone is 1 month. Increasing the resolution leads to a stepwise drop in spatial coverage until arriving at 1 day of observations for which less than 20% coverage is achieved for latitudes below 83° N, with tracks averaged to a 25 km^2 grid spacing (Lawrence et al. 2019; see also Tilling et al. 2016 for a similar analysis using 2 days of observations). This poses a significant limitation in our ability to both understand physical processes that occur on sub-monthly timescales, and in capturing the temporal variability of sea ice thickness, which could provide information on sea ice forecasts.

Recent studies have found ways to improve temporal coverage through the merging of different satellite products. Ricker et al. (2017) for example, merged thickness observations from CryoSat-2 and Soil Moisture and Ocean Salinity (SMOS) satellites to produce pan-Arctic thickness fields at weekly timescales. Furthermore, Lawrence et al. (2019) showed that radar freeboard observations from CryoSat-2, Sentinel-3A, and Sentinel-3B can be merged to produce pan-Arctic coverage below 81.5° N every 10 days. While both of these studies are significant improvements in temporal resolution, a daily pan-Arctic freeboard and/or thickness product does not yet exist. The

Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) sea ice thickness model (Zhang & Rothrock, 2003) is a commonly used substitute for observations when evaluating sea ice thickness in climate models, and is available at daily resolution. PIOMAS assimilates daily sea ice concentration and sea-surface temperature fields, and despite assimilating no information on sea ice thickness, it has been shown to be generally consistent with in situ and submarine observations (Schweiger et al., 2011), as well as exhibiting similar mean trends and annual cycles to both ICESat and CryoSat-2 satellites (Schweiger et al. 2011; Laxon et al. 2013; Schröder et al. 2019). However it generally over-estimates thin ice regions and under-estimates thick ice regions (Schweiger et al. 2011; Stroeve et al. 2014b), giving a mis-representation of the true ice thickness distribution.

This chapter exploits the consistency between CryoSat-2 (CS2), Sentinel-3A (S3A) and Sentinel-3B (S3B) radar freeboards (Lawrence et al., 2019) in order to produce a gridded pan-Arctic freeboard product, hereafter referred to as CS2S3, at daily resolution between December 2018 and April 2019. The method, Gaussian process regression (see chapter 2), aims to learn functional mappings between pairs of observation points in space and time by updating prior probabilities in the presence of new information - Bayes' theorem. Through this novel supervised learning approach, it is possible to move beyond the typical assessment of satellite-based radar freeboard variability with simple running means, and instead move closer towards understanding drivers of radar freeboard variability on sub-weekly time-scales with a daily product (see section 4.5). While many previous studies have reported statistical interpolation methods under a variety of names, e.g., Gaussian process regression (Paciorek & Schervish 2005; Rasmussen & Williams 2006), kriging (Cressie & Johannesson 2008; Kang et al. 2010; Kostopoulou 2021), objective analysis (Le Traon et al., 1997), and optimal interpolation (Ricker et al., 2017), each of these methods contain the same approach to learning functional mappings and the same key set of predictive equations. Despite this, the method is almost infinitely flexible in terms of its application and model setup. Section 4.3 for example will show how the model here differs from e.g., Ricker et al. (2017) through the unique choices of input-output pairs, as well as choice of prior over functions, and approach to learning model hyperparameters.
The chapter is structured as follows: section 4.2 introduces the data sets which are used within this study, section 4.3 outlines the specific implementation of Gaussian process regression unique to this study and presents an example of how pan-Arctic radar freeboard estimates can be achieved on any given day. Section 4.4 evaluates the interpolation performance through a comparison with the training inputs, and cross-validation experiments. Section 4.5 then provides an assessment of the improved temporal variability achieved by the use of a daily product, before finally ending with conclusions in section 4.6.

4.2 Data

The following section outlines the processing steps applied to CS2, S3A and S3B along-track data for generating the radar freeboard observations used as inputs to the Gaussian process regression model, as well as listing auxiliary data sets used.

4.2.1 Freeboard

Note that radar freeboard (the height of the radar scattering horizon above the local sea surface) is distinct from sea ice freeboard (the height of the snow-ice interface above the local sea surface). To convert radar freeboard to sea ice freeboard, a priori information on snow depth, density, and radar penetration depth are required. This study focuses solely on radar freeboard so as not to impose new sources of uncertainty. Radar freeboard can be considered as the 'base product', from which pan-Arctic daily estimates of sea ice freeboard and thickness can later be derived. For this study, along-track radar freeboard was derived for CS2 (synthetic aperture radar (SAR) and synthetic aperture radar interferometric (SARIN) modes), S3A and S3B in a two-stage process that is detailed in full in Lawrence et al. (2019). First, raw Level-0 (L0) data were processed to Level-1B (L1B) waveform data using ESA's Grid Processing On Demand (GPOD) SARvatore service (Dinardo et al., 2014). At the L0 to L1B processing stage, Hamming-weighting and zero-padding were applied, both of which have been shown to be essential for sea ice retrieval and which are not included in the ESA standard processing of S3 data at this time (Lawrence et al., 2019) – they are however applied during ESA's CS2 processing chain. Next, L1B waveforms were processed into radar freeboard following the methodology outlined in Lawrence et al. (2019) – based on that of Tilling et al. (2018) – and accounting for the Sentinel-3A and -3B (S3) retracking bias, as suggested in their conclusion. After subtracting the extra 1 cm retracking bias, 2018–2019 winter-average freeboards from CS2, S3A and S3B fall within 3 mm of one another (see Lawrence 2019). Notably, the standard deviation on the S3A(B)CS2 difference is comparable to the standard deviation on S3AS3B (σ=6.0(6.0)(5.9) cm for S3ACS2(S3BCS2)(S3AS3B)). Since S3A and S3B are identical in instrumentation and configuration, differing only in orbit, this suggests that any CS2S3A(B) radar freeboard differences are the result of noise and the fact that the satellites sample different sea ice floes along their different orbits, rather than due to biases relating to processing. Such consistency between data from individual satellites permits the combination of data from all three; thus in the following methodology, data from the three satellites are propagated as a single data set (CS2S3). It is also worth noting that while GPOD-derived CS2 radar freeboards are not compared with the ESA L2 baseline D product here, Lawrence et al. (2019) applied the same L1B $\rightarrow$ L2 processing to GPOD L1B and ESA L1B

(baseline C) data and found a radar freeboard difference of ~ 6 mm, which can be attributed to the fact that the GPOD L1B data do not contain the stack standard deviation parameter which is used for filtering lead and floe waveforms in the ESA L1B $\rightarrow$ L2 processing chain. This means that the product presented in this chapter can effectively be considered as version 1.0, while awaiting the availability of ESA Sentinel-3 Level-2 data which are processed to be consistent with CS2 (i.e., with Hamming weighting and zero padding applied).

4.2.2 Auxiliary data

In order to produce pan-Arctic estimates of radar freeboard on a given day, information on the maximum sea ice extent on that day is required. For this, daily sea ice concentration fields were extracted between December 2018 and April 2019 from the National Snow and Ice Data Center (NSIDC). These fields correspond to the NASA Team sea ice algorithm applied to passive microwave brightness temperatures from the Nimbus-7/SMMR (Scanning Multichannel Microwave Radiometer), DMSP (Defense Meteorological Satellite Program)/SSM/I (Special Sensor Microwave/Imager), and DMSP/SSMIS (Special Sensor Microwave Imager/Sounder), which are provided on a 25×25 km polar stereographic grid (Cavalieri et al., 1996). Grid cells containing sea ice were then selected as those with a concentration value $\geq 15\%$.
Section 4.3 makes distinctions between first-year ice (FYI) and multi-year ice (MYI) zones in order to perform computations related to the model setup. For this, the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) daily ice type product (OSI-403-c; Aaboe et al. 2016), derived from DMSP/SSMIS, Metop/ASCAT and GCOM-W/AMSR-2 satellites is used. These data are provided on a 10×10 km polar stereographic grid.
Finally, section 4.5 utilises the NSIDC-defined sea ice regions (Fetterer et al., 2010) in order to perform a regional analysis of the temporal variability of the daily CS2S3 product, as well as making comparisons with daily ERA5 snowfall reanalysis data (ERA5, 2017). Daily snowfall data were generated for each day between December 2018 and April 2019 by computing the sum of 24-hourly reanalysis fields.

4.3 Methods

This section revisits the method of Gaussian process regression, and outlines how it can be used to produce gridded pan-Arctic radar freeboard observations on any given day. In the example presented here, along-track freeboard observations from each satellite are averaged on a 25×25 km polar stereographic grid. NSIDC sea ice concentration, and OSI-SAF ice type data are also re-gridded to the same grid.

For a given day t, we have gridded freeboard observations from CS2, S3A and S3B satellites ($\mathbf{z}_1(t)$, $\mathbf{z}_2(t)$ and $\mathbf{z}_3(t)$ respectively). The number of observations from each satellite varies such that $\mathbf{z}_1(t) = \{z_1(t)_i\}_{i=1}^{n_1(t)}$, $\mathbf{z}_2(t) = \{z_2(t)_i\}_{i=1}^{n_2(t)}$, $\mathbf{z}_3(t) = \{z_3(t)_i\}_{i=1}^{n_3(t)}$ where some, but not all, of the observations between $\mathbf{z}_1(t)$, $\mathbf{z}_2(t)$ and $\mathbf{z}_3(t)$ are co-located in space. Let us then define $\mathbf{z}(t)$ as a $n(t) \times 1$ vector $\big(n(t) = n_1(t) + n_2(t) + n_3(t)\big)$ which is generated by concatenating the freeboard observations from each of the three satellites. Repeating this step for $\pm\ \tau$ consecutive days in a window around day t, means that $\mathrm{T} = 2\tau + 1$ number of $\mathbf{z}(t)$ vectors are combined in order to produce a single $n \times 1$ vector $\big(n = n(t-\tau) + \ldots + n(t) + \ldots + n(t+\tau)\big)$ of all observations $\mathbf{z}$. In this case τ is chosen to be 4, hence 9 days of observations are used in the model training. This results in the majority of grid cells having been sampled at least once over this period, thus reducing the prediction uncertainty at any given grid cell. Figure 4.1 shows how, in this example, the spatial coverage is improved from $\sim 23\%$ to $\sim 72\%$ by using 9 days of observations instead of 1. It is also worth noting that freeboard measurements from different satellites which are co-located in space and time are treated as separate observation points in the proposed workflow, where the average percentage of co-located points on any given day between CS2, S3A, and S3B; CS2 and S3A; CS2 and S3B; and S3A and S3B is $< 1\%$, $\sim 4\%$, $\sim 4\%$, and $\sim 8\%$ respectively. The aim is to then understand the function which maps the freeboard observations to their respective space-time positions in order to make predictions at unobserved locations on day t. This corresponds to a 3-D implementation of the supervised learning problem outlined in e.g., Equation (2.30), which is given here as:

$$\mathbf{z} = f(\mathbf{x}, \mathbf{y}, \mathbf{t}) + \boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2). \tag{4.1}$$

Once again, $\boldsymbol{\varepsilon}$ represents independent and identically distributed Gaussian noise with mean 0 and variance σ^2. The zonal and meridional grid positions of the freeboard observations are then given by the $n \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$ respectively, and finally $\mathbf{t}$ is a $n \times 1$ vector which contains the time index of each observation point $\mathbf{t} = \big(\{(t-\tau)_i\}_{i=1}^{n(t-\tau)}\ ,\ \ldots\ ,\ \{(t)_i\}_{i=1}^{n(t)}\ ,\ \ldots\ ,\ \{(t+\tau)_i\}_{i=1}^{n(t+\tau)}\big)$. For convenience, the collective training inputs are also defined as $\boldsymbol{\Phi} = (\mathbf{x}, \mathbf{y}, \mathbf{t})$, such that $f(\mathbf{x}, \mathbf{y}, \mathbf{t}) \equiv f(\boldsymbol{\Phi})$.

Section 2.2.1 showed how Gaussian Process Regression (GPR) enables the function f to be learned from the training set of inputs $\boldsymbol{\Phi}$ and outputs $\mathbf{z}$, and subsequently enables predictions to be made for a new set of test inputs $\boldsymbol{\Phi}_* = (x_*, y_*, t)$. Here the test inputs correspond to the zonal and meridional grid positions of where the predictions will be made. For now, let us assume that this corresponds to one grid cell which contains sea ice on day t (i.e., a grid cell within the grey mask shown in Figure 4.1). Recall from Equation (2.31) that by assuming f is a Gaussian Process

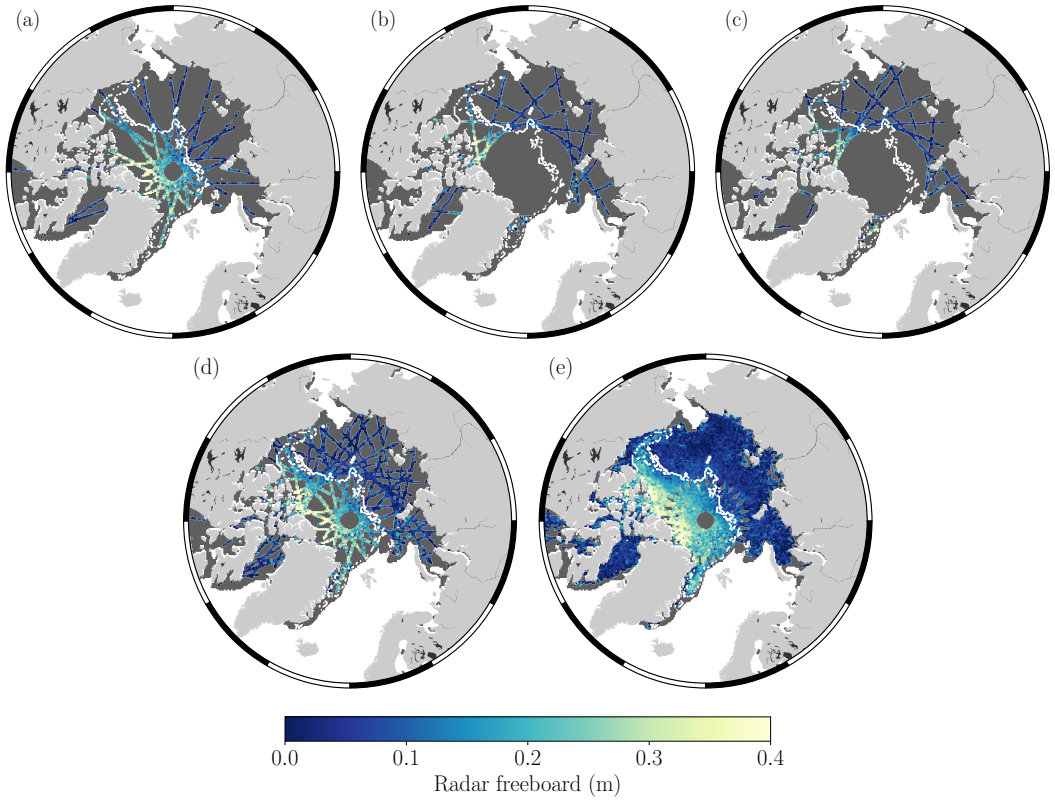


Figure 4.1: Gridded tracks from (a) CS2, (b) S3A, and (c) S3B on a 25 × 25 km polar stereographic grid, respectively covering approximately 11%, 7% and 7% of the total NSIDC NASA Team sea ice extent (grey mask) on day $t = 1^{\text{st}}$ of December 2018. By combining the three satellites (d), this coverage is increased to approximately 23%. Combining $t \pm 4$ days of gridded tracks (e), the coverage is increased further to approximately 72% of the total sea ice extent. The white ice type contour (from OSI-SAF) shows the boundary between thick MYI and thin FYI, on day t.

($\mathcal{GP}$), a particular class of mean and covariance function must be chosen in order to specify which type of functions to give preference to:

$$f(\mathbf{\Psi}) \sim \mathcal{GP}(m(\mathbf{\Psi}), k(\mathbf{\Psi}, \mathbf{\Psi}')),$$

where $\mathbf{\Psi}$ represents arbitrary function inputs (either the training $\mathbf{\Phi}$ or test inputs $\mathbf{\Phi}_*$). In this case, a constant value is assigned for the prior mean $m(\mathbf{\Psi})$, which is given as the mean of CS2 FYI freeboards from the 9 days prior to the first day of training data, i.e., from $(t - \tau - 9)$ to $(t - \tau - 1)$. The reason that only FYI freeboards are used is that with fewer observation points, GPR has less evidence to support significantly different freeboards from $m(\mathbf{\Psi})$. From Figure 4.1e we can see that there are generally fewer observation points in the (FYI) coastal margins, hence predictions here will likely remain close to $m(\mathbf{\Psi})$. In the MYI zone however, there are a large number of observations (more evidence) to support freeboards which may be different from $m(\mathbf{\Psi})$, so the choice of $m(\mathbf{\Psi})$ will likely have little effect on the

prediction values here. The prior covariance $k(\boldsymbol{\Psi}, \boldsymbol{\Psi}')$ is given here as the anisotropic Matérn covariance function:

$$k(\boldsymbol{\Psi}, \boldsymbol{\Psi}') = \sigma_f^2\big(1 + \sqrt{3}d(\boldsymbol{\Psi}, \boldsymbol{\Psi}')\big) \exp\big(-\sqrt{3}d(\boldsymbol{\Psi}, \boldsymbol{\Psi}')\big), \tag{4.2}$$

with Euclidean distance $d(\boldsymbol{\Psi}, \boldsymbol{\Psi}')$ in space and time given by:

$$d(\boldsymbol{\Psi}, \boldsymbol{\Psi}') = \frac{\|\boldsymbol{\psi}_x - \boldsymbol{\psi}'_x\|}{\ell_x} + \frac{\|\boldsymbol{\psi}_y - \boldsymbol{\psi}'_y\|}{\ell_y} + \frac{\|\boldsymbol{\psi}_t - \boldsymbol{\psi}'_t\|}{\ell_t}.$$

In the above definition, σ_f^2, ℓ_x, ℓ_y, and ℓ_t (and also σ^2 from Equation 4.1) are the model hyperparameters, each taking a real value > 0. Similar to the example outlined in section 2.2.2, σ_f^2 controls the overall variance of the function values, while ℓ_x, ℓ_y and ℓ_t are the correlation length scales (one for each dimension), describing how far in the input space (metres or days) the observations are expected to remain correlated. Let us now refer to $\boldsymbol{\theta} = (\sigma_f^2, \ell_x, \ell_y, \ell_t, \sigma^2)^{\mathrm{T}}$ as the collection of all hyperparameters. Here the empirical Bayes approach (see section 2.2.2) is used to select values of $\boldsymbol{\theta}$ which maximise the log marginal likelihood function:

$$\ln p(\mathbf{z}|\boldsymbol{\Phi}, \boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{K}| - \frac{1}{2}(\mathbf{z} - m(\boldsymbol{\Phi}))^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{z} - m(\boldsymbol{\Phi})), \tag{4.3}$$

where

$$\mathbf{K} = k(\boldsymbol{\Phi}, \boldsymbol{\Phi}) + \sigma^2\boldsymbol{I}.$$

It is also worth noting that as well as optimising the model hyperparameters, the choice of prior covariance function was established by evaluating Equation (4.3) on a variety of different covariance functions, where the Matérn was found to be favourable, producing the highest log marginal likelihood.

With the model fully determined, predictions of radar freeboard can then be generated at the test locations $\boldsymbol{\Phi}_*$. As outlined in Equation (2.39), this corresponds to evaluating the mean $\bar{f}_*$ and variance $\sigma_{f_*}^2$ of the posterior predictive distribution:

$$\begin{aligned} \bar{f}_* &= m(\boldsymbol{\Phi}_*) + k(\boldsymbol{\Phi}, \boldsymbol{\Phi}_*)\mathbf{K}^{-1}(\mathbf{z} - m(\boldsymbol{\Phi})) \\ \sigma_{f_*}^2 &= k(\boldsymbol{\Phi}_*, \boldsymbol{\Phi}_*) - k(\boldsymbol{\Phi}, \boldsymbol{\Phi}_*)\mathbf{K}^{-1}k(\boldsymbol{\Phi}_*, \boldsymbol{\Phi}). \end{aligned} \tag{4.4}$$

Now that the framework is in place to generate a predictive distribution of radar freeboard values for a given set of training and test locations, let us explore how this is implemented in practice. Due to the need to invert a matrix of size $n \times n$ (i.e., $\mathbf{K}$), GPR has run time complexity $\mathcal{O}(n^3)$, which means that if the number of training points is doubled, then the run time is increased by a factor of 8. As $\mathbf{K}$ must be inverted at every iteration step when optimising the model hyperparameters,

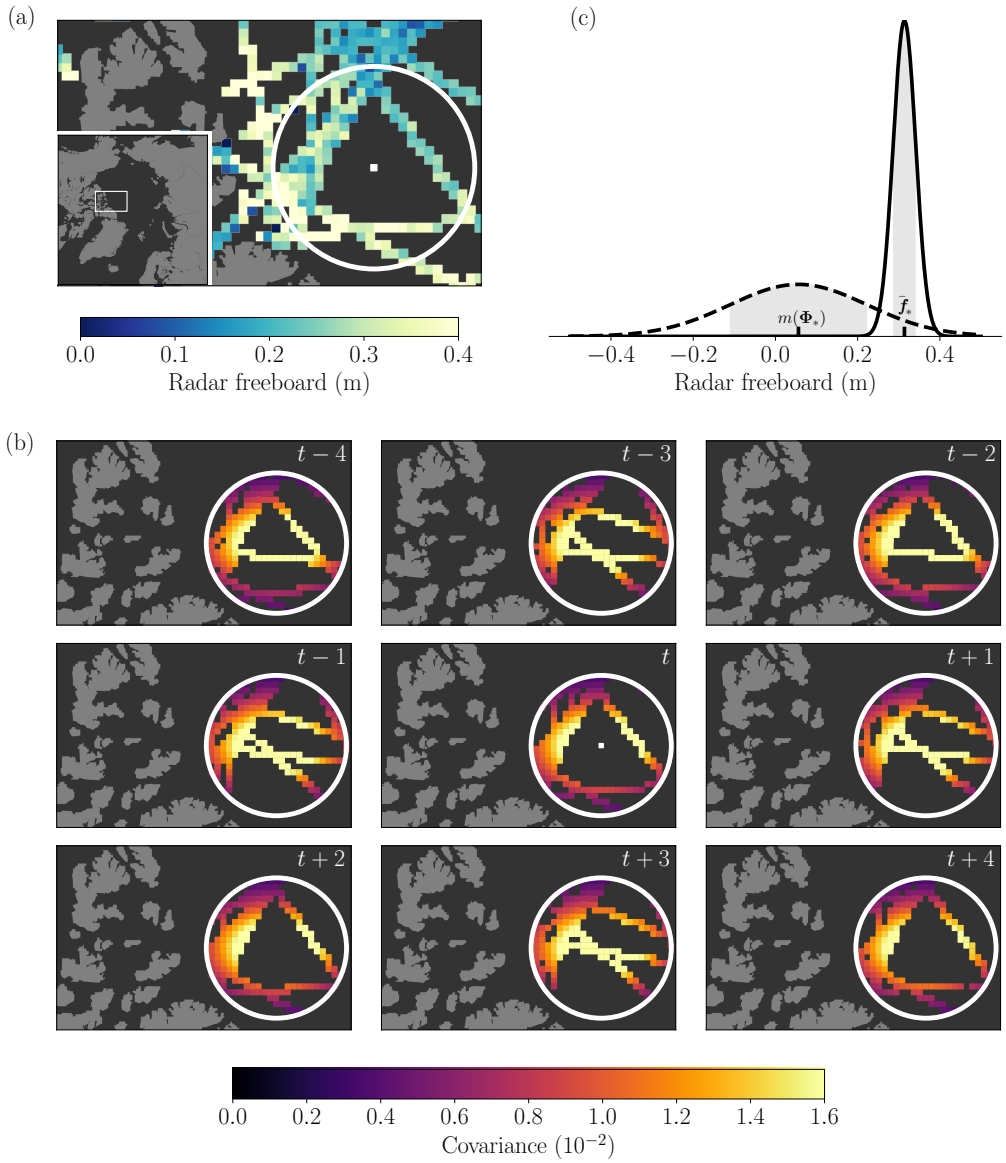


Figure 4.2: Estimating freeboard for one grid cell. (a) A posterior distribution of freeboard values is to be estimated at the location of the white pixel on day t. The white circle corresponds to a distance of 300 km from the pixel, and the gridded tracks are the CS2 and S3 observations on day t. (b) The prior covariance $k(\mathbf{\Phi}_*, \mathbf{\Phi})$, which is the covariance between the test input $\mathbf{\Phi}_*$ (the white pixel), and all the training inputs $\mathbf{\Phi}$ that lie within a 300 km radius, for each of the 9 days of training data. (c) Probability density functions showing the prior distribution of function values (dashed), with mean $m(\mathbf{\Phi}_*) = 0.056$ m and 1 standard deviation $\sqrt{k(\mathbf{\Phi}_*, \mathbf{\Phi}_*)} = 0.167$ m, as well as the posterior predictive distribution (solid) showing the estimated freeboard value at the location of the white pixel on day t as $\bar{f}_* = 0.314$ m, with 1 standard deviation of $\sigma_{f_*} = 0.028$ m.

GPR becomes increasingly computationally expensive with increasing n. For this reason, an iterative approach to the predictions is taken here: for a particular day t, predictions of radar freeboard are generated at each grid cell by using only the available training data that exist within a 300 km radius (see Figure 4.2). The

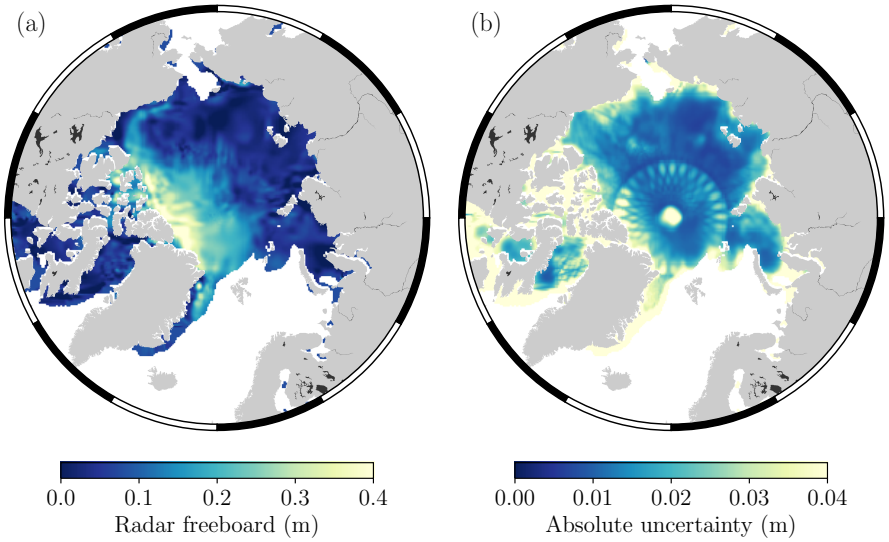


Figure 4.3: (a) Gridded CS2S3 radar freeboard from Gaussian Process Regression. (b) The absolute uncertainty (1 standard deviation), corresponding to the square root of the predictive variance from Equation (4.4). Both images correspond to estimates on the 1st of December 2018, at a grid resolution of 25 $\times$ 25 km.

process shown in Figure 4.2 is essentially repeated for every grid cell which contains sea ice, until producing a pan-Arctic field on day t. While this does effectively mean that observations beyond 300 km distance are considered to be uncorrelated, it does have the advantage of both computational efficiency, and it allows spatial variation to be freely incorporated into the length scales ℓ_x, ℓ_y, ℓ_t (and hence spatial variation into the smoothness of the function values). This seems sensible given the different scales of surface roughness that exist between FYI and MYI (Nolin et al., 2002). Furthermore, 300 km can be considered a reasonable distance in estimating the spatial covariance between inputs given that freeboard observations are correlated up to a distance of at least 200 km due to along-track interpolation of sea level anomalies (Tilling et al. 2018; Lawrence et al. 2018).

Figure 4.3 shows the result of repeating the steps in Figure 4.2 for every grid cell which contains sea ice on day t. The freeboard values here correspond to the mean of the posterior distribution $\bar{f}_*$, and the uncertainty as the square root of the variance term σ_{f_*}. Notice that the CS2S3 freeboard appears smoother than if we were to simply average the training observation points (e.g., Figure 4.1e). This is because GPR estimates the function values f, while it was stated in Equation (4.1) that the observations $\mathbf{z}$ are the function values corrupted by random Gaussian noise. We also notice how the uncertainty in the estimation of the freeboard values increases in locations where there are fewer training data, and is typically highest in areas where there are no data at all (e.g., the polar hole above 88° N). Section 4.5 provides further discussion relating to the predictive uncertainty of the CS2S3 field, looking particularly at the Canadian Archipelago and the Greenland, Iceland and Norwegian

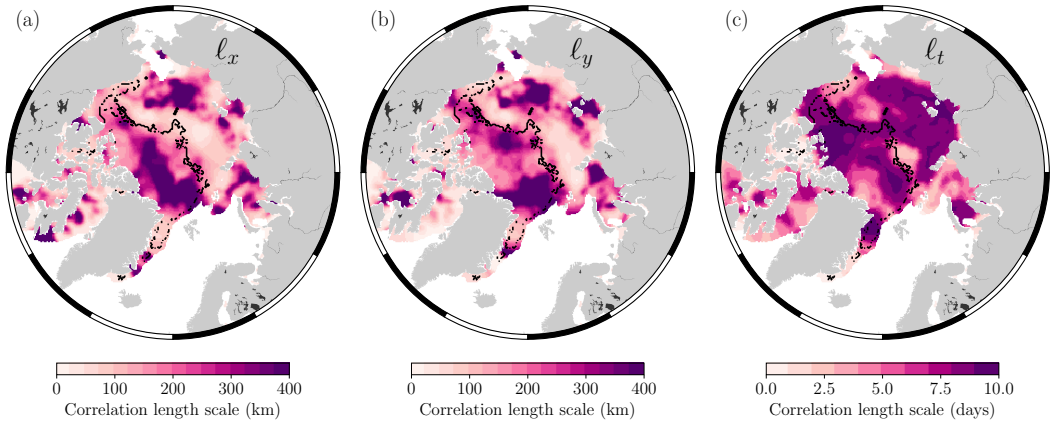


Figure 4.4: The 25×25 km gridded (a) zonal, (b) meridional, and (c) temporal correlation length scales which maximise the log marginal likelihood function (Equation 4.3) for the 1st of December 2018. The black contour also corresponds to the FYI/MYI contour for the 1st of December 2018.

seas, where uncertainties are consistently higher than other regions of the Arctic. See also Figures B.1 and B.2 for sensitivity tests, which shows how the predictions and uncertainty estimates in Figure 4.3 are affected by varying the number of days of observations used to train the model.

As a final point related to the methodology, given that the predictions are generated iteratively, it is possible to construct daily spatial maps of each of the model hyperparameters that maximise the log marginal likelihood function. Of particular importance are perhaps the correlation length scale parameters, as these could be of use in other applications such as localisation techniques in sea ice data assimilation systems (Sakov & Bertino 2011; Massonnet et al. 2015; Zhang et al. 2018, 2021). Previous studies have estimated the average spatial length scales of sea ice thickness anomalies across a range of coupled climate models (Blanchard-Wrigglesworth & Bitz, 2014) and reanalysis products (Ponsoni et al., 2019) to be typically in the range of 500–1000 km. In this approach, the maximum spatial length scales of radar freeboard are capped at 600 km as this is the maximum distance between any pair of grid cells used during the model training (similarly the temporal length scales are capped at 9 days). Nevertheless, from Figure 4.4 we notice some interesting features. Specifically, that larger spatial length scales typically coincide with the thicker MYI zone, as well as some localised features around the East Siberian–Chukchi seas, and the Kara and Laptev seas. We also notice some presence of anisotropy (i.e., at a given pixel, $\ell_x \neq \ell_y$), particularly in the area north of Greenland, and some regions of the Beaufort Sea. This could be explained by the presence of ridges and other deformation features which form in a preferential orientation relative to the apparent stress regime (Kwok 2015; Petty et al. 2016b). Some of the lowest correlation length scales occur in the peripheral seas, which is perhaps related to more dynamic ice conditions and uncertainty related to spatial sampling in these regions (see section

Date	μ (m) CS2-CS2S3	σ (m) CS2-CS2S3	μ (m) S3A-CS2S3	σ (m) S3A-CS2S3	μ (m) S3B-CS2S3	σ (m) S3B-CS2S3	RMSE (m) CS2-CS2S3	RMSE (m) S3A-CS2S3	RMSE (m) S3B-CS2S3
201812	0.001	0.051	0.000	0.057	-0.001	0.057	0.051	0.057	0.057
201901	0.001	0.049	0.001	0.056	-0.002	0.055	0.049	0.056	0.055
201902	0.000	0.050	0.000	0.055	-0.001	0.055	0.050	0.055	0.055
201903	0.001	0.050	0.000	0.056	-0.001	0.057	0.050	0.056	0.047
201904	0.001	0.053	0.000	0.061	-0.001	0.061	0.053	0.061	0.061
all months	0.001	0.051	0.000	0.057	-0.001	0.057	0.051	0.057	0.057

Table 4.1: Mean (μ), standard deviation (σ), and RMSE of the daily difference between gridded CS2, S3A and S3B tracks and co-located CS2S3 points, for all days between the 1st of December 2018 and the 24th of April 2019. Also computed for all days in each respective month.

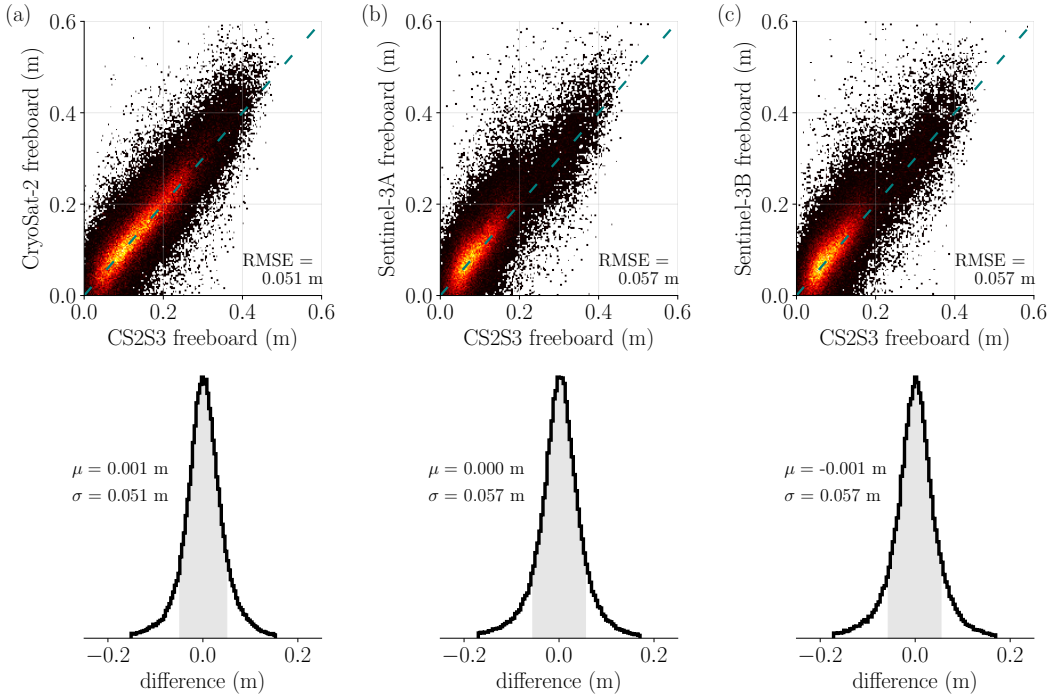


Figure 4.5: Comparison of daily gridded CS2S3 freeboard against daily (a) CS2, (b) S3A, and (c) S3B gridded freeboard tracks, for all days between the 1st of December 2018 and the 24th of April 2019. In each case, the freeboard values are directly compared in the scatter-density plot (dashed line shows $y = x$). The distribution of the error is given by the histograms below each scatter plot (1σ either side of the mean μ is shaded in grey). Only values within $\pm 3\sigma$ of the mean are shown for the histogram plots.

4.5 for further discussion). With regards to temporal length scales, we see that radar freeboard observations are correlated over much of the Arctic over the 9 day training period, except for the Canadian Archipelago and some areas of the Greenland, Iceland and Norwegian Seas (also discussed further in section 4.5).

4.4 Validation

In this section various metrics are used to assess the GPR model in terms of training and prediction. First, CS2S3 daily freeboards are compared against the

training inputs in order to derive average errors across the 2018–2019 winter season, before evaluating the predictive performance of the model through cross-validation experiments. Note that hereafter, all CS2S3 fields are generated at 50×50 km resolution, and CS2 and S3 along-track data are averaged to the same 50×50 km polar stereographic grid, for computational efficiency.

4.4.1 Comparison with training inputs

Here the performance of the GPR model is assessed during training; that is, an assessment of the errors between the gridded CS2S3 field, and CS2 and S3 gridded tracks, for all days between the 1$^{\text{st}}$ of December 2018 and the 24$^{\text{th}}$ of April 2019. In general, caution should be applied when evaluating training performance, as a model which fits the training data well may be fitting to the noise in the observations (over-fitting), leading to poor predictive performance. Meanwhile, a model which under-fits the training data is equally undesired (recall from section 2.2.1 however that over-fitting is inherently mitigated in the GPR method). In the model outlined in Equation (4.1), the freeboard observations from CS2 and S3 corresponded to the function values plus random Gaussian noise. Therefore by comparing the CS2S3 freeboard values from Equation (4.4) to the training data, the difference should correspond to zero-mean Gaussian noise. In Figure 4.5 differences are computed (CS2-CS2S3, S3A-CS2S3, S3B-CS2S3) for all days, where we can see that the difference (error) follows a normal distribution, centred approximately on 0. The average daily difference μ between CS2S3 and CS2 (or S3) is ≤ 1 mm, with standard deviation on the difference $\sigma < 6$ cm (see also Figure B.3 where training errors are compared for interpolations run at different spatial grid resolutions for 1 day). Furthermore, we can see that the Root Mean Square Error (RMSE) between CS2S3 and the training inputs is equivalent to the standard deviation of the difference, which can only occur when the average bias is approximately 0. Notably, each of the standard deviations is approximately equal to the ~ 6 cm uncertainty on 50 km grid-averaged freeboard measurements, determined from a comparison of S3A and S3B data during tandem phase of operation (see Figure B.4). A breakdown of average errors for each month are given in Table 4.1, however it should be noted that rounding is likely to play a role in many of the presented statistics. For example, we can see that the CS2-CS2S3 mean difference for the 'all months' case is reported as 0.001 m, and similarly for S3A-CS2S3 we have 0.000 m. If these values are expanded, we arrive at 0.00078 m and 0.00024 m respectively, hence the apparent difference between means CS2-CS2S3 and S3A-CS2S3 is almost halved.

4.4.2 Cross-validation

The previous section showed that CS2S3 freeboards closely resemble CS2 and S3 observations (the training data), however this does not give an indication as to how well the model performs in predicting at unobserved locations. An independent set of radar freeboard observations would allow CS2S3 freeboards to be validated at locations unobserved by CS2 or S3 on any given day, however in the absence of these, alternative metrics must be used to evaluate the predictions. Commonly, data from the airborne mission Operation IceBridge are used to validate sea ice freeboard and/or thickness observations, however as only 6 days were collected in a small area around north of Greenland in 2019, there are insufficient data points here to realistically draw any conclusions about the CS2S3 freeboards on larger temporal and spatial scales. K-fold cross-validation (see appendix A.1) is a useful tool when validation data are limited, however models must be run K number of times (ideally where $K = n$), incurring significant computational expense when n is large (as it is in this case). A pragmatic solution is opted for here: the predictive performance of the model is tested by removing different combinations of each of the S3 satellites from the training data, re-generating the predictions with the remaining subset, and subsequently evaluating predictions on the withheld set. For example, in the first instance daily pan-Arctic predictions are generated across the 2018–2019 period, except that both S3A and S3B are withheld from the training set. Predictions are then evaluated from each day against observations from S3A and S3B. In the next instance, the same process is repeated except that only S3A are withheld from the training set (whilst retaining S3B), and S3A are used as the validation set. In the final scenario, S3B are withheld from the training set and used as the validation set. It should be noted however, that this approach only allows predictions to be validated below 81.5° N, and at locations where sea ice concentration values are $\geq$ 75%. This is because 81.5° N is the limit of spatial coverage of S3 satellites, and during the processing of both CS2 and S3 along-track data, diffuse waveforms which sample grid cells with lower than 75% concentration are discarded (Lawrence et al., 2019). Figure 4.6 compares predictions for each of the previously mentioned scenarios, across all days in the 2018–2019 winter period. As expected, the mean and spread of error in the predictions is largest in the case where only CS2 is used to train the model, with $\mu = -0.002$ m, $\sigma = 0.074$ m and $\mu = -0.004$ m, $\sigma = 0.073$ m relative to S3A and S3B observations respectively. The mean and spread of error then decreases slightly with the incorporation of either S3A or S3B, with $\mu = -0.001$ m, $\sigma = 0.072$ m and $\mu = -0.003$ m, $\sigma = 0.072$ m respectively. The mean error across all validation tests is consistently ≤ 4 mm, showing that the model is able to make reliable predictions at unobserved locations. It therefore seems reasonable to expect that with the inclusion of all 3 satellites, the predictions at unobserved locations are

Date	μ (m) S3A-CS2S3(-S3)	σ (m) S3A-CS2S3(-S3)	μ (m) S3B-CS2S3(-S3)	σ (m) S3B-CS2S3(-S3)	μ (m) S3A-CS2S3(-S3A)	σ (m) S3A-CS2S3(-S3A)	μ (m) S3B-CS2S3(-S3B)	σ (m) S3B-CS2S3(-S3B)
201812	-0.002	0.073	-0.004	0.072	0.001	0.072	-0.002	0.072
201901	-0.001	0.071	-0.004	0.071	0.002	0.070	-0.003	0.070
201902	-0.002	0.072	-0.003	0.071	0.000	0.071	-0.002	0.070
201903	-0.003	0.074	-0.005	0.075	0.000	0.072	-0.004	0.073
201904	-0.002	0.079	-0.005	0.076	0.001	0.076	-0.003	0.076
all months	-0.002	0.074	-0.004	0.073	0.001	0.072	-0.003	0.072

Table 4.2: Mean (μ) and standard deviation (σ) of the daily difference between gridded S3A and S3B tracks and co-located CS2S3 points, with different combinations of S3 observations removed from the training data.

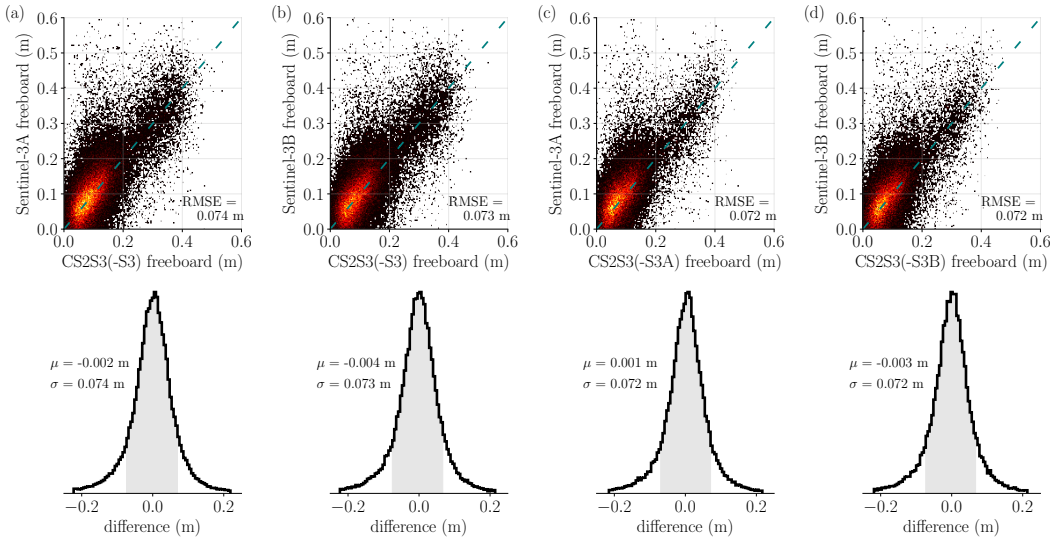


Figure 4.6: Scatter-density plots and error distributions of the GPR predictions for all days between the 1st of December 2018 and the 24th of April 2019, with different combinations of S3 satellites removed from the training data. (a) Model trained using only CS2 observations, validated against S3A. (b) Model trained using only CS2 observations, validated against S3B. (c) Model trained using only CS2 and S3B observations, validated against S3A. (d) Model trained using only CS2 and S3A obsevations, validated against S3B. Grey shading is used to indicate 1σ either side of the mean μ for each histogram. Only values within $\pm 3\sigma$ of the mean are shown for the histogram plots.

at least as good as ≤ 4 mm error, if not better. A breakdown of equivalent statistics is presented for each month in Table 4.2, although RMSE is not included here as it is again consistent with the standard deviation in each case. See also Figures B.5 and B.6 which show the uplift in the actual predictions and uncertainty estimates brought by the inclusion of all three satellites, as opposed to any one of the cross-validation experiments outlined above.

4.5 Assessment of temporal variability

Finally, to showcase the capabilities of the CS2S3 product, this section turns to an assessment of temporal variability. Within 9 sectors of the Arctic the mean CS2S3

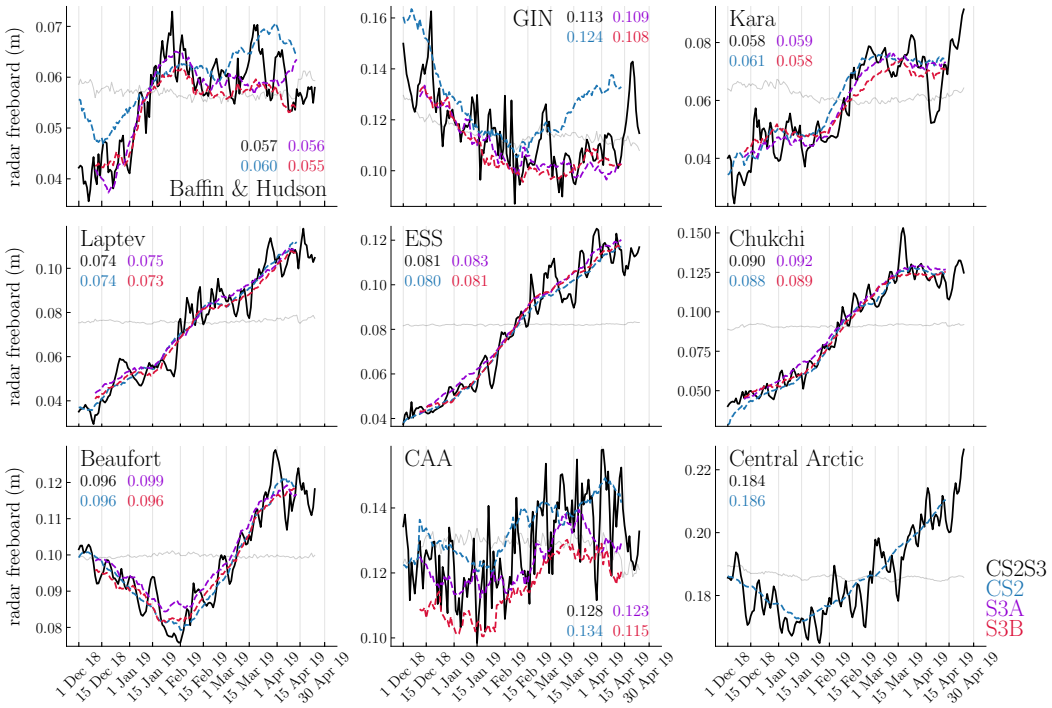


Figure 4.7: Comparison of CS2S3 time series and 31-day running means of CS2 and S3 across 9 Arctic sectors. Sectors are defined based on Fetterer et al. (2010), and are available from NSIDC. GIN: Greenland, Iceland, and Norwegian Seas; ESS: East Siberian Sea; CAA: Canadian Archipelago. Values presented for each graph correspond to the mean (in metres) of each time series. All means are computed across the period 11^{th} of December 2018 to 12^{th} of April 2019 (the period where all four time series are available). Note that S3 time series are not included in the Central Arctic as much of this region overlaps with the S3 polar hole $> 81.5°$ N. Grey lines are the benchmark time series to test the efficacy of the model (see main text).

freeboard is calculated for each day of the 2018–2019 winter season to produce a time series of radar freeboard evolution for each sector. The 31-day running mean freeboard from CS2, S3A and S3B is also plotted to demonstrate the increase in variability when moving from a monthly to a daily product. The 9 Arctic sectors are taken from NSIDC (Fetterer et al., 2010) and include Baffin and Hudson Bays, Greenland, Iceland and Norwegian (GIN), Kara, Laptev, East Siberian, Chukchi, Beaufort seas, the Canadian Archipelago (CAA), and the central Arctic.

In Figure 4.7 we can see how the day-to-day variability is increased with the CS2S3 product, compared to the CS2 and S3 31-day running means. Generally, the mean of the CS2S3 time series lies within 3 mm of CS2 and S3 (approximately in line with the results of the cross-validation presented in section 4.4.2), however large discrepancies exist in the GIN Seas (up to 1.1 cm), the CAA (up to 1.3 cm), and Baffin and Hudson Bays (up to 1.2 cm in December 2018). This is perhaps not surprising given the larger uncertainty in radar freeboard in shallow-shelf seas and at coastal margins, relating to higher uncertainty in interpolated sea level anomalies

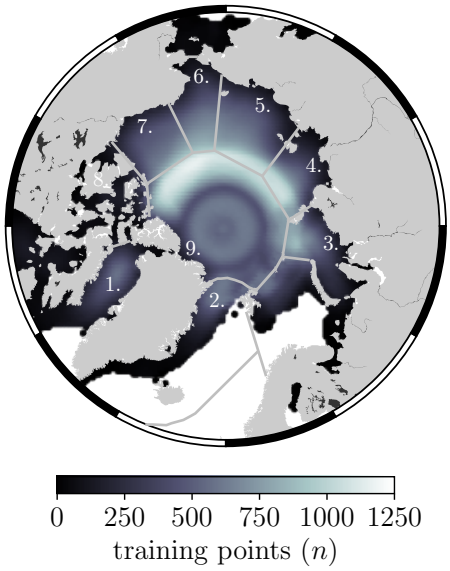


Figure 4.8: The number of training points used to predict radar freeboard at each grid cell at 50×50 km resolution, shown as an average across all days between the 1$^{\text{st}}$ of December 2018 and the 24$^{\text{th}}$ of April 2019. The contour lines mark the boundaries between the 9 NSIDC regions: (1) Baffin and Hudson bays, (2) Greeland, Iceland, and Norwegian (GIN) Seas, (3) Kara Sea, (4) Laptev Sea, (5) East Siberian Sea (ESS), (6) Chukchi Sea, (7) Beaufort Sea, (8) Canadian Archipelago (CAA), (9) central Arctic.

(Lawrence et al., 2019). Indeed, we can see that the difference between the 31-day running means (CS2-S3A(B)) in the CAA is also large, at 1.1(1.9) cm. The GIN region includes the area east of Greenland, one of the most dynamic regions which carries MYI exported out of the Fram Strait southward. Being such a dynamic region, it is likely that the difference in spatial sampling of the three satellites may drive discrepancies between the 31-day running means. Note that the GIN Seas, the CAA, and Baffin and Hudson Bays also coincide with where some of the largest uncertainty in the CS2S3 daily field is observed (e.g., Figure 4.3b), as well as the lowest spatial and temporal freeboard correlation length scales (Figure 4.4), which may be, in part, a reflection of the discrepancies between CS2 and S3 freeboards in these locations, but also due to the limited amount of data available in these regions across the 9-day training period. We can see from Figure 4.8 that the number of training points used to inform on predictions in the coastal margins (including the CAA, GIN Seas, and Baffin and Hudson Bays) is, on average $\lesssim 200$ points, whereas in the central regions (central Arctic, Beaufort, Chukchi, East Siberian, Laptev, and Kara seas), the number of training points is typically > 500. A natural question is then whether the variability seen in the time series in Figure 4.7 represents real physical signal, or is just noise related to observational uncertainty. To address this question, the potential issue of spatial sampling is first considered by creating a 'benchmark' time series (see Figure 4.7). For this, a 'static' background field is initially computed by averaging all CS2 and S3 gridded tracks between the 1$^{\text{st}}$ of

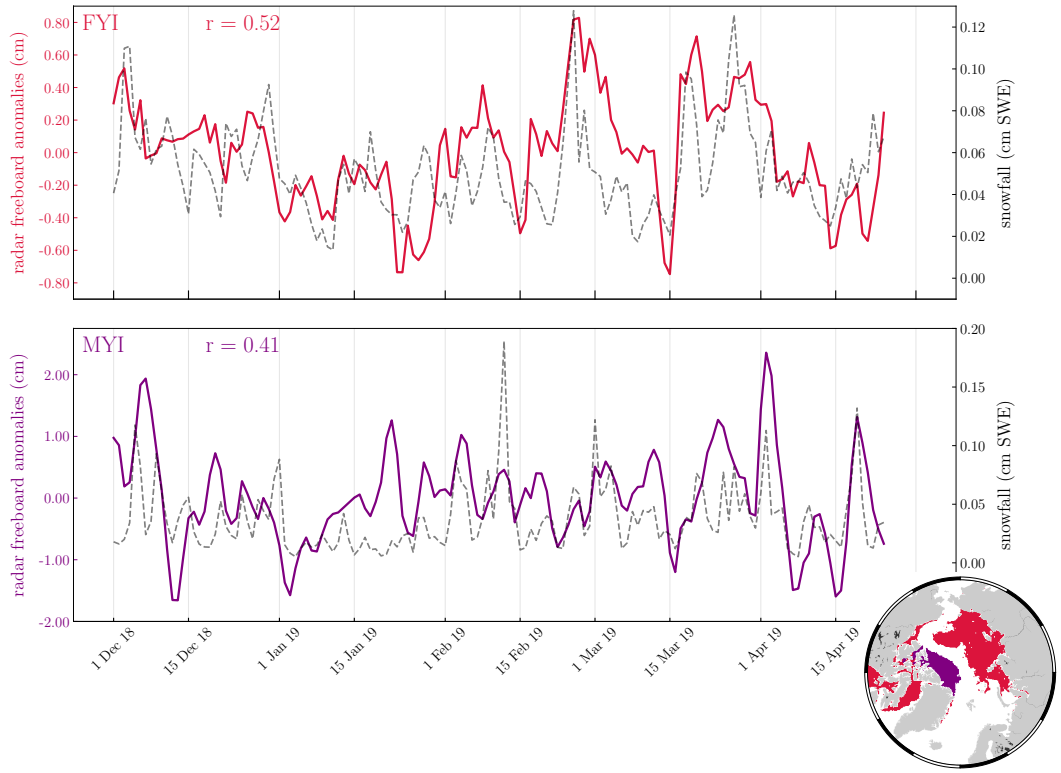


Figure 4.9: Time series of CS2S3 daily radar freeboard anomalies with daily ERA5 snowfall (cm snow water equivalent; SWE). The corresponding contour for FYI and MYI zones are given in the accompanying spatial plot, where each pixel represents a location which has remained either FYI or MYI for the entire 2018–2019 winter season. The linear correlation coefficient (r) between radar freeboard anomalies and snowfall are given for each zone.

December 2018 and the 24th of April 2019, and then re-generating the predictions for each day as per the GPR model, except that the training data now sample from the background field along the CS2 and S3 track locations of the respective days used in the model training. It should therefore be expected that the benchmark predictions for each day be approximately equivalent in areas where there are a significant amount of training data and to see variability in areas where there are less data, which can be explained by tracks sampling different locations of the background field on different days. This is typically what is seen in Figure 4.7, where there is almost zero day-to-day variability of the benchmark time series in regions such as the Beaufort, Chukchi, East Siberian and Laptev seas, but slight variability in the coastal margins. As a second test, the evolution of daily CS2S3 radar freeboard is also compared to daily ERA5 snowfall data, building on the work of Lawrence (2019). Assuming that the radar pulse fully penetrates the sea ice snow cover, radar freeboard is expected to change with snowfall by two distinct mechanisms: (a) snow loading of sea ice will depress the sea ice floe into the ocean, reducing the sea ice freeboard and therefore also the radar freeboard, (b) because of the slower speed of radar propagation through snow compared to air, additional snowfall will further slow

down the pulse, resulting in delayed receipt of the echo at the satellite, equating to a reduced radar freeboard. Thus, under the assumption of full snow penetration, radar freeboard is expected to decrease with increasing snowfall according to these two mechanisms. In Figure 4.9 the average of each field is computed over FYI and MYI zones for each day, since surface roughness and snow depth typically show different characteristics over the two ice types (Tilling et al., 2018), and therefore variability of radar freeboard with snowfall may also show different signals. Freeboard anomalies are then generated by subtracting a second-order polynomial fit to each of the time series. In contrast to expectations, a strong positive correlation of 0.52 is observed over FYI, and a medium positive correlation of 0.41 over MYI. While it goes beyond the scope of this study to investigate the specific drivers of this correlation, these results appear to challenge conventional assumptions of full snow penetration at Ku-band frequency. The derived correlation of 0.52 and 0.41, together with the benchmark test, suggest that the CS2S3 data are able to capture physical radar freeboard variability at sub-weekly time-scales.

4.6 Conclusions

This chapter presented a methodology for deriving daily gridded pan-Arctic radar freeboard estimates through Gaussian process regression, a Bayesian inference technique. An example was shown of how this method uses 9 days of gridded freeboard observations from CryoSat-2 (CS2), Sentinel-3A (S3A) and Sentinel-3B (S3B) satellites in order to model spatio-temporal covariances between observation points, and make pan-Arctic predictions of radar freeboard, with uncertainty estimates, on any given day at 25×25 km resolution. This product is referred to as CS2S3. The empirical Bayes approach was also highlighted as a way to estimate the hyperparameters which define the covariance function, and to subsequently derive pan-Arctic estimates of the spatial and temporal correlation length scales of radar freeboard. An evaluation of the model performance was then carried out for both training and predictions at 50×50 km resolution for computational efficiency. For the training points, CS2S3 freeboards were compared to gridded CS2 and S3 freeboards at co-located points for all days across the 2018–2019 winter season, where the differences were found to follow a normal distribution, with mean errors ≤ 1 mm, and standard deviations < 6 cm. The predictive performance of the model was then evaluated for latitudes below 81.5° N and at locations where sea ice concentration is $\geq 75\%$, based on a cross-validation approach whereby different combinations of S3 freeboards were withheld from the training set of observations, the daily predictions were then re-generated, and the withheld data were used to validate the predictions. The general prediction error was also found to be normally distributed, with mean errors ≤ 4 mm, and standard deviations < 7.5 cm. Finally, the improved temporal

variability of a daily pan-Arctic freeboard product was presented by comparing time series of CS2S3 freeboards, with 31-day running means from CS2 and S3 observations, in 9 different Arctic sectors. The mean of the CS2S3 time series were generally within 3 mm of the 31-day running mean time series from CS2 and S3, except for the Canadian Archipelago and the Greenland, Iceland and Norwegian seas, where prediction uncertainty is large due to significant discrepancies between CS2 and S3 freeboards. Two pieces of analysis were then presented to conclude that the variability seen in the CS2S3 radar freeboard time series is related to real physical signal rather than noise related to observational uncertainty. One based on a benchmark where predictions were generated from a static field for all days across the 2018–2019 winter season, and another where time series of daily radar freeboard anomalies were compared against daily ERA5 snowfall data over first-year ice (FYI) and multi-year ice (MYI) zones. Linear correlation coefficients between freeboard and snowfall in these regions were given as 0.52 over FYI, and 0.42 over MYI, suggesting that the daily fields produced by Gaussian process regression are able to capture real radar freeboard variability. Interestingly, the positive correlation between radar freeboard seemingly contradicts conventional assumptions of full snow penetration at Ku-band radar frequency, and remains the subject of future investigation. The improved temporal variability from a daily product is a hopeful prospect for improving the understanding of physical processes that drive radar freeboard and/or thickness variability on sub-monthly timescales. Of course, an investigation into the drivers of the temporal variability in the CS2S3 field would be an additional way to validate the product, although this goes beyond the scope of the study here. This chapter concludes that the Gaussian process regression method is an extremely robust tool for modelling a wide range of statistical problems, from interpolation of geo-spatial data sets, as presented here and in other works (Le Traon et al. 1997; Ricker et al. 2017), to time series forecasting (e.g., chapter 3). The Gaussian assumption holds well in many environmental applications, and the fact that the Gaussian process prior can take any number of forms, so long as the covariance matrix over the training points is symmetric and positive semi-definite, means that the model can be tailored very specifically to the problem at hand.

Chapter 5

Climate connectivity in general circulation models

The following chapter uses the unsupervised learning complex networks methodology to establish how well the latest generation of coupled climate models participating in phase 6 of the World Climate Research Programme Coupled Model Intercomparison Project (CMIP6) are able to reflect the spatio-temporal patterns of variability in northern-hemisphere winter sea-level pressure and Arctic summer sea ice concentration over the period 1979–2020. Two specific global metrics are introduced as ways to compare patterns of variability between CMIP6 sea-level pressure networks and ERA5 atmospheric reanalysis, as well as between CMIP6 sea ice concentration networks and three satellite-derived observational data sets. These metrics include the Adjusted Rand Index, and a network distance metric which compares the degree of connectivity between two geographic regions. Subsequently, the regional connectivity between the winter Arctic Oscillation and summer sea ice concentration network nodes is evaluated in the observations and CMIP6 models over the same period, as well as an assessment of how this teleconnection may be changing over time as the ice cover changes. A discussion is then presented outlining the implications of any differences in the representation of this teleconnection between observations and models, in the context of seasonal dynamical model predictions of summer sea ice.

5.1 Introduction

Understanding the drivers of inter-annual sea ice variability has important consequences for our ability to make reliable sea ice predictions on seasonal to inter-annual time-scales (e.g., section 1.1.3). A number of studies for example have highlighted various climatological teleconnections as key drivers of sea ice variability, including

land-ice interactions (typically via the atmosphere; Serreze et al. 1995; Overland et al. 2012; Matsumura et al. 2014; Crawford et al. 2018), atmosphere-ice (Deser et al. 2000; Kapsch et al. 2013; Park et al. 2018; Olonscheck et al. 2019), ocean-ice (Venegas & Mysak 2000; Vinje 2001; Zhang 2015), and ice-ice (Schröder et al. 2014; Bushuk et al. 2017), suggesting inherent sources of sea ice predictability across the various components of the climate system. Furthermore, coupled climate models have demonstrated a horizon of sea ice predictability beyond 12 months lead time based on so-called 'perfect-model' experiments (Holland et al. 2011; Tietsche et al. 2014; Day et al. 2014; Bushuk et al. 2019), however in reality, regional operational summer sea ice forecasts (in models) appear to be strongly controlled by a 'spring predictability barrier' (Bonan et al., 2019), which is governed by the date of melt onset in the preceding spring (Bushuk et al., 2020). This leads to the question as to how well climate models reflect the teleconnections known to drive summer sea ice variability, given the gap between perfect-model and operational regional forecast skill in those models (Bushuk et al., 2019). Recent work has gone into investigating physically-based mechanisms for sea ice predictability (Bushuk et al. 2017; Bonan & Blanchard-Wrigglesworth 2020; Giesse et al. 2021) in order to assess whether the shortfalls in operational climate model forecasts can, in part, be attributed to a lack of representation of such mechanisms across a wide range of general circulation models (GCMs).

This chapter pursues a similar line of investigation, looking specifically at the Arctic Oscillation (AO) teleconnection (Thompson & Wallace, 1998), whose winter pattern has been shown to explain up to 22% of the variability in pan-Arctic September sea ice extent (Park et al., 2018). Historically, the dominant spatio-temporal modes of winter sea-level pressure variability have been somewhat mis-represented in the majority of GCMs participating in previous phases of the Coupled Model Intercomparison Project (CMIP), e.g., CMIP3 (Miller et al. 2006; Cattiaux & Cassou 2013) and CMIP5 (Jin-Qing et al. 2013; Gong et al. 2016), which naturally has implications for the representation of the AO to sea ice teleconnection in those models. Here, the spatio-temporal patterns (modes) of variability in both winter sea-level pressure and summer sea ice concentration, and also the presence of the winter AO to summer sea ice teleconnection are assessed over the period 1979–2020, in 31 of the latest generation of GCMs submitted to CMIP6 (Eyring et al., 2016), as well as how this teleconnection may be changing over time as the ice cover thins and is more susceptible to atmospheric forcing (Maslanik et al. 1996; Mioduszewski et al. 2019). The complex networks methodology (section 2.3.3) is used here to derive climatological teleconnections which are traditionally defined as the leading mode of variability in their respective climate fields, in order to assess which regions the models show are the dominant locations of either winter sea-level pressure or summer sea ice concentration variability, and also how similarly the models reflect the regional

responses of summer Arctic sea ice to different phases of the winter AO.
This chapter is structured as follows: section 5.2 introduces the sea ice observations, atmospheric reanalysis, and CMIP6 model data that are used to generate complex networks of the respective sea ice and atmospheric fields. In section 5.3 the complex networks methodology is revisited, as well as outlining two global metrics which are used to describe similarities and differences between networks. Section 5.4 presents the results of the networks generated from the 31 different CMIP6 models and discusses their similarities and differences relative to the observations and reanalysis data, and furthermore evaluates the presence of the winter AO to summer sea ice teleconnection across all models. A discussion and conclusions are then given in section 5.5.

5.2 Data

5.2.1 Observations

For analysing summer sea ice concentration variability in the observations, the average of monthly mean June, July, August and September (JJAS) sea ice concentration fields are computed between 1979 and 2020 from three separate observational data sets based on the series of multi-frequency passive microwave satellite observations since October 1978. These include the National Snow and Ice Data Center (NSIDC) NASA Team (Cavalieri et al., 1996), and Bootstrap (Comiso, 2017) products, as well as the Ocean and Sea Ice Satellite Application Facility (OSI-SAF) OSI-450 (1979–2015) and OSI-430-b (2016–2020) products (OSI-SAF 2017; Lavergne et al. 2019). Three different products are used as each has subtle variations in their summer variability, from how each account for new melt-pond formation (Comiso et al., 2017). Each of the three data sets apply separate processing algorithms to passive microwave brightness temperatures derived from multiple satellites across the historical record: Nimbus-7 SMMR (1979–1987), the DMSP F-8, F-11 and F-13 SSM/Is (1987–2007), and finally the DMSP F-18 SSM/I (2007–2020). These data are provided on 25 $\times$ 25 km polar stereographic (NASA Team and Bootstrap), and EASE (OSI-450 and OSI-430-b) grids, which are re-gridded to a common 50 $\times$ 50 km polar stereographic grid using a nearest neighbour interpolation here for computational reasons. Grid cell area information (used to generate area-weighted time series, see section 5.3.1) was also extracted from NSIDC's pixel area tools library.

5.2.2 Atmospheric Reanalysis

The AO is typically defined as the leading mode of variability in mean sea-level pressure data north of 20° N (Thompson & Wallace, 1998). As a proxy for an

Table 5.1: CMIP6 models used in this study.

Model	No. Ensembles
ACCESS-CM2	1
ACCESS-ESM1-5	1
BCC-CSM2-MR	1
CAMS-CSM1-0	1
CanESM5-CanOE	1
CanESM5	20
CAS-ESM2-0	1
CESM2	1
CESM2-WACCM	3
CMCC-CM2-SR5	1
CMCC-ESM2	1
CNRM-CM6-1	6
CNRM-ESM2-1	1
EC-Earth3	1
EC-Earth3-Veg	2
FGOALS-g3	1
FIO-ESM-2-0	1
GFDL-CM4	1
GFDL-ESM4	1
HadGEM3-GC31-LL	3
HadGEM3-GC31-MM	4
IPSL-CM6A-LR	6
MIROC6	3
MIROC-ES2L	1
MPI-ESM1-2-HR	1
MPI-ESM1-2-LR	1
MRI-ESM2-0	1
NESM3	1
NorESM2-LM	1
NorESM2-MM	1
UKESM1-0-LL	5

observational record of the winter AO here, the average of monthly mean December, January, February and March (DJFM) mean sea-level pressure data is computed north of 20° N, from ERA5 reanalysis (ERA5, 2017) between 1979–2020. Only one reanalysis product is used due to the high consistency of sea-level pressure fields between different reanalyses over the Arctic region (Graham et al., 2019). As December data are not available for the year 1978 for ERA5, the winter period in 1979 corresponds to the average of January, February and March data. Sea-level pressure fields are output on a $2° \times 4°$ latitude-longitude grid.

Section 5.5 also makes use of the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) sea ice thickness model (Zhang & Rothrock, 2003) to help explain some of the features related to the winter AO to summer sea ice teleconnection. PIOMAS is a coupled ice-ocean model that assimilates observed sea ice concentration and sea-surface temperatures (open water only), and is forced by NCEP-NCAR atmospheric reanalysis. Although it is a model, it has been shown to be relatively

consistent with in-situ and submarine observations (Schweiger et al., 2011), and generally has consistent biases with CMIP3/5 models relative to observational data, in terms of its ice thickness distribution (Stroeve et al., 2014b). Furthermore, it is able to provide consistent coverage over the observational period.

5.2.3 CMIP6 model outputs

The seasonal patterns of variability in sea ice concentration, and mean sea-level pressure are assessed in 31 different GCMs participating in CMIP6. In order to compare with recent observations, monthly averaged model outputs from historical runs (1979–2014) are combined with ScenarioMIP run SSP5-8.5 (Gidden et al., 2019) to extend the analysis period to 2020, hence only model ensembles which contain historical and ScenarioMIP outputs for both sea ice concentration and mean sea-level pressure are considered in this work. As detailed above, the corresponding winter (DJFM) and summer (JJAS) averages are computed for mean sea-level pressure and sea ice concentration outputs respectively. Sea ice concentration outputs are also re-gridded to the same 50 $\times$ 50 km polar stereographic grid as the observational data sets, and mean sea-level pressure outputs are re-gridded to a $2^\circ \times 4^\circ$ latitude-longitude grid. The chosen models, along with their respective number of available ensembles are summarised in Table 5.1.

5.3 Method

5.3.1 Complex networks

Generating complex networks follows the methodology outlined in section 2.3.3. The key steps are summarised here, as well as presenting an example DJFM sea-level pressure network from ERA5, and JJAS sea ice concentration networks from the observations.

Recall that a climate network of N nodes corresponds to time series data $\boldsymbol{G} = \{\boldsymbol{g}_k\}_{k=1}^{N}$ representing n regularly sampled observations in time $\boldsymbol{g}_k = (g_{1k}, g_{2k}, \ldots, g_{nk})$ at N fixed geographical locations, and the links represent statistical interdependencies between any pair of node time series $\boldsymbol{g}_k$ and $\boldsymbol{g}_l$. In more detail, let us define $\mathbf{x} = \{\boldsymbol{x}_p\}_{p=1}^{P}$ as a de-trended (zero-mean) time series data set (e.g., DJFM mean sea-level pressure anomalies or JJAS sea ice concentration anomalies), which represents $n = 42$ regularly sampled observations in time $\boldsymbol{x}_p = (x_{1p}, x_{2p}, \ldots, x_{np})$ at P fixed geographical locations, such that $\mathbf{x} \in \mathbb{R}^{P \times n}$. The N network nodes are then derived by implementing the grid-based clustering algorithm to the input data set (section 2.3.2), so that the dimensionality of $\mathbf{x}$ is reduced from P to N. Links between the nodes are then generated by first computing the cumulative anomaly time

series of each network node, which for a given node $\mathcal{C}_k$, is taken as the sum of the grid-cell-weighted de-trended time series of all cells within that node:

$$\boldsymbol{g}_k = \sum_{p \in \mathcal{C}_k} \mathbf{x}_p \sqrt{\mathrm{w}_p}, \tag{5.1}$$

where $\mathrm{w}_p = \cos(\theta_p)$ for a regular latitude-longitude grid (θ_p is the latitude of grid cell p), or simply $\mathrm{w}_p = d_p$ for a polar stereographic area grid (d_p is the area in km^2 of grid cell p). Subsequently, the link weight between two nodes k and l is calculated as the temporal covariance between two network node anomaly time series:

$$w_{kl} = \frac{1}{n-1} \sum_{i=1}^{n} (g_{ik} - \mathbb{E}[\boldsymbol{g}_k])(g_{il} - \mathbb{E}[\boldsymbol{g}_l]). \tag{5.2}$$

Finally, the strength of a given network node $\mathcal{S}_k$ is defined as the sum of the absolute value of all its associated link weights:

$$\mathcal{S}_k = \sum_{l=1}^{N} |w_{kl}|. \tag{5.3}$$

Recall that the node with the highest strength generally represents the leading mode of variability of the input data set $\mathbf{x}$, and by this definition, the node with the highest strength belonging to the network of mean sea-level pressure data is equivalent to the AO. Note also that this network framework allows for weighted links between nodes of a single network (as detailed above), and also between nodes of multiple networks (i.e., between nodes of sea-level pressure and sea ice concentration), which is used to assess the winter AO to summer sea ice teleconnection in section 5.4.3.

Figure 5.1a shows the network structure of DJFM mean sea-level pressure data from ERA5. In this case we can see how the network nodes correspond to a set of spatially contiguous areas, where for a given node, each grid cell is weighted by the strength of the node in which that cell belongs. As mentioned, the node with the highest strength in Figure 5.1a is equivalent to the spatial pattern of variability of the AO, therefore from this map we can see that the AO corresponds to the large node situated over the majority of the Arctic Ocean, Greenland, the Canadian Archipelago, and parts of northern Russia. The weighted links then illustrate how each of the nodes have co-varied relative to each other over the period 1979–2020, and indeed we notice the out-of-phase relationship (negative covariance) between the AO node and the mid-latitude Atlantic sector; highlighting the dipole nature of the North Atlantic Oscillation (Hurrell et al., 2003).
It is also possible to extract the temporal component of variability from the 'AO node' (Figure 5.1b), which produces a very consistent signal with the standard AO

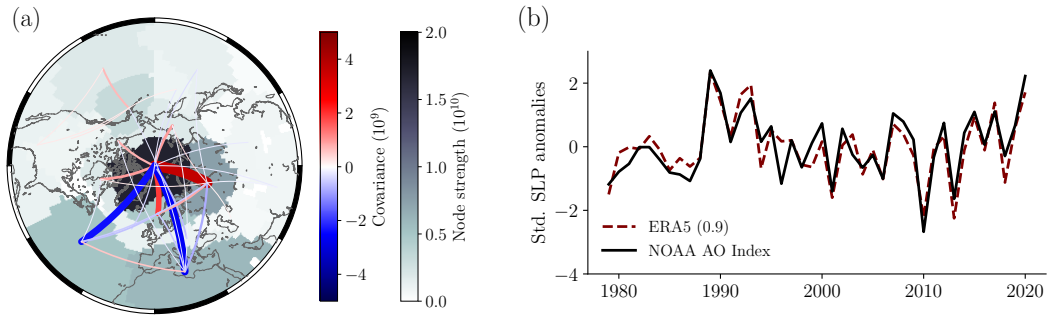


Figure 5.1: (a) Complex network of DJFM mean sea-level pressure from ERA5. Only links which have a corresponding p-value < 0.10 are shown here to aid visualisation. (b) The standardised (Std.) leading temporal mode of variability (dashed curve), extracted from the ERA5 network node with the highest strength. The number in parentheses corresponds to the linear correlation coefficient with the DJFM AO index from NOAA (see main text), available from https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml.

index as defined by the National Oceanic and Atmospheric Administration (NOAA), highlighting the robustness of the complex networks method. It is worth noting however that in Figure 5.1b, and indeed for the rest of this manuscript, the sign of the temporal component of each node of the winter sea-level pressure networks (from ERA5 and CMIP6 data) is reversed in order to be consistent with the standard AO index, for which positive AO index values correspond to low atmospheric pressure, and similarly negative AO index values correspond to high atmospheric pressure.

Figure 5.2 shows similar networks for JJAS sea ice concentration, from each of the observational products. Here we can see how the leading spatio-temporal modes of summer sea ice variability (i.e., highest node strengths) are typically in the East Siberian and Laptev seas, as well as the Canada basin. Each observational product generally shows the same structure of largely positive covariance between network nodes, and the out-of-phase connection between the Fram Strait and the Pacific sector. The magnitude of the node strengths between the observational data sets somewhat varies in the dominant regions of variability, with the Bootstrap product showing the largest strengths in the East Siberian and Laptev seas. In the next section we introduce two global metrics for deriving quantitative measures of similarity between networks.

5.3.2 Metrics for comparing networks

Before introducing the two metrics which are used to compare similarities between complex networks, it is worth saying a few words about what information we can expect to obtain when comparing models and observations/reanalysis. Due to the fact that any CMIP6 model ensemble is in its own phase of internal variability (e.g.,

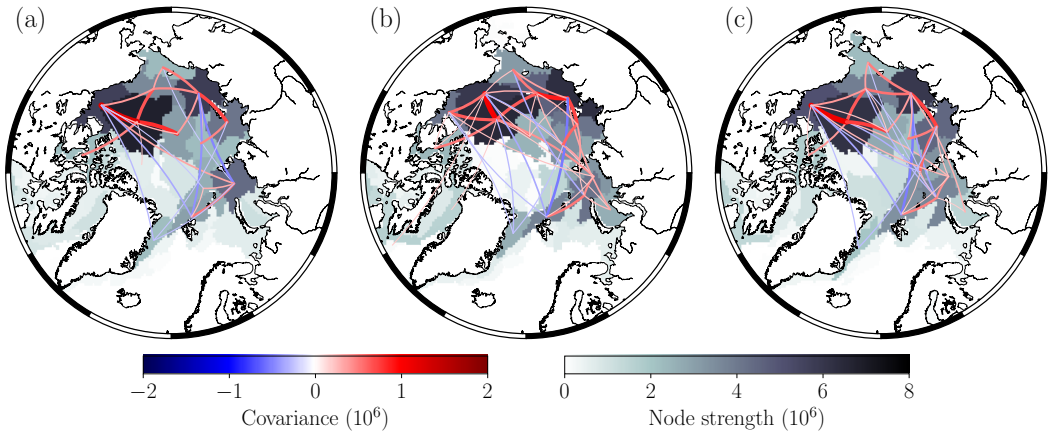


Figure 5.2: Complex networks of JJAS sea ice concentration from (a) NASA Team, (b) Bootstrap, and (c) OSI-SAF data sets, computed between 1979–2020. Only links which have a p-value < 0.10 are shown here to aid visualisation.

Hawkins & Sutton 2009; Notz 2015), we cannot expect to find consistency in the sign and magnitude of anomalies between e.g., the ERA5 AO time series, and that of any one model ensemble (and similarly for sea ice), therefore it would not be prudent to perform any analysis which makes direct comparisons of the temporal modes between any observational and CMIP6 network nodes. We can however expect a model with accurate physics to reproduce similar dominant regions of variability as the observations, and the same sign and magnitude of the inter-connected links between nodes, e.g., the strong negative coupling between the AO and sea-level pressure anomalies in the north Atlantic, and the weak negative coupling with the north Pacific (see Figure 5.1a), and similarly we can expect the same regional responses of Arctic sea ice to different phases of the AO between observations and models. The two metrics introduced in the coming sections provide a way to quantify similarities in the locations which the observations and models define as the dominant regions of variability, and the connectivity of these regions, without explicitly comparing network node time series.

Adjusted Rand Index The Adjusted Rand Index (ARI; Hubert & Arabie 1985) is a metric which is often used to evaluate similarities between sets of clusters (Steinley, 2004), and as such it is used here to compare how two networks have clustered grid cells together to form their spatially contiguous set of network nodes, and subsequently their spatial patterns of either sea-level pressure or sea ice concentration variability. To understand the ARI, it is worth briefly introducing the (un-adjusted) Rand Index by following the example outlined by Rand (1971). First, consider two different synthetic networks which have clustered grid cells together in two distinct

Table 5.2: Synthetic example of cell clusters to illustrate the concept of the Rand Index (see main text), after Rand (1971).

Cell pairs	ab	ac	ad	ae	af	bc	bd	be	bf	cd	ce	cf	de	df	ef	**Total**
Together in both	✓															1
Separate in both			✓	✓	✓		✓	✓	✓		✓	✓		✓		9
Mixed		✓				✓				✓			✓		✓	5

ways:

$$\text{Network}_1\ (\text{N}_1) = [(a, b, c), (d, e), (f)]$$
$$\text{Network}_2\ (\text{N}_2) = [(a, b), (c, d), (e, f)].$$

In these two simple network constructions, there are 3 nodes in each network, where each node contains a clustering of cells labelled a–f. The Rand Index then measures similarities and differences in the clustering of these cells by analysing all the possible cell pairings between the two networks (see Table 5.2). In this example there are a total of 10 similarities (grid cells which are clustered together in both networks *and* grid cells which are separate in both networks) out of a possible 15 pairings, which gives a Rand Index score of $10/15 = 0.67$. The ARI is then an update of the Rand Index, which takes into account the fact that grid cells could be clustered together by chance. This can be understood more clearly by creating a contingency table (also known as a confusion matrix) of the synthetic cell clusters above:

$\text{N}_1 \backslash \text{N}_2$	Node 1	Node 2	Node 3	Sums
Node 1	$P_{11} = 2$	$P_{12} = 1$	$P_{13} = 0$	$a_1 = 3$
Node 2	$P_{21} = 0$	$P_{22} = 1$	$P_{23} = 1$	$a_2 = 2$
Node 3	$P_{31} = 0$	$P_{32} = 0$	$P_{33} = 1$	$a_3 = 1$
Sums	$b_1 = 2$	$b_2 = 2$	$b_3 = 2$	

where each entry describes the number of grid cells which are common to network nodes i and j from networks N_1 and N_2 respectively. The ARI is then given as:

$$\text{ARI} = \frac{\sum_{ij} \binom{P_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{P}{2}}{\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] / 2 - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{P}{2}} = 0.07, \tag{5.4}$$

where P is the total number of grid cells (in this case $P = 6$). In plain terms, the ARI is equivalent to the normalised deviation of the Rand Index from the expected number of cell pairs between all nodes (Steinhaeuser & Chawla, 2010), and therefore varies between 0 (totally dissimilar clustering) and 1 (identical clustering). Computing the ARI between e.g., the NASA Team and OSI-SAF summer sea ice concentration networks produces a value of 0.69, showing relatively consistent clustering (as seen qualitatively in Figure 5.2).

Network distance metric The network distance metric (D; Fountalis et al. 2015) provides a way to compare networks in terms of both the spatial extent of their network nodes, and also their node strengths. Recall that node strength incorporates information about the connectivity of a particular node (i.e., the magnitude of all its connected links), hence when comparing the strength of a particular region between models and observations, it is possible to infer which one has a larger degree of co-variability across the network. This allows us to deduce whether models over- or under-estimate the magnitude of variability of a particular region without comparing node time series. Consider for example the underlying map of node strengths in Figure 5.2. D can be computed by first taking the sum of the absolute difference between two of these 'strength maps' M_1 and M_2 (e.g., NASA Team and OSI-SAF), and then normalising by the sum of the absolute difference between random permutations of both network strength maps, $\hat{M}_1$ and $\hat{M}_2$:

$$D = 1 - \frac{\sum_{p=1}^{P} |M_{1p} - M_{2p}|}{\sum_{p=1}^{P} |\hat{M}_{1p} - \hat{M}_{2p}|}. \tag{5.5}$$

A value of $D = 1$ means that both networks are identical in their node strengths and spatial extent of nodes, whereas a value close to $D = 0$ implies that the two networks are as similar as a random assignment of node strengths to grid cells. Computing D between the NASA Team and OSI-SAF summer sea ice concentration networks produces a value of 0.86.

The combination of ARI and D allows us to infer various properties between two networks. For example, when ARI $= 1$ and $D = 0$, this suggests that two networks agree in terms of which grid cells have behaved homogeneously over the length of the time series record in order to cluster together to form network nodes, however they disagree in terms of the magnitude of variability of the nodes. On the other hand, if ARI is close to 0 and D is close to 1, then this implies that the magnitude of variability across the networks are relatively consistent, however the geographic areas which are clustered together to form network nodes are considerably different. Two networks can then be considered identical if ARI and $D = 1$.

5.4 Results

5.4.1 Sea-level pressure networks in CMIP6

For every available ensemble from each of the CMIP6 models outlined in Table 5.1, individual complex networks of DJFM sea-level pressure are computed between 1979 and 2020, and then ARI and D metrics are computed relative to the ERA5

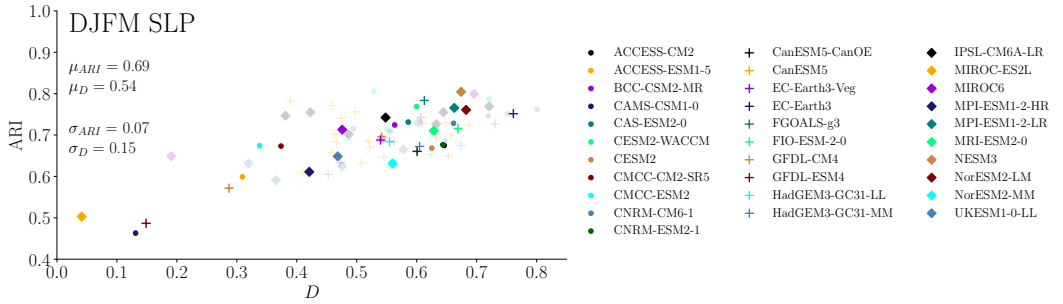


Figure 5.3: ARI and D metrics for winter sea-level pressure (SLP) networks computed for every ensemble member for 31 different CMIP6 models (74 realisations), relative to ERA5 atmospheric reanalysis. The semi-transparent colours represent individual ensemble members (where the number of ensembles is greater than 1), and the opaque colours are the mean of all ensemble members. The mean and standard deviation across all points are given by μ and σ respectively.

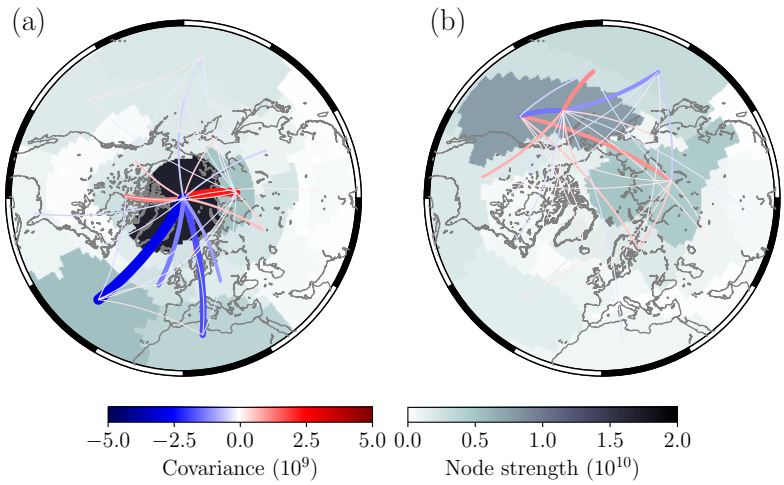


Figure 5.4: Winter sea-level pressure networks from (a) CNRM-CM6-1 (ensemble: r6i1p1f2) and (b) MIROC-ES2L (ensemble: r1i1p1f2). The CNRM-CM6-1 model produces ARI and D values 0.76 and 0.80 respectively, while the MIROC-ES2L model produces values 0.50 and 0.04 respectively. Only links which have a corresponding p-value < 0.10 are shown here to aid visualisation.

sea-level pressure network. In Figure 5.3 we can see how the spread in D values across all model ensembles is over twice as large as for the ARI values, which suggests large inter-model disagreement on the degree of connectivity of network nodes, and hence the magnitude of regional sea-level pressure variability. Note that the apparent linear relationship between ARI and D is to be expected, given that both metrics encapsulate information related to the spatial agreement of network nodes between models and ERA5. Across all models, CNRM-CM6-1 produces the most similar network structure to ERA5, with ARI $= 0.76$ and $D = 0.80$. Figure 5.4 shows the corresponding network for CNRM-CM6-1, and also MIROC-ES2L. The MIROC-ES2L model produces the most dissimilar network structure relative to ERA5, with ARI and D of 0.50 and 0.04 respectively. The networks show how

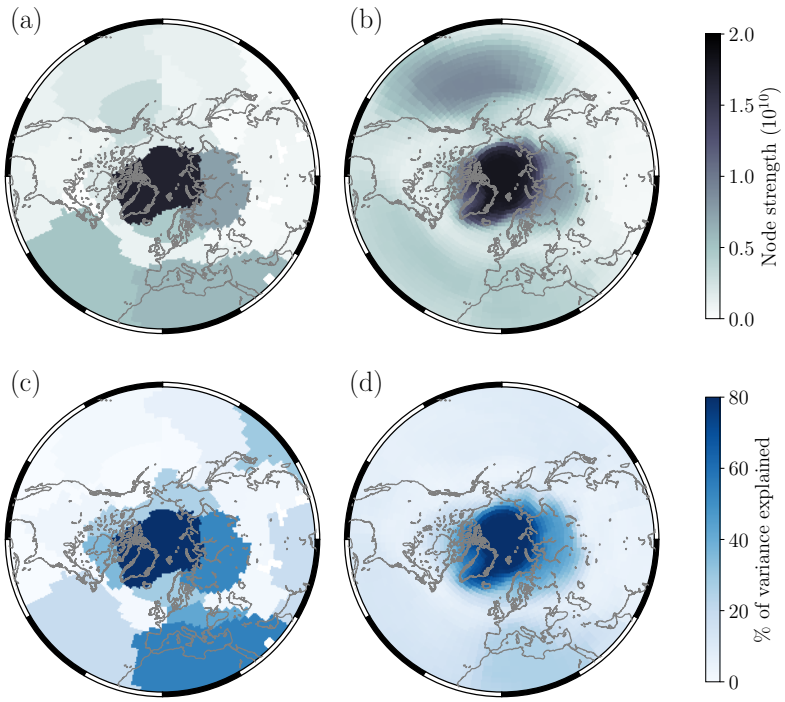


Figure 5.5: (a)-(b) The spatial patterns of winter sea-level pressure variability (node strengths) from (a) ERA5, and (b) the average of all 74 CMIP6 model ensemble members. (c)-(d) The percentage of variance in mean northern-hemisphere winter sea-level pressure explained (e.g., Björnsson & Venegas 1997) by network nodes in (c) ERA5, and (d) the average of all 74 CMIP6 model ensemble members.

the CNRM-CM6-1 model produces the relatively consistent node of high strength over the Arctic Ocean (similar to ERA5), and also shows the same strong negative linkage with the mid-latitude Atlantic sector, and weak linkage with the Pacific sector. On the other hand, the MIROC-ES2L model shows significantly different regions of variability than ERA5, and also weaker connectivity, with overall weaker link weights and very low strength over the Arctic Ocean. Figure 5.5a-b shows the average of each of the network strength maps across all of the CMIP6 model ensemble members, and compares this with the ERA5 strength map. While this removes the ability to identify individual network nodes and their links, it does allow us to qualitatively assess how CMIP6 models, on average, represent the spatial patterns of winter sea-level pressure variability and their degree of connectivity. We notice for example, that on average CMIP6 models represent the spatial pattern of the AO relatively well, although slightly under-estimate its node strength. Furthermore, node strengths in the north-western Pacific Ocean appear to be over-estimated on average, while they are under-estimated over north Africa, and southern Europe. In Figure 5.5c-d, the percentage of variance in mean northern-hemisphere sea-level pressure anomalies that is explained by each ERA5 network node, and the average of CMIP6 nodes is also shown (the mean sea-level pressure anomalies in CMIP6 models are

computed for each individual model ensemble, and then the percentage of variance is computed between this signal and its own respective sea-level pressure network nodes). We can see that the nodes centred over the Arctic Ocean explain the highest percentage of variance in northern-hemisphere sea level pressure in both ERA5 and CMIP6 networks, however the models under-estimate the relative importance of the north Atlantic and north Africa–southern Europe region in explaining winter sea-level pressure variability. It is also worth mentioning that although the CMIP6 models identify the north Pacific as a region of strong co-variability (Figure 5.5b), the percentage of variance explained by this region is relatively low. This can occur due to the fact that network nodes which are larger in spatial extent will naturally show higher covariance with other regions (and hence node strengths), because the temporal component of variability of a given node corresponds to the sum of all grid cell time series within that node (see Equation 5.1). If however a node's physical correlation (i.e., standardised covariance) with the mean sea-level pressure signal is relatively weak, then this results in a squared reduction in percentage of variance explained (recall, percentage of variance explained = correlation^2).

5.4.2 Sea ice concentration networks in CMIP6

Individual complex networks of JJAS sea ice concentration are now computed between 1979 and 2020 for every available ensemble from each of the CMIP6 models, and then ARI and D metrics are computed relative to the NASA Team, Bootstrap and OSI-SAF sea ice concentration networks. In Figure 5.6 we see a lower spread in ARI values than compared to the D values which, similar to the sea-level pressure networks, suggests that the models show large disagreement on the degree of connectivity of network nodes, and hence the magnitude of regional sea ice concentration variability. What is perhaps noticeable is that the models which appear to perform better in terms of their summer sea ice ARI and D scores, are not necessarily the same as those that score well for their winter sea-level pressure networks – discussed further in section 5.5. Figure 5.7 shows sea ice concentration networks from the MIROC6 and the CAMS-CSM1-0 models. The MIROC6 model produces closer patterns of variability to the observations than other CMIP6 models, with ARI values of 0.48 (NASA Team), 0.48 (Bootstrap), 0.47 (OSI-SAF), and D values of 0.66 (relative to each observational network). Having said that, we can see that the spatial extent and strength of the node in the Beaufort Sea–Canada basin is somewhat under-estimated, and the node strength in the Laptev Sea is over-estimated, and interestingly its link between the Beaufort Sea and the East Siberian Sea is negative. It does however produce consistent out-of-phase network links between the Fram Strait and the Eurasian–Pacific sectors of the Arctic. The CAMS-CSM1-0 model produces a more dissimilar score with ARI values of 0.33 (NASA Team), 0.30 (Bootstrap),

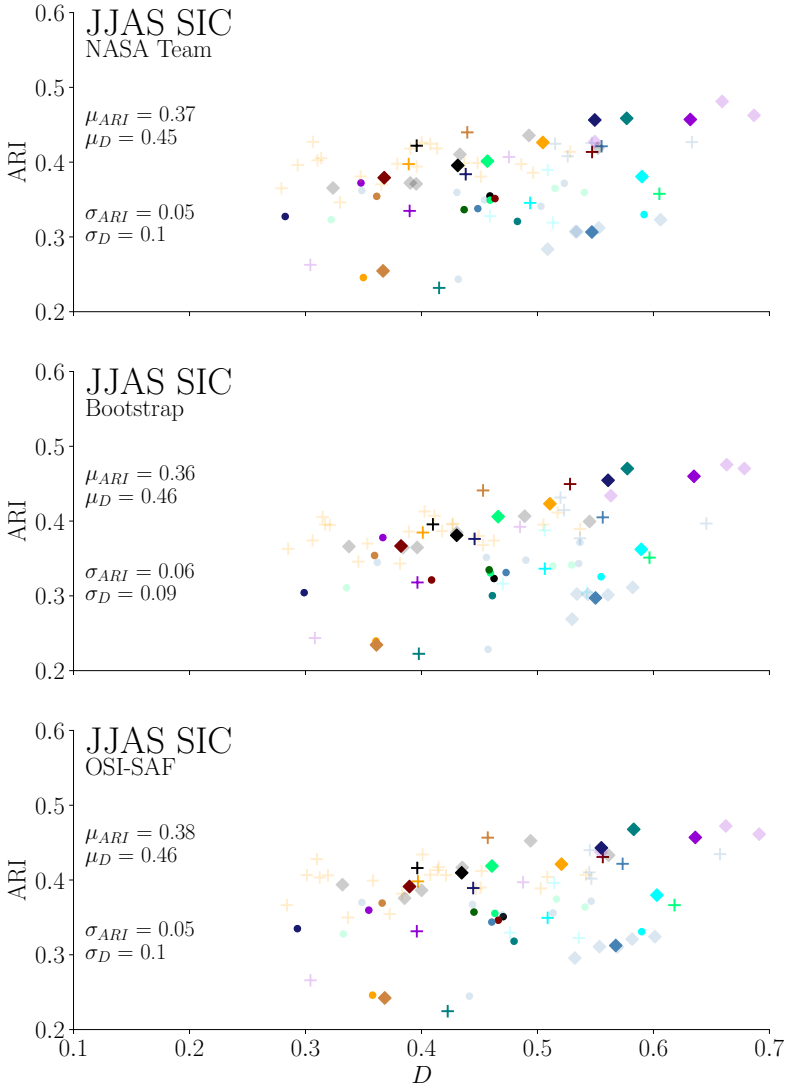


Figure 5.6: ARI and D metrics for summer sea ice concentration (SIC) networks computed for every ensemble member for 31 different CMIP6 models (74 realisations). ARI and D are computed relative to NASA Team (top), Bootstrap (middle) and OSI-SAF (bottom) observational networks. The symbols and colours of each point are consistent with Figure 5.3.

0.33 (OSI-SAF), and D values of 0.28 (NASA Team), 0.30 (Bootstrap), and 0.29 (OSI-SAF). The low D values are being caused by the significant over-estimation in link weights, and hence node strengths, in the Greenland, Iceland, and Norwegian seas, Barents Sea, East Siberian Sea and Laptev Sea (notice how the link weights and node strengths in this model are in some cases an order of magnitude higher than the observational networks).

Figure 5.8a-b shows the average of each of the network strength maps across both the observational data, and across all of the CMIP6 model ensemble members. Here we notice that, on average, the models show the dominant regions of variability are

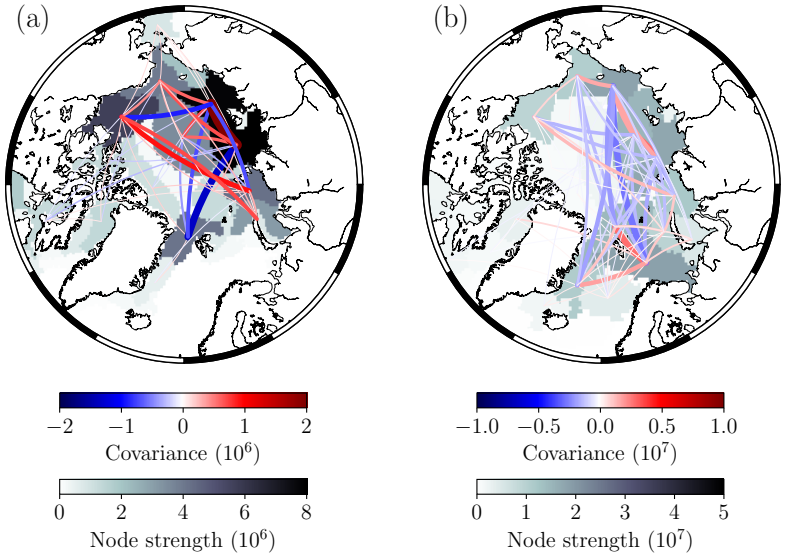


Figure 5.7: Summer sea ice concentration networks from (a) MIROC6 (ensemble: r1i1p1f1) and (b) CAMS-CSM1-0 (ensemble: r1i1p1f1). The MIROC6 model produces average ARI and D values of 0.48 and 0.66 respectively (average of metrics computed relative to NASA Team, Bootstrap and OSI-SAF networks), while the CAMS-CSM1-0 model produces average values of 0.32 and 0.29 respectively. Only links which have a corresponding p-value < 0.10 are shown here to aid visualisation.

in the East Siberian and Laptev seas, although the node strengths are somewhat over-estimated relative to the observations. Furthermore, while the observations outline the Beaufort Sea–Canada basin as the region of highest connectivity (more so than the East Siberian–Laptev seas), the models show relatively little connectivity here on average. Figure 5.8c-d also shows the percentage of variance in pan-Arctic summer sea ice area that is explained by each observational network (averaged), and the average of CMIP6 model ensembles. In this case the models generally under-estimate the importance of regions such as the Beaufort, East Siberian and Laptev seas in explaining the variance in pan-Arctic summer sea ice area, and over-estimate the percentage of variance explained in regions such as the Barents Sea and parts of the Eurasian basin. Once again, it is also noticeable how the regions of highest strength are not necessarily the ones which explain the highest percentage of variance in pan-Arctic sea ice area in the models, which suggests that the models may be over-estimating the spatial extent of the network nodes in the Eurasian seas; causing them to covary more strongly with other nodes despite having perhaps weaker physical correlation with the pan-Arctic sea ice area signal.

5.4.3 AO to sea ice teleconnection

This section now turns to an investigation of the winter AO to summer sea ice teleconnection, beginning by illustrating how the network framework can be used

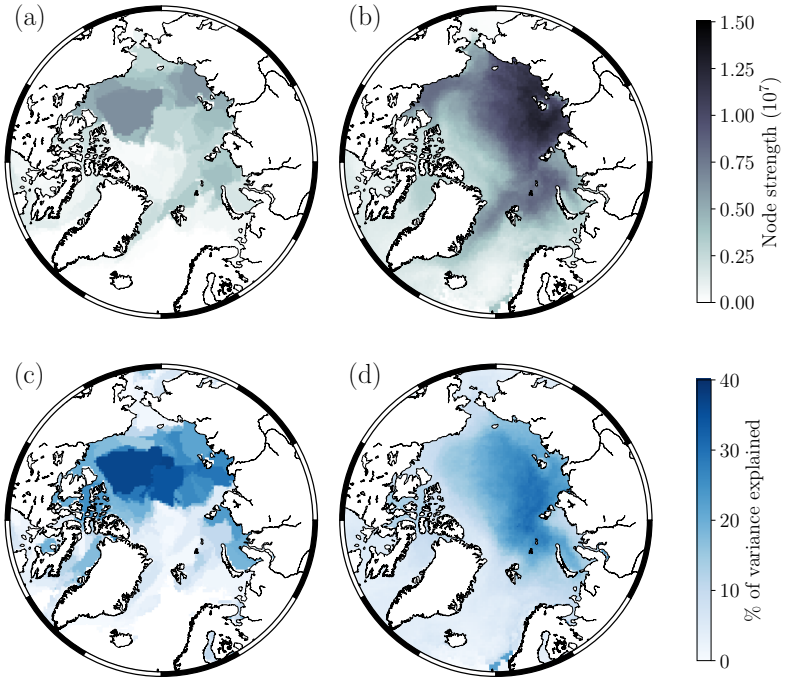


Figure 5.8: (a)-(b) The average spatial patterns of summer sea ice concentration variability (node strengths) from (a) the three observational data sets, and (b) all 74 CMIP6 model ensemble members. (c)-(d) The percentage of variance in pan-Arctic summer sea ice area explained by network nodes in (c) the three observational data sets, and (d) the average of all 74 CMIP6 model ensemble members.

to exploit this relationship in the observational and reanalysis data by effectively considering both winter sea-level pressure and summer sea ice concentration networks as individual layers within a *multi-layer* network (Boccaletti et al., 2014). After which, a brief investigation of whether this teleconnection may be changing over time is carried out, before ultimately performing the same analysis for each of the CMIP6 models. A discussion of the results in this section is then presented in section 5.5.

Observations In this section, the leading temporal mode of variability from the ERA5 sea-level pressure network is used to define the time series corresponding to the winter AO (i.e., the dashed time series in Figure 5.1b). Links are then generated between the winter AO and summer sea ice as the temporal covariance (Equation 5.2) between this AO time series, and each of the nodes of the summer sea ice concentration networks from each of the observational data sets. In Figure 5.9, the same concept as the strength maps shown previously is used, but instead each grid cell is weighted by the link weight (temporal covariance) between the AO and sea ice concentration node time series. In the first row of Figure 5.9 the link weights are computed using the entire observational period (1979–2020), where we notice a very strong anti-correlation between the winter AO and summer sea ice in the East Siberian Sea across all observational products (standardising the link weight

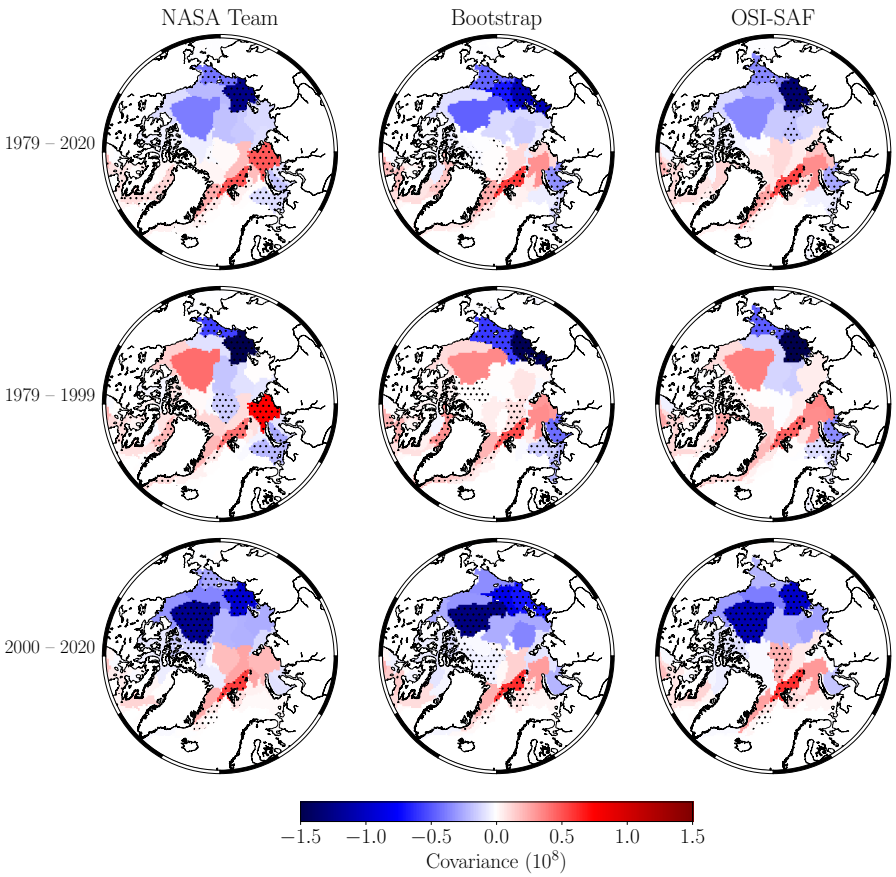


Figure 5.9: Network link weight between the DJFM ERA5 'AO node' (dashed time series from Figure 5.1b), and each of the JJAS sea ice concentration network nodes, computed between (first row) 1979–2020, (second row) 1979–1999, and (third row) 2000–2020. The columns from left to right show the corresponding maps for NASA Team, Bootstrap, and OSI-SAF data sets. Stippling denotes links with p-values less than 0.05.

for this node produces correlation coefficients of −0.65, −0.57, and −0.66 for the NASA Team, Bootstrap and OSI-SAF data sets respectively). Furthermore, all of the summer sea ice nodes in the Eurasian–Pacific sector of the Arctic exhibit varying degrees of anti-correlation with the winter AO, while the Atlantic sector shows largely positive covariance, and is particularly strong in the Fram Strait region. If the covariance is then analysed between the first half (1979–1999) and second half (2000–2020) of the observational record (second row and third row of Figure 5.9 respectively) we notice some interesting patterns. In particular, the correlation across the whole Eurasian–Pacific sector of the Arctic has been more strongly negative since the year 2000, especially within the Canada basin. The reverse of sign in the Canada basin may not be significant given the moderate degree of positive correlation between 1979–1999, however the strong negative correlation between 2000–2020 implies that positive AO winters (anomalously low sea-level pressure) now typically lead to anomalously low summer sea ice concentration anomalies across both the eastern and western Arctic, whereas previously this typically only occurred

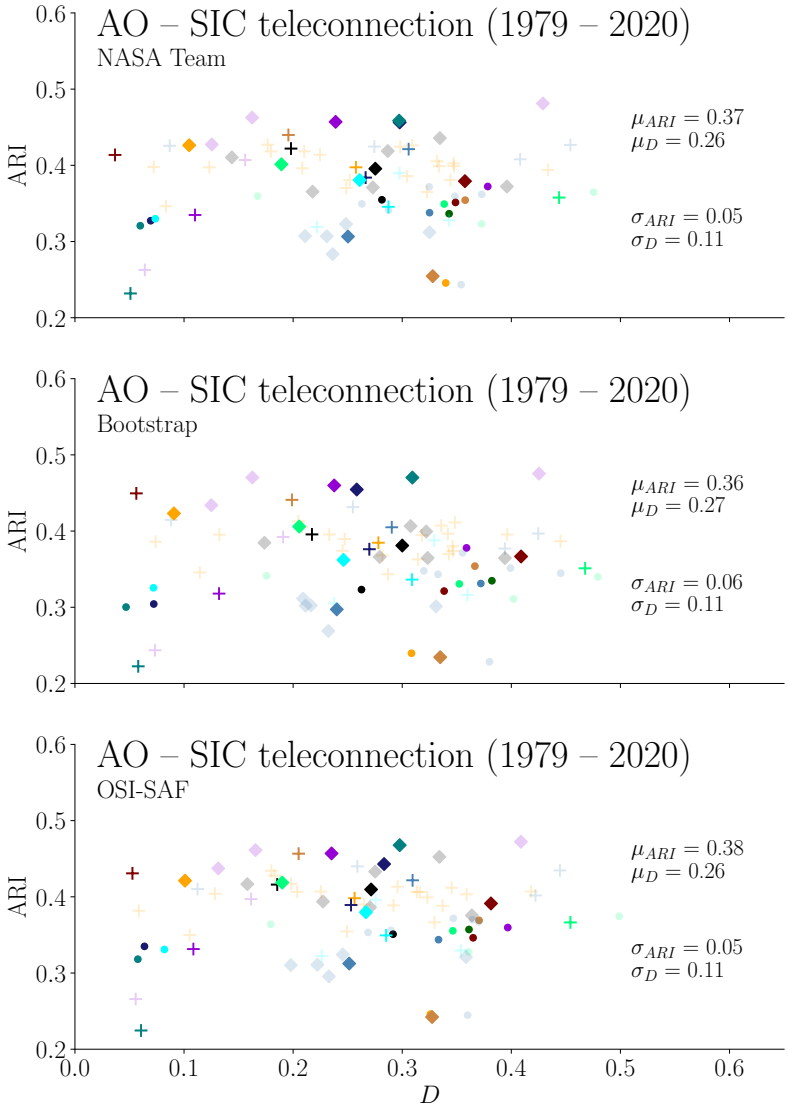


Figure 5.10: ARI and D metrics for comparing observation and CMIP6 model summer sea ice concentration networks and the winter AO to summer sea ice teleconnection, for every ensemble member for 31 different CMIP6 models (74 realisations). ARI and D are computed relative to NASA Team (top), Bootstrap (middle), and OSI-SAF (bottom) observational networks. Network distance values (D) are computed from observation and model 'link maps' as shown in Figure 5.9. The symbols and colours of each point are consistent with Figure 5.3.

in the eastern Arctic. It is also worth noting that summer sea ice in the Atlantic sector has generally remained positively correlated with the winter AO over both halves of the observational period – see section 5.5 for further discussion.

CMIP6 models For each CMIP6 model ensemble, the temporal component is extracted from the leading mode of winter sea-level pressure variability (the node with the highest strength), and the covariance-based link weight is computed with each node of its respective summer sea ice concentration network. Figure 5.10 shows

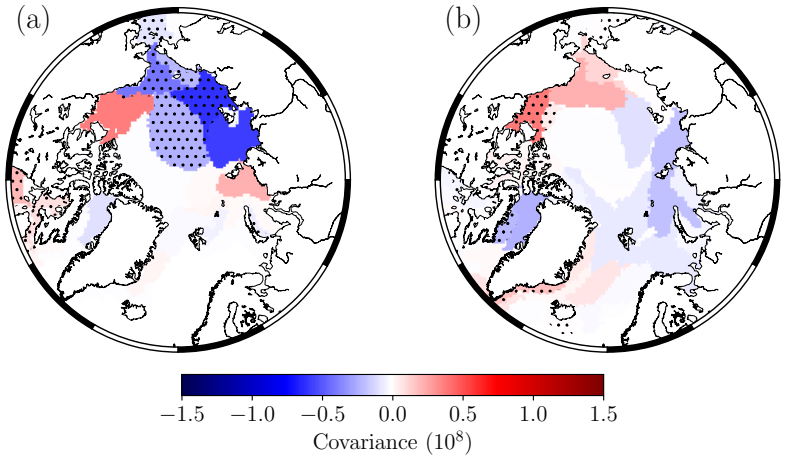


Figure 5.11: Covariance-based link weights between the winter AO node time series and each node of the summer sea ice concentration network, between 1979 and 2020 for (a) MIROC6 (ensemble: r1i1p1f1) and (b) EC-Earth3-Veg (ensemble: r4i1p1f1). The MIROC6 model produces average ARI and D values 0.47 and 0.56 respectively, while the EC-Earth3-Veg model produces values 0.26 and 0.97 respectively. Stippling denotes links with p-values less than 0.05.

an adaptation of the network comparison metrics shown in Figure 5.6. In this case, rather than computing the distance metric D as the normalised sum of the difference between observational and CMIP6 model strength maps, the corresponding 'link maps' are created for each model ensemble (i.e., the equivalent of the maps shown in Figure 5.9), and D is computed relative to the observational link maps. The ARI metric is computed as before, hence ARI values presented in Figures 5.6 and 5.10 are identical. The values reported in Figure 5.10 are for link weights computed over the entire period (1979–2020), and with an average distance values of 0.26, 0.27 and 0.26 relative to NASA Team, Bootstrap and OSI-SAF respectively, we can see that the models perform quite poorly at replicating the observed network links between the winter AO and summer sea ice – recall that for $D = 0$, the two maps are as dissimilar as a random assignment of link weights to grid cells. The equivalent plots for the periods 1979–1999 and 2000–2020 are shown in Figures C.1 and C.2 respectively. Figure 5.11 shows two examples of CMIP6 ensemble member link maps for the winter AO to summer sea ice teleconnection between 1979 and 2020. The MIROC6 model was shown in Figure 5.7 to be a network which produced relatively similar patterns of summer sea ice variability compared to the observations, and here is one of the models with the highest similarity score in terms of its AO to sea ice teleconnection, with D values of 0.43 (NASA Team), 0.43 (Bootstrap), and 0.41 (OSI-SAF). We can see that it also captures the strong negative covariance linkage with the East Siberian Sea, however it over-estimates the connection within the Laptev Sea, and does not capture the negative link with the Beaufort Sea, or strong positive link with the Fram Strait. The EC-Earth3-Veg model produces the

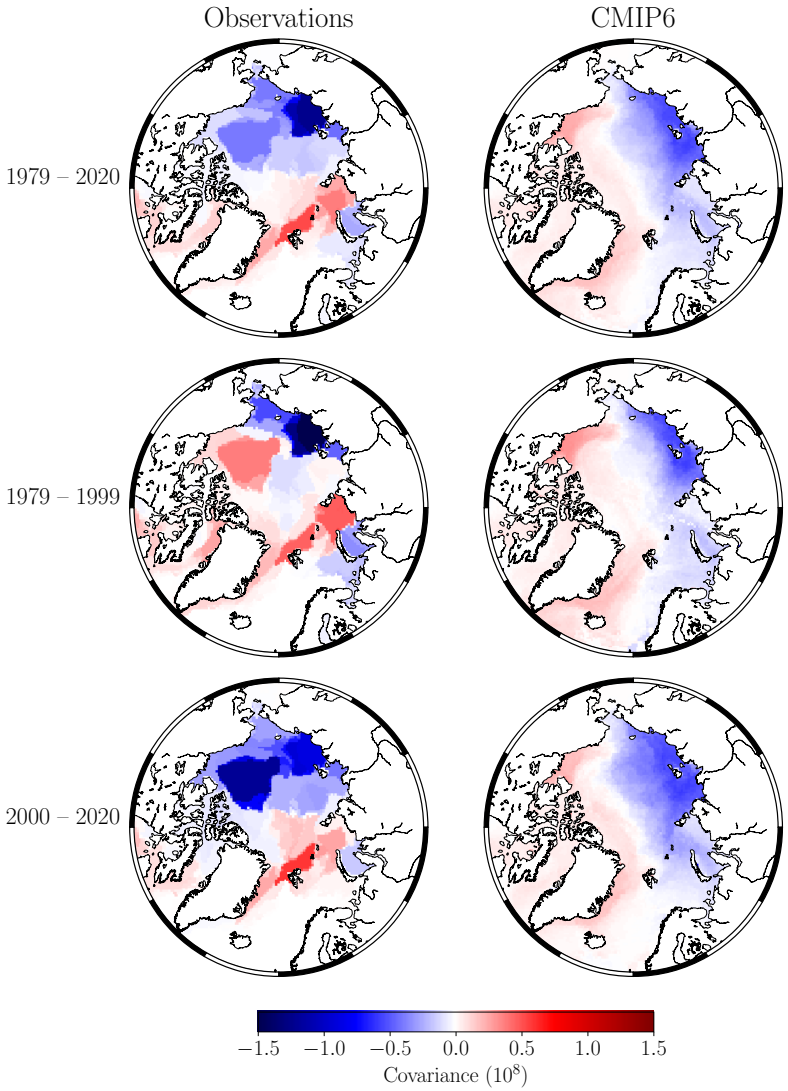


Figure 5.12: The average covariance-based link weights between the winter AO node time series and each node of the summer sea ice concentration networks, across both the observations (left column) and all CMIP6 model ensemble members (right column). Each row shows the link weights computed over different parts of the time series record: 1979–2020 (first row), 1979–1999 (second row) and 2000–2020 (third row).

lowest D scores, at 0.06 (NASA Team), 0.07 (Bootstrap), and 0.06 (OSI-SAF). This is both due to the difference in sign of many of the AO to sea ice node link weights compared to the observations (e.g., Kara and Beaufort seas), and also due to its inability to represent the similar regions of sea ice variability as the observations. Figure 5.12 shows the average teleconnection link weights between the winter AO and summer sea ice concentration node time series, for both the observations and the average of all CMIP6 ensemble members. Generally, the models agree on the sign of the network links between the winter AO and summer sea ice in the East Siberian and Laptev seas, however the magnitude of this connection is under-estimated on average. The models also do not capture the positive connection with the Kara

and Barents seas, and also do not show the same transition to an overall negative connection in the Eurasian–Pacific sectors between 2000–2020. Instead, the Canada basin region remains moderately positively correlated over the entire record.

5.5 Discussion and Conclusions

In this chapter, the complex networks methodology was used to derive spatio-temporal patterns of variability in northern-hemisphere winter sea-level pressure and Arctic summer sea ice concentration over the period 1979–2020, and to subsequently understand the spatio-temporal network connectivity between the winter Arctic Oscillation (AO) and summer sea ice cover over the same period. These patterns were analysed in both satellite observational data sets and ERA5 atmospheric reanalysis, and also from 31 of the latest generation of General Circulation Models (GCMs) participating in the most recent phase of the Coupled Model Intercomparison Project (CMIP6). Two global metrics were also introduced for comparing patterns of variability between two networks: the Adjusted Rand Index and a network distance metric. Together these allowed an assessment of how CMIP6 models perform at replicating the patterns of both winter sea-level pressure, and summer sea ice concentration variability, relative to ERA5 and the observations respectively.
Recall from section 1.1.3 how the mechanism which relates the winter AO to summer sea ice was outlined: a positive winter AO (anomalously low mean sea-level pressure) is coincident with (a) a weakening of the Beaufort Gyre, which reduces the amount of west-to-east ice advection, (b) a strengthening of the Transpolar Drift Stream (TDS), which increases ice export out of the Fram Strait, and (c) an increase in cyclonic ice motion in the Eurasian–Pacific sectors of the Arctic (Rigor et al., 2002), which causes increased ice divergence and facilitates new ice formation. Once the melt season begins, these expanses of relatively thin ice are then more susceptible to melting, thus generally leading to anomalously low sea ice area by the end of summer. Figure 5.9 showed how the observations support various aspects of this hypothesis, by the fact that the strong negative covariance in the East Siberian Sea means that following a positive winter AO, this region typically experiences anomalously low sea ice area in the summer, as the ice has undergone thinning and subsequent melting in the spring–summer. The positive covariance in the Fram strait region suggests that following positive AO winters we see an increase in sea ice in this area, which is due to positive AO events strengthening the TDS, resulting in large quantities of ice being advected towards the Atlantic sector (Rigor et al. 2002; Ricker et al. 2018). The fact that we see an overall shift towards more strongly negative covariance between 1979–1999 and 2000–2020 across the whole Eurasian–Pacific sector is likely due to the significant reductions in the thicker multi-year ice cover that have occurred in this region over recent decades (Maslanik et al. 2007; Kwok 2018). Between 1979–1999

the substantially thicker ice cover in the western Arctic was able to withstand the thinning caused by increased ice divergence from a weakened Beaufort Gyre during positive AO events, thus allowing it to survive through the melt season. More recently however, a thinner ice cover means areas of open water are more likely to form during the periods of increased ice divergence in the western Arctic, leading to the growth of new ice which is more susceptible to dramatic ice melt throughout the spring–summer.

The inability of certain CMIP6 models to accurately reflect the dominant regions of winter sea-level pressure or summer sea ice concentration variability, and their connectivity structure, could be due to a number of factors. Figure 5.5 showed that, compared to the reanalysis, on average CMIP6 models replicate the spatial patterns of winter sea-level pressure relatively well, however they generally over-estimate the magnitude of variability over the north Pacific, and under-estimate the magnitude over the Arctic Ocean and the north Atlantic; consistent with previous analysis of CMIP5 models (Jin-Qing et al. 2013; Gong et al. 2016). A recent study by Gong et al. (2019) suggested that the strong north Pacific pattern of variability in GCMs is likely due to the over-estimation of the strength of the stratospheric polar vortex (interestingly, a persistent feature of models with lower vertical resolutions in their atmospheric components), which causes enhanced coupling of atmospheric circulation between the north Pacific and north Atlantic.

In terms of summer sea ice, Figure 5.8 illustrated how CMIP6 models show discrepancies in the regions which govern summer sea ice variability, and that they generally under-estimate the contributions from regions such as the Beaufort, East Siberian and Laptev seas (Pacific sector) in explaining pan-Arctic summer sea ice area variability; and similarly over-estimate contributions from the Barents Sea and Eurasian basin (Atlantic sector). The biases in the Atlantic sector are likely related to the model's over-estimation of the sea ice extent in these regions, as in reality these regions are now largely ice-free in summer (hence the observations show little variability). Meanwhile, biases in the Pacific sector are more likely due to the poor representation of the spatial sea ice thickness distribution in models, which was previously shown to be an issue in CMIP5 models (Stroeve et al., 2014b), and also recently for a subset of CMIP6 models (Watts et al., 2021). The sea ice thickness distribution strongly determines how susceptible regions are to melting in the summer (Massonnet et al., 2018), as thicker ice effectively dampens the amount of energy transfer between the atmosphere and ocean. Figure 5.13 shows the average regional summer sea ice thickness in the Beaufort, East Siberian and Laptev seas from both PIOMAS and 25 of the CMIP6 models used in this study (thickness outputs were not available for BCC-CSM2-MR, CAMS-CSM1-0, CAS-ESM2-0, FGOALS-g3, FIO-ESM-2-0 and CanESM5-CanOE at the time of this study). On average, CMIP6 models report higher average thickness than PIOMAS in each region, which could

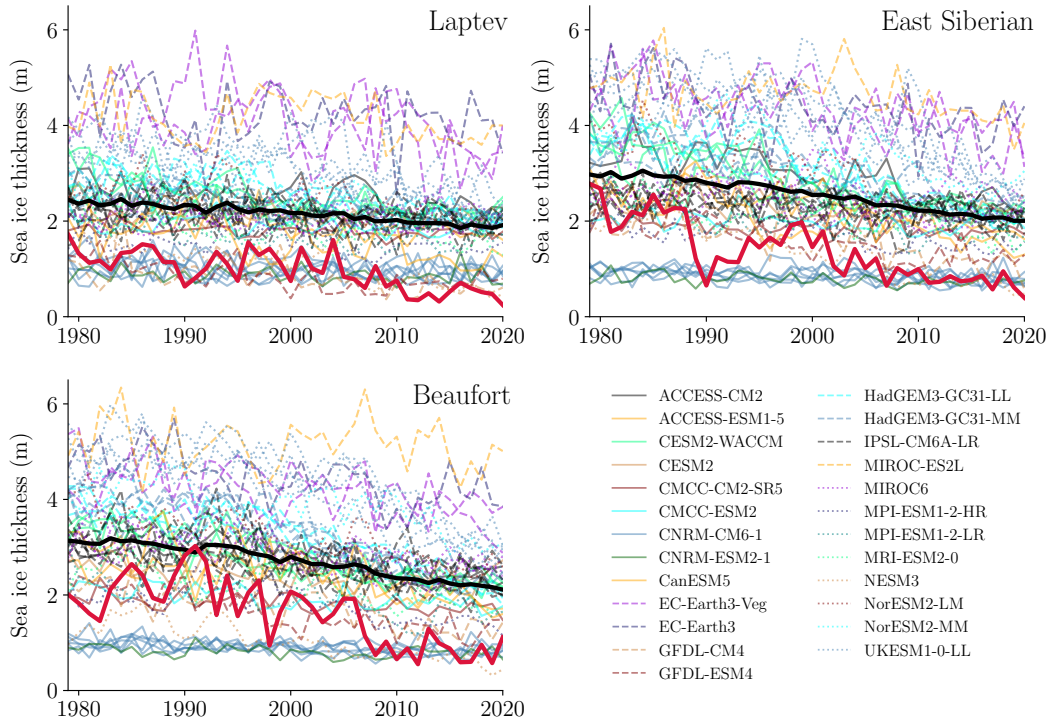


Figure 5.13: The average summer sea ice thickness from 25 CMIP6 models (49 realisations) and PIOMAS, in the Laptev Sea (top left), East Siberian Sea (top right) and Beaufort Sea (bottom). PIOMAS is given by the bold red curve, and the average of all CMIP6 models is given by the bold black curve.

explain their under-estimation of sea ice concentration variability.

The regional biases in sea ice thickness estimates from CMIP6 models could be related to several factors which determine sea ice transport and hence the ice thickness distribution, including biases in surface winds, ice rheology, and ocean heat fluxes (Stroeve et al. 2014b; Watts et al. 2021). Given then that positive winter AO events typically act to pre-condition the ice for increased melting (Williams et al., 2016), models which may perhaps reflect the spatio-temporal patterns of winter sea-level pressure variability well, may still mis-represent the effects of the winter AO on summer sea ice because the ice is too thick, and subsequently, they therefore under-estimate the amount of variability that these sea ice regions explain in terms of pan-Arctic summer sea ice area. To briefly test this hypothesis, Figure 5.14 shows the average CMIP6 winter AO to summer sea ice teleconnection (as in Figure 5.12), although this time computed only for a subset of 15 model ensembles which show the lowest average Root Mean Square Error (RMSE) in terms of their mean sea ice thickness relative to PIOMAS in the East Siberian, Laptev, and Beaufort seas. Comparing this with Figure 5.12, we notice that when only considering the models with thinner regional sea ice, the magnitude of covariance between the winter AO and summer sea ice in the East Siberian–Laptev seas is increased, and that between 1979–1999 and 2000-2020 there is evidence of the Beaufort Sea becoming

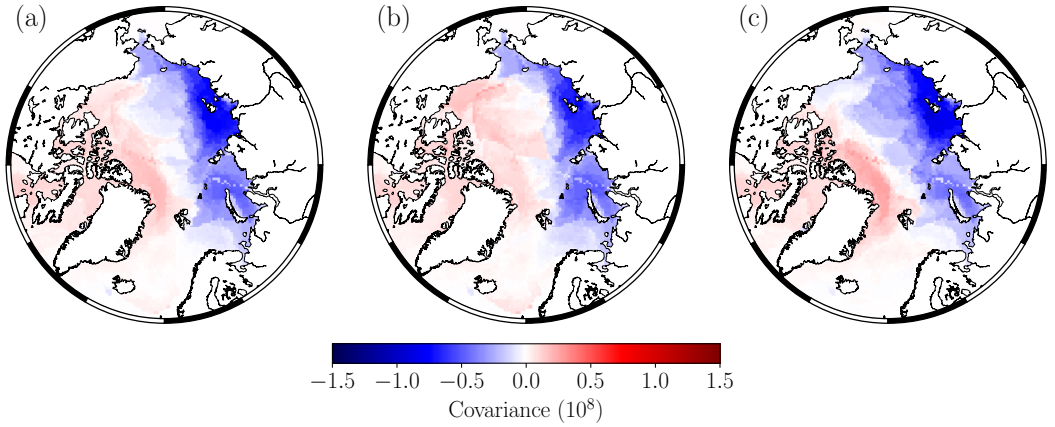


Figure 5.14: Average winter AO to summer sea ice teleconnection for 15 CMIP6 model ensembles with the lowest average Root Mean Square Error (RMSE) in mean sea ice thickness relative to PIOMAS in the East Siberian, Laptev, and Beaufort Seas. (a) Links computed between 1979–2020, (b) 1979–1999, (c) 2000–2020.

more negatively correlated, although still with a lower magnitude than shown in the observations.

Framing these results in the perspective of dynamical sea ice forecasts, the accuracy of seasonal to inter-annual sea ice predictions in GCMs ultimately hinges upon their ability to reproduce the physical processes that drive sea ice variability, and subsequently their ability to reflect the geographic regions which are responsible for explaining the overall variability in summer sea ice area. Recent years have shown the improvement in seasonal predictions brought by initialising dynamical models with observations of sea ice thickness (Chevallier & Salas-Mélia 2012; Doblas-Reyes et al. 2013; Day et al. 2014; Collow et al. 2015; Bushuk et al. 2017; Allard et al. 2018; Blockley & Peterson 2018; Schröder et al. 2019; Ono et al. 2020; Balan-Sarojini et al. 2021). Therefore, reducing sea ice thickness biases in GCMs could improve the representation of the winter AO to summer sea ice teleconnection and begin to bridge the gap between perfect-model and operational sea ice forecasts.

Chapter 6

Summary and outlook

This chapter provides a general summary of the key results presented in this thesis, as well as outlining areas for potential development related to each theme. Where applicable, a discussion is also presented on how the methodologies can be transferred to alternative applications for possible future work.

6.1 Chapter 3: Skilful sea ice forecasting

In chapter 3 a statistical framework was introduced for producing skilful seasonal forecasts of regional and pan-Arctic September sea ice extents, by combining the complex networks methodology with Gaussian process regression, through the use of a random walk graph kernel. Forecasts were made up to 3 months lead time by exploiting the predictability mechanism of sea ice persistence, whereby networks of sea ice concentration were created in the 3 preceeding months (June, July, August), and then used to predict September extents. The forecast results produced the highest skill scores of de-trended pan-Arctic September sea ice extent of any statistical or dynamical model presented in the literature at that time, at 0.53, 0.62, and 0.81 for 3-, 2-, and 1-month lead times respectively.

Section 3.5 also briefly mentioned how when increasing to 4 months lead time with this methodology, the forecast skill drops considerably – which was owed to the inability of May sea ice anomalies to persist through to September (e.g., Figure 1.11). It was also discussed how the network framework could be adapted to incorporate additional components of the climate system, as a way to potentially improve the forecast skill beyond 3 months. Figure 6.1 illustrates this by including networks of northern-hemisphere ($> 40^\circ$ N) sea-surface temperature anomalies in the forecast model. We notice how using only sea ice concentration information in May yields a forecast skill of 0.00, however by including sea-surface temperatures, this increases to

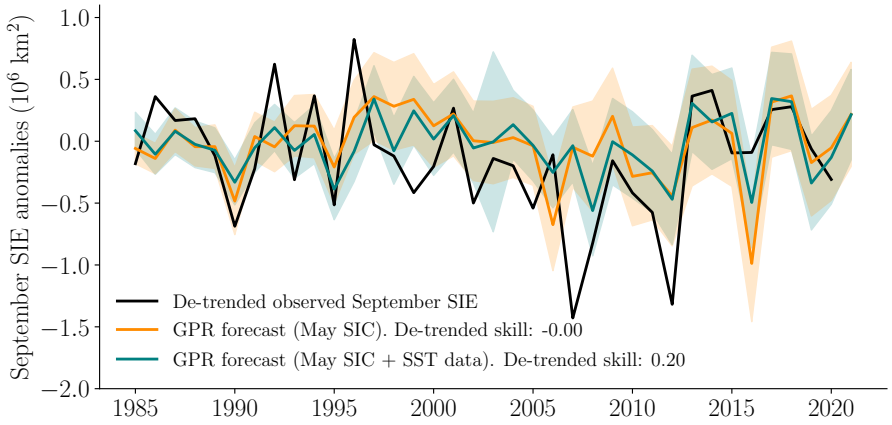


Figure 6.1: May predictions of pan-Arctic September sea ice extent anomalies based on networks of monthly mean Sea Ice Concentration (SIC) data (orange), or SIC *and* sea-surface temperature (SST) networks (teal). Shaded area reflects the predictive uncertainty ($\pm$ 1σ around the mean).

0.20. This enhanced predictability from sea-surface temperatures is likely related to the sea ice reemergence teleconnection whereby spring sea ice concentration anomalies imprint anomalies in sea-surface temperatures around the location of the ice edge, which are then able to persist through to the summer minimum.

Also outlined in chapter 3 was the need for spatial sea ice forecasts of alternative metrics, such as sea ice concentration, or sea ice probability, which would perhaps prove more useful to end-users of sea ice forecasts. In the simplest case, the Gaussian Process Regression (GPR) methodology could be used to independently predict September sea ice concentration anomalies or sea ice probability for each grid cell of e.g., a sea ice concentration time series data set. Alternatively, it would seem reasonable to expect grid cells that are within a given length scale of one another to behave in a similar way, and what would be useful is if this degree of co-variability among nearby target grid cells could be used to aid the learning process. This is known as *multi-task* learning, and its principles fit naturally into the Gaussian process framework (e.g., Bonilla et al. 2008; Ashton & Sollich 2012; Stegle et al. 2011; Rakitsch et al. 2013). Recall the standard equation for the predictive mean corresponding to the (single-task) GPR problem:

$$\bar{f}_* = \mathbf{k}_{x*}^{\mathrm{T}}(\mathbf{K}_x + \sigma^2 \boldsymbol{I})^{-1}\mathbf{y},$$

where $\mathbf{y}$ is the $n \times 1$ vector of target observations, $\mathbf{K}_x = k(\mathbf{x}, \mathbf{x})$ is the $n \times n$ covariance matrix over the training inputs, and $\mathbf{k}_{x*} = k(\mathbf{x}, \mathbf{x}_*)$ is the $n \times n_*$ covariance between the training and test inputs. In the multi-task learning approach, $\mathbf{K}_f$ is introduced as a $T \times T$ inter-task covariance matrix, which encodes the prior covariance between the T tasks that are to be predicted, as well as a $T \times T$ diagonal matrix $\boldsymbol{D}$, which contains the noise variance of each task $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_T^2)$. The equation for the multi-task

predictive mean can then be written as:

$$\bar{\boldsymbol{f}}_* = (\mathbf{K}_f \otimes \mathbf{k}_{x*})^{\mathrm{T}}(\mathbf{K}_f \otimes \mathbf{K}_x + \boldsymbol{D} \otimes \boldsymbol{I})^{-1}\mathbf{Y}, \tag{6.1}$$

where $\otimes$ is the Kronecker (tensor) product, $\mathbf{Y}$ is the $nT \times 1$ vector of the concatenated target observations for all T targets (tasks), and subsequently $\bar{\boldsymbol{f}}_*$ are the predictions of size $n_*T \times 1$. In terms of spatial sea ice forecasting, $\bar{\boldsymbol{f}}_*$ would contain the sea ice concentration anomalies of the year being forecast, for all T number of grid cells.
One of the limitations of the multi-task method is that for large T, learning becomes increasingly slow due to the need to invert a $nT \times nT$ matrix; which has run time complexity $\mathcal{O}(n^3T^3)$, although this burden can be mitigated somewhat by diagonalising all covariance matrices and exploiting certain 'tricks' related to the Kronecker product (Stegle et al. 2011; Rakitsch et al. 2013), so the run time complexity reduces to $\mathcal{O}(n^3 + T^3)$.
A benefit of the multi-task approach is that the definition of the inter-task covariance matrix can be very flexible, so long as it is positive semi-definite. For example, we may wish to incorporate spatially varying length scale information into $\mathbf{K}_f$, as we might expect sea ice concentration anomalies in some regions of the Arctic to be correlated over larger or shorter distances than in other regions (e.g., shorter over first-year ice regions and longer over multi-year ice regions, as was the case with radar freeboards in chapter 4). One way to achieve this could be through implementing one of the many non-stationary covariance functions outlined in Paciorek & Schervish (2005), for example the squared exponential function:

$$\mathrm{K}_f(\mathbf{z}_p, \mathbf{z}_q) = \sigma_f^2 \left|\boldsymbol{\Sigma}_p\right|^{1/4} \left|\boldsymbol{\Sigma}_q\right|^{1/4} \left|\frac{\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q}{2}\right|^{-1/2} \exp(-Q_{pq}), \tag{6.2}$$

where

$$Q_{pq} = (\mathbf{z}_p - \mathbf{z}_q)^{\mathrm{T}} \left(\frac{\boldsymbol{\Sigma}_p + \boldsymbol{\Sigma}_q}{2}\right)^{-1} (\mathbf{z}_p - \mathbf{z}_q). \tag{6.3}$$

Here $\mathbf{z}_p$ is a vector of length 2 containing the zonal and meridional grid coordinates of grid cell p, and $\boldsymbol{\Sigma}_p$ is a 2×2 diagonal matrix containing the zonal and meridional correlation length scale parameters ℓ_x and ℓ_y (which can be optimised through e.g., the empirical Bayes approach). Figure 6.2 shows an example non-stationary squared exponential inter-task covariance matrix, illustrating how the length scale of correlation varies depending on the location of each grid cell, for 4 different grid cell locations within the Arctic. Figure 6.3 then implements this inter-task covariance structure in a multi-task GPR forecast (based on the efficient approach outlined in Rakitsch et al. 2013) of September sea ice concentration for the year 2011, at 1-month lead time (using an August sea ice concentration network as the predictor, see chapter 3). Generally the forecast follows the observed ice edge contour, although with largest over-predictions around the Severnaya Zemlya group of islands in the

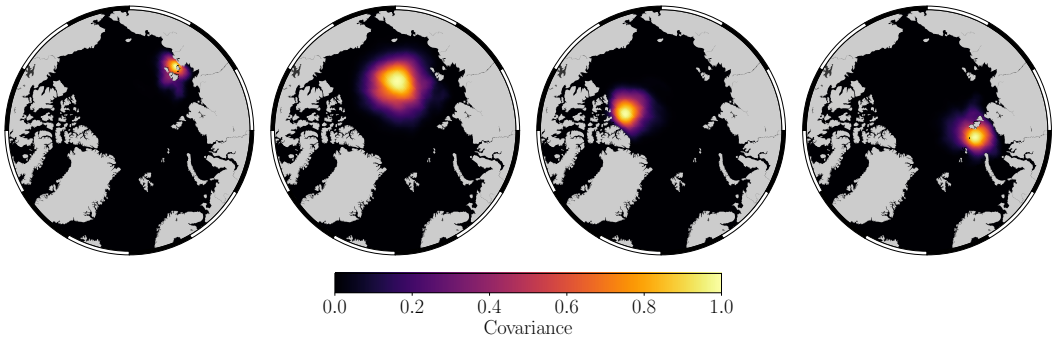


Figure 6.2: Example of a non-stationary squared exponential inter-task (sea ice concentration) covariance matrix for 4 different grid cell locations within the Arctic. Notice how the length scale of correlation varies depending on the location of each grid cell.

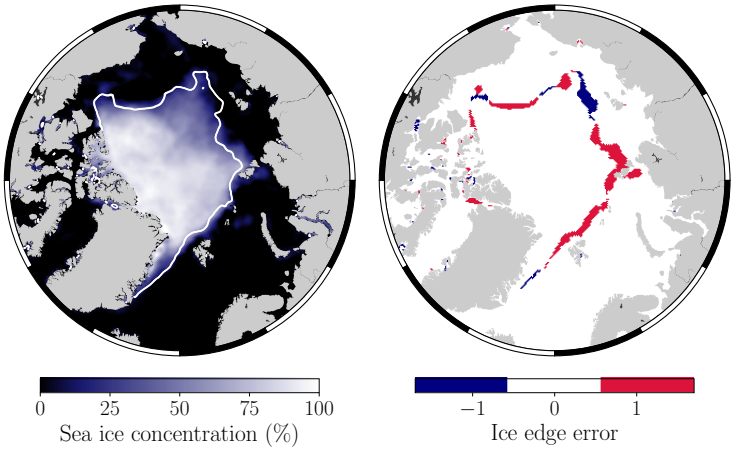


Figure 6.3: (Left) Forecast of 2011 September sea ice concentration at 1-month lead time using multi-task GPR with a non-stationary squared exponential inter-task covariance matrix. The white contour is the observed sea ice edge (15% concentration) from September 2011. (Right) Forecast error in the location of the ice edge. Red signifies over-estimation relative to the observations, and vice versa for blue.

western Laptev sea. Goessling et al. (2016) introduced a metric for quantifying the error in spatial sea ice forecasts, called the Integrated Ice Edge Error (IIEE). This metric corresponds to the total area for which the sea ice edge is either over- or under-predicted (i.e., the total area of red and blue pixels in Figure 6.3 right), and for this particular forecast corresponds to 0.75×10^6 km^2. Using this method operationally however would require further forecasts of additional years and lead times to evaluate its general performance beyond the example given here. As well as this, a cost-benefit analysis would be required to establish whether any uplift achieved by the multi-task approach over iterative predictions of each grid cell, outweighs the computational expense of the method – an iterative GPR approach would have run time complexity $\mathcal{O}(n^3T)$, whereas already mentioned in the multi-task case it is $\mathcal{O}(n^3 + T^3)$.

More recently, deep learning approaches such as convolutional neural networks have

shown to be successful in terms of spatial summer sea ice forecasts, and could be explored as a potential avenue for future work. A recent study by Andersson et al. (2021) for example trained a convolutional neural network using both climate model simulations and observational data in a binary classification forecast of the sea ice edge, with relatively successful predictions up to 6 months lead time. In this particular model however, predictions are not probabilistic, and so a way to improve this methodology could be in the form of Bayesian neural networks which places prior distributions over the various parameters in each network layer (see e.g., section 5.7 of Bishop 2006).

6.2 Chapter 4: Data fusion and optimal interpolation of radar freeboards

Chapter 4 showed how the GPR framework can be used to generate daily pan-Arctic radar freeboard and uncertainty estimates, through the optimal interpolation of CryoSat-2 (CS2), Sentinel-3A (S3A), and Sentinel-3B (S3B) observational data sets. Furthermore, by using the empirical Bayes approach, daily pan-Arctic fields of the radar freeboard correlation length scales could also be derived. Analysis of training and prediction errors showed that the interpolated freeboards (CS2S3) were consistent with CS2, S3A and S3B, to within 1 mm, and that the average prediction errors based on cross-validation across the 2018–2019 winter season were generally ≤ 4 mm. Subsequent analysis of regional daily radar freeboard evolution showed the uplift in temporal variability brought by moving from a monthly-averaged to a daily product, and further analysis with daily ERA5 snowfall showed strong positive correlation between freeboard anomalies and snowfall over first-year ice; challenging the conventional assumptions of full Ku-band radar penetration through snow.

One of the major sources of uncertainty in derived sea ice thickness estimates from satellite altimeters is the depth of the snow layer that rests atop the sea ice. Many current operational sea ice thickness products use a simple climatology snow model based on spatially-interpolated Soviet drifting stations, collected over multi-year ice between 1954–1991 (Warren et al., 1999). Recent studies have made concerted efforts to produce more representative estimates of the snow depth distribution across the Arctic, based on model studies (Petty et al. 2018; Liston et al. 2020; Stroeve et al. 2020a), and through the differencing of observations from dual-frequency radar altimeters (Lawrence et al., 2018), or similarly laser and radar altimeters (Kwok et al., 2020). Recall that laser altimeters such as ICESat-2 measure total freeboard, which is the height of the top of the snow surface above the adjacent ocean, whereas radar freeboard measures the height of the radar scattering horizon

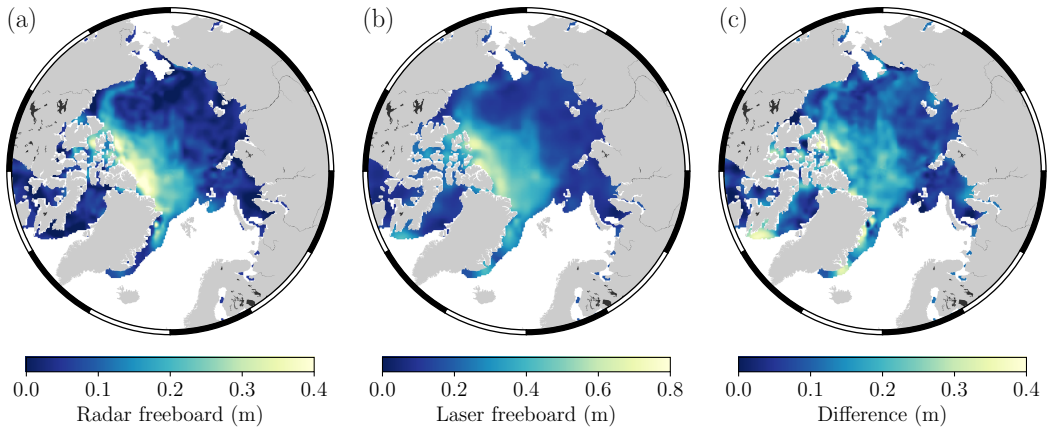


Figure 6.4: Example of using the Gaussian process regression methodology to produce pan-Arctic fields of (a) radar freeboard (CS2S3), (b) laser freeboard (ICESat-2), and (c) the difference (proxy for snow depth), for one day (1st of December 2018).

above the adjacent ocean; the location of which may vary considerably depending on the dielectric properties of the snow (Willatt et al., 2011). In any case, under cold and dry conditions the radar scattering horizon corresponds to somewhere close to the snow-ice interface, and so differencing laser and radar freeboards can give a rough proxy for snow depth. In chapter 4, daily pan-Arctic fields of radar freeboard were generated, however this approach could also be used to generate daily fields of total (snow) freeboard from e.g., ICESat-2. Differencing these two fields could then provide an initial proxy for daily pan-Arctic snow depth on sea ice. An example of this is shown for one day in Figure 6.4 by using the same methodology outlined in chapter 4 to optimally combine 9 days of ICESat-2 data. While further consideration of the snow properties are required for estimating the *true* snow depth (which has been one of the focal points of the recent 2019–2020 MOSAiC expedition, e.g., Stroeve et al. 2020b), this approach could be a considerable step forward in improving sea ice thickness estimates from satellite altimeters.

A significant advancement to the methodology outlined in chapter 4 would come in the form of computational speed-ups when generating each daily field. The current implementation requires the exploitation of high-performance computers with parallel processing capabilities, however it was found during testing that optimisation problems (e.g., maximising the log marginal likelihood) often face bottleneck issues during parallel processing, which results in longer run times. A simple way to alleviate the computational burden of this method is to use fewer data points during the model training – recall that GPR has run time complexity $\mathcal{O}(n^3)$. A first attempt of this was outlined in chapter 4 by generating the predictions iteratively and only using training data within a distance of 300 km, however this could be taken further. A recent study by Landy et al. (2021) used an approach whereby an arbitrary subset

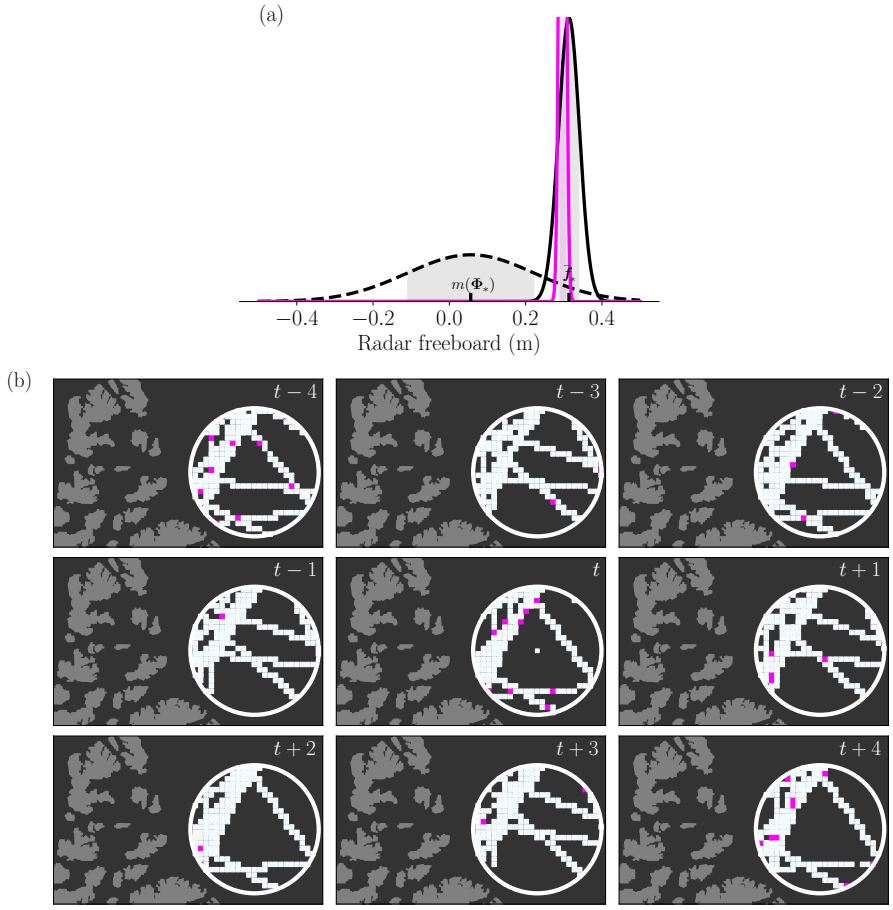


Figure 6.5: As in Figure 4.2, although also showing the prediction generated via an RVM. (a) The posterior distribution of radar freeboard generated by GPR (black, solid) and RVM (magenta). (b) The $n = 1855$ data points used to inform on predictions in the GPR model (pale blue), while the magenta points are the 47 'relevant vectors' identified via an RVM.

of N training points were chosen based on those which exhibit the largest prior covariance between training and test inputs (e.g., the largest N correlations in Figure 4.2b). An alternative, and perhaps more principled approach to this could be through the use of Relevance Vector Machines (RVMs) for regression (Tipping 2001). RVMs are probabilistic linear models which follow the same principles as the Bayesian linear regression model outlined in section 2.2.1, although result in sparse solutions such that only a subset of the training data are used to inform on predictions. In this type of model, function values $f(\mathbf{\Psi})$ correspond to linear combinations of a given kernel function (or equally a set of fixed basis functions as in section 2.2.1):

$$f(\mathbf{\Psi}) = \sum_{i=1}^{n} \beta_i k(\mathbf{\Psi}, \mathbf{\Psi}_i) + \beta_0. \tag{6.4}$$

In the context of our GPR model in chapter 4, $k(\mathbf{\Psi}, \mathbf{\Psi})$ could either be the $n \times n$ covariance matrix between the training inputs $k(\mathbf{\Phi}, \mathbf{\Phi})$, or the $n_* \times n$ covariance

between the training and test inputs $k(\boldsymbol{\Phi}_*, \boldsymbol{\Phi})$, and then $\boldsymbol{\beta}$ is a vector of length n containing weights for each column of $k(\boldsymbol{\Psi}, \boldsymbol{\Psi})$. An RVM then typically assigns a zero-mean prior distribution over the weight vector:

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}) = \prod_{i=1}^{n} \mathcal{N}(0, \gamma_i^{-1}), \tag{6.5}$$

which, when combined with the likelihood over the outputs (Equation 2.14), results in the posterior distribution over the weights, with mean $\tilde{\boldsymbol{\mu}}$ and covariance $\tilde{\boldsymbol{\Sigma}}$ given by:

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \left(\mathbf{K}\boldsymbol{V}^{-1}\mathbf{K}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1}\right)^{-1}\mathbf{K}\boldsymbol{V}^{-1}\mathbf{y} \\ \tilde{\boldsymbol{\Sigma}} &= (\mathbf{K}\boldsymbol{V}^{-1}\mathbf{K}^{\mathrm{T}} + \boldsymbol{\Sigma}^{-1})^{-1}. \end{aligned} \tag{6.6}$$

Here $\mathbf{K} = k(\boldsymbol{\Phi}, \boldsymbol{\Phi})$ is the covariance matrix over the training inputs, $\boldsymbol{V} = \sigma^2\boldsymbol{I}$, and $\boldsymbol{\Sigma}$ is a diagonal matrix with entries $(\gamma_1^{-1}, \gamma_2^{-1}, \ldots, \gamma_n^{-1})$. The difference between the solution above and the one presented in the discussion on Bayesian linear regression (i.e., Equation 2.22) is that in the RVM case, each parameter β_i is assigned its own corresponding prior variance term γ_i, which is treated as a hyperparameter. Through maximising the log marginal likelihood function with respect to these hyperparameters, it is generally found that the majority of the variance terms are infinite (Tipping & Faul, 2003), which means that the corresponding entries of $\tilde{\boldsymbol{\mu}}$ will be zero (i.e., the posterior is infinitely peaked around zero), thus leading to a sparse model. Figure 6.5 revisits the example of generating a single prediction of radar freeboard for one pixel, as was shown in Figure 4.2. In this example the number of training points is $n = 1855$, which means that the matrix $\mathbf{K} \in \mathbb{R}^{1855\times1855}$ must be inverted in order to generate the GPR predictive distribution. By taking an RVM approach however, only 47 (out of 1855) of the columns of $\mathbf{K}$ are identified as the 'relevant vectors', and hence only 47 of the entries of $\tilde{\boldsymbol{\mu}}$ contain values greater than zero. This means that, at most, a matrix of size 47×47 must be inverted in order to generate $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$, at which point the predictive mean and covariance are given as $k(\boldsymbol{\Phi}_*, \boldsymbol{\Phi})\tilde{\boldsymbol{\mu}}$ and $k(\boldsymbol{\Phi}_*, \boldsymbol{\Phi})\tilde{\boldsymbol{\Sigma}}k(\boldsymbol{\Phi}, \boldsymbol{\Phi}_*)$ respectively. It is worth noting that while the predictive mean generated via RVM and GPR are relatively comparable (Figure 6.5a) at 0.30 and 0.31 m respectively, the predictive variance in the RVM approach is under-estimated relative to GPR. This is due to the particularly counter-intuitive property of the standard RVM method, whereby as the distance from the relevant vectors to the test input location increases, the predictive variance decreases. This problem has however been addressed in previous studies, e.g., Candela & Hansen (2004); Rasmussen & Candela (2005).

6.3 Chapter 5: Climate connectivity in general circulation models

In chapter 5, the complex networks methodology was used to compare the spatio-temporal patterns of variability in winter sea-level pressure and summer sea ice concentration between observations and outputs from 31 of the latest generation of coupled climate models participating in phase 6 of the World Climate Research Programme Coupled Model Intercomparison Project (CMIP6). As well as this, an analysis of the representation of the winter Arctic Oscillation (AO) to summer sea ice teleconnection was also carried out for both observations and models using two specific global network comparison networks; the Adjusted Rand Index and a network distance metric. It was generally found that CMIP6 models reflect the spatio-temporal pattern of the AO relatively well, although over-estimate the magnitude of sea-level pressure variability over the north-western Pacific Ocean, and under-estimate the magnitude over the north Atlantic sector. Furthermore, models and observations generally showed disagreement in terms of which regions are responsible for explaining the largest percentage of variance in pan-Arctic summer sea ice area, with the observations showing largely the Canada basin region and the East Siberian Sea, and the models on average showing considerable contributions from parts of the Eurasian basin. It was hypothesised that this discrepancy was both due to the poor representation of the sea ice thickness distribution and sea ice extent in the models. Finally, the observational data showed that (likely due to ice thinning) positive winter AO events between 2000–2020 now lead to anomalously low summer sea ice concentration in both the eastern and western sectors of the Arctic Ocean, whereas between 1979–1999, this typically only occurred in the eastern Arctic. CMIP6 models however did not record this change on average, and generally under-estimated the magnitude of covariance between the winter AO and summer sea ice in the Eurasian seas, and showed moderately positive correlations in the western Arctic across the entire observational record. This was also attributed to the models over-estimation of sea ice thickness in these regions.

The statistical relationships established between network nodes in chapter 5, and subsequently the regional responses of summer Arctic sea ice to different phases of the winter AO, were based purely on linear covariance. Although these were computed as lagged correlations (winter $\rightarrow$ summer), this still only permits us to make statements about the *association* between these two fields. A way to develop the methodology could be through adopting a causal inference approach in order to trace the effects that the winter AO has on the various intermediate physical processes that eventually lead to the regional summer sea ice responses seen in chapter 5, and subsequently better isolate the point at which this teleconnection breaks down across CMIP6

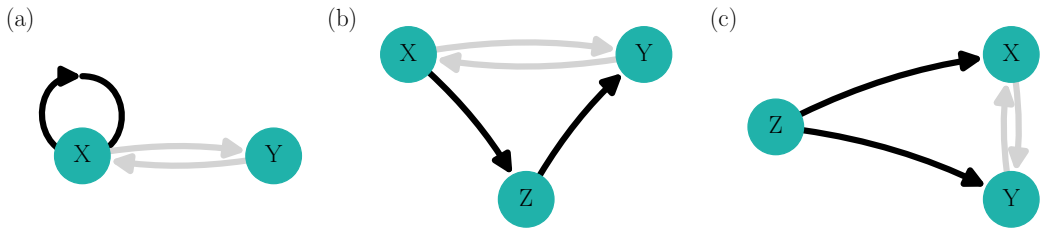


Figure 6.6: Examples of how spurious correlations can arise between two independent variables X and Y. Black arrows signify causal links while grey arrows signify spurious correlation. (a) Correlation between X and Y caused by large autocorrelation of X. (b) Indirect correlation between X and Y via a mediator variable Z. (c) Z is a confounder (common driver) of both X and Y. Plot adapted from Kretschmer et al. (2016).

models.

Causal inference presents a principled framework for making statements about physical drivers within the climate system, and is an approach which has gained much recent attention in time series analysis (Runge et al. 2014, 2019a,b); already with applications in polar climate studies (Kretschmer et al., 2016). The movement towards establishing causality typically begins with understanding the ways in which spurious correlations between two independent physical processes X and Y can arise, through the use of causal diagrams (e.g., Pearl 1995). Figure 6.6 highlights three examples of this, although perhaps most significant to this study is Figure 6.6b which shows how spurious correlations can arise due to indirect links to an intermediate process (or processes) Z. In the context of the AO to sea ice teleconnection, a number of these intermediate processes were summarised in sections 1.1.3 and 5.5. The aim of causal inference techniques is then to establish whether X and Y are *conditionally independent* (i.e., not directly related; Dawid 1979), given Z; which can be written as $\mathrm{X} \perp\!\!\!\perp \mathrm{Y} \mid \mathrm{Z}$. In terms of probabilities, this implies that the joint distribution of X and Y given Z, $p(\mathrm{X}, \mathrm{Y}|Z)$, is completely defined by the product of their marginal distributions (conditioned on Z), i.e., $p(\mathrm{X}, \mathrm{Y}|Z) = p(\mathrm{X}|Z)p(\mathrm{Y}|Z)$. When considering this framework in the context of Figures 6.6b and 6.6c, we notice that the statement $\mathrm{X} \perp\!\!\!\perp \mathrm{Y} \mid \mathrm{Z}$ is true for both cases. For example in Figure 6.6b, the joint probability of X and Y given Z is expressed as $p(\mathrm{X}, \mathrm{Y}|\mathrm{Z}) = p(\mathrm{X})p(\mathrm{Z}|\mathrm{X})p(\mathrm{Y}|\mathrm{Z})/p(\mathrm{Z})$, where the terms $p(\mathrm{X})p(\mathrm{Z}|\mathrm{X})/p(\mathrm{Z})$ then simplify to $p(\mathrm{X}|\mathrm{Z})$ through Bayes' Law, and hence we arrive at conditional independence $p(\mathrm{X}, \mathrm{Y}|\mathrm{Z}) = p(\mathrm{X}|\mathrm{Z})p(\mathrm{Y}|\mathrm{Z})$.

In practice, establishing (linear) conditional independence is typically achieved via partial correlation tests (Spirtes et al., 2000), which in the simplest case corresponds to computing the linear Pearson correlation between two sets of residuals from ordinary least squares regression; those generated from regressing X (outputs) against Z (inputs), and those generated from regressing Y (outputs) against Z (inputs). Recent studies by Runge et al. (2019a) have extended this approach to large multivariate time series data sets and have shown success in isolating climate drivers related

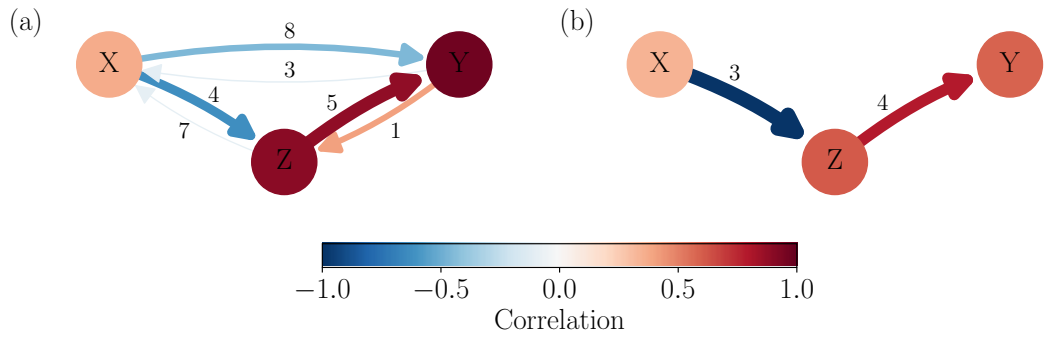


Figure 6.7: Identifying causal links through a causal discovery algorithm (TiGraMITE). (a) Links generated by simple lagged correlations (no causal principles implemented). The colour of each node corresponds to its autocorrelation at a time lag of 1. Otherwise the lag associated with each connection is given by the number adjacent to each link. (b) Links generated from causal inference.

to teleconnections such as the tropospheric Walker cell (Runge et al., 2014). As a motivating example, a synthetic time series case is presented here which is analogous to the AO to sea ice teleconnection, and is based on the causal inference algorithm TiGraMITE[1] (Runge et al., 2019a).

Consider the case where **X**, **Y** and **Z** are time series data such that:

$$\begin{aligned} \mathrm{X}_t &= 0.4\mathrm{X}_{t-1} + \varepsilon_t^{\mathrm{X}} \\ \mathrm{Z}_t &= 0.8\mathrm{Z}_{t-1} - 5\mathrm{X}_{t-3} + \varepsilon_t^{\mathrm{Z}} \\ \mathrm{Y}_t &= 0.7\mathrm{Y}_{t-1} + 1.2\mathrm{Z}_{t-4} + \varepsilon_t^{\mathrm{Y}}, \end{aligned}$$

where $t = \{1, 2, \ldots, n\}$ and $n = 504$ – analogous to monthly sampled data for all years between 1979 and 2020. In this case, the state of the variable **X** at time t is a linear function of its state at time $t-1$, plus random noise $\varepsilon_t^{\mathrm{X}}$. This is also the case for **Z** and **Y**, although **Z** is also a function of **X** at a time lag of 3 (months) and **Y** a function of **Z** at a time lag of 4 (months). Graphically, these data could be represented by the causal diagrams in Figures 6.6a and 6.6b (each time series has an element of autocorrelation), and in the context of the AO to sea ice teleconnection, the time series **X** might correspond to the AO, **Y** to sea ice concentration at a particular node of a complex network, and **Z** an intermediate process which connects the two. Note also that in the definitions above **Y** is independent of **X**, although this is assumed unknown a-priori.

By computing simple lagged correlations between each of the variables independently (Figure 6.7a) we find significant correlations (p-values < 0.05) between all variables, and in both directions. Furthermore, we can see how this approach leads to spurious

[1]Time Series Graph Based Measures of Information Transfer https://github.com/jakobrunge/tigramite

correlations due to both autocorrelation and the mediator process $\mathbf{Z}$, and also incorrect estimation of the lag times from $\mathbf{X} \to \mathbf{Y}$ and $\mathbf{Z} \to \mathbf{Y}$.

The causal inference algorithm TiGraMITE establishes linear conditional independence between two variables via iterative partial correlation measures, and adjusts for inflated and/or deflated correlations due to autocorrelation and/or dimensionality respectively (with only 3 processes here, dimensionality may not be a significant issue, although in the case of multi-layer climate networks there may be tens or hundreds of processes to consider. See Runge et al. 2019a for further details). In the example here, conditional *dependence* between e.g., $\mathbf{X}$ at a time lag $t - \tau$ and $\mathbf{Y}$ at time t corresponds to $\mathrm{X}_{t-\tau} \not\perp\!\!\!\perp Y_t \mid \mathbf{P}_t \setminus \{\mathrm{X}_{t-\tau}\}$, which in plain terms would mean $\mathrm{X}_{t-\tau}$ drives Y_t when controlling for (conditioning on) the common past of all variables $\mathbf{P}_t = \{\mathrm{X}_{t-1}, \mathrm{Y}_{t-1}, \mathrm{Z}_{t-1}, \mathrm{X}_{t-2}, \ldots, \mathrm{Y}_{t-\tau}, \mathrm{Z}_{t-\tau}\}$, excluding $\mathrm{X}_{t-\tau}$. This is achieved through finding which time lag produces significant partial correlations $\rho(\mathrm{X}_{t-\tau}, \mathrm{Y}_t \mid \mathbf{P}_t \setminus \{\mathrm{X}_{t-\tau}\}) \neq 0$, for all lags up to a maximum $\mathrm{X}_{t-\tau_{\max}}$ (in this case $\tau_{\max} = 12$). If no significant correlations are found then the two variables are said to be conditionally independent. Note that both $\mathbf{X}$ and $\mathbf{Y}$ appear in the conditioning set $\mathbf{P}_t$ here in order to account for autocorrelation effects which may cause spurious correlations between $\mathrm{X}_{t-\tau}$ and Y_t. Furthermore, in reality $\mathbf{P}_t$ would only contain a subset of the whole set of variables and lags, which are called the 'parents' of the process Y_t (e.g., Spirtes & Glymour 1991), although for simplicity let us continue with the definition of $\mathbf{P}_t$ as a generic conditioning set of past variables.

The result of implementing this algorithm to the synthetic data above is shown in Figure 6.7b, where the p-values associated with each link are < 0.05. Noticeably, the spurious correlation $\mathbf{X} \to \mathbf{Y}$ is removed, as well as the other spurious links $\mathbf{Y} \to \mathbf{Z}$ and $\mathbf{Z} \to \mathbf{X}$. Furthermore the inflated correlation between $\mathbf{Z}$ and $\mathbf{Y}$ in 6.7a (due to autocorrelation) is corrected for in 6.7b, and the correct time lag has been identified for both links.

To implement this procedure with real observational data would require further consideration of the different types of variables to be analysed, as well as any required pre-processing of the data, choice of temporal frequency (e.g., daily, weekly, monthly, or even seasonal), and assumptions about the interactions between physical processes, such as linearity *et cetera* (Kretschmer et al., 2016). Even so, the synthetic example outlined here should provide as a motivating example of ways to move the complex networks methodology forward.

Appendix A

Chapter 2 supplementary material

This appendix contains figures and which are supplementary to chapter 2, and also a brief outline of the cross-validation method.

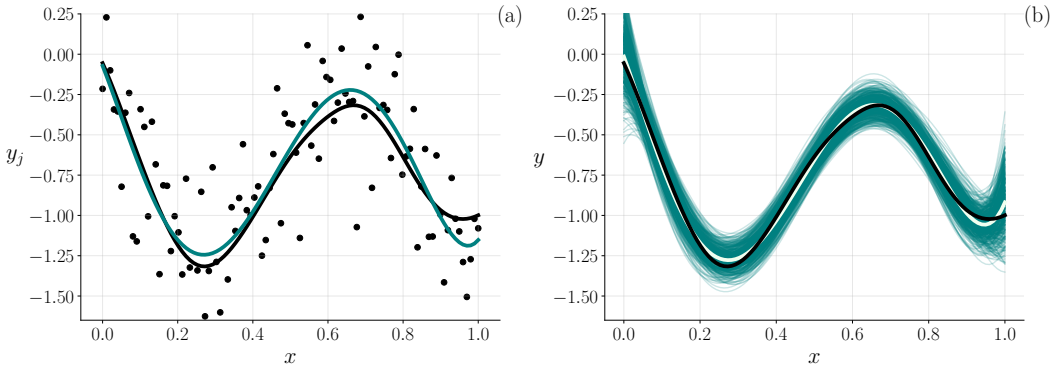


Figure A.1: Synthetic example of 1-D polynomial OLS regression, as in Figure 2.3, except that the number of observations is increased from $n = 10$ to $n = 100$. Comparing both figures, both the bias and variance of the estimated function values are reduced by using more observations, for a given model complexity.

A.1 *K*-fold cross-validation

Generating predictions with the Ordinary Least Squares (OLS) polynomial basis function model in section 2.2.1 required a decision on the optimal order of the polynomial basis functions (and hence number of model parameters), and similarly for the value of the regularisation coefficient λ in the case of Ridge regression.

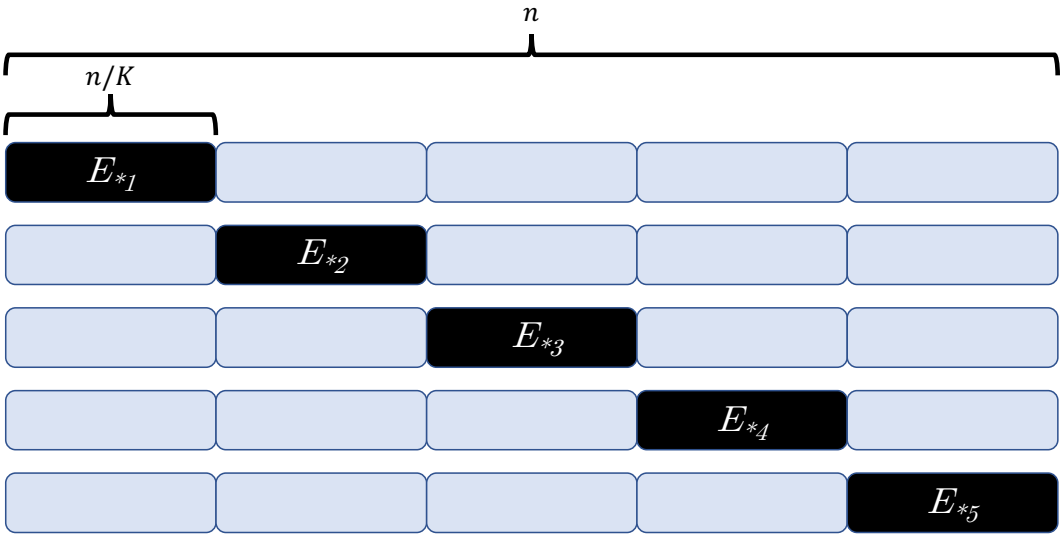


Figure A.2: K-fold cross-validation where $K = 5$. The total number of training points is given by n, and the prediction error at each iteration step is given by E_{*i}, where $i = \{1, \ldots, K\}$. Blue cells are the data used for training and the black cells are those held aside to evaluate the predictions.

In many cases it will be necessary to compare a range of models with perhaps a variety of complexity parameters and/or different types of basis functions, where ultimately the one which produces the best predictions is desired. This is often achieved by separating the data into disjoint training and validation sets whereby the various different models are trained on a subset of the overall data and predictions are validated against the withheld set. Subsequently, the model with the lowest prediction error is then chosen. In this approach however, only a subset of the data are used to train the model, and if the number of points used for validation is small, then the variance of the predictions will be large (Rasmussen & Williams, 2006). A way to circumvent this is through K-fold cross-validation (Stone, 1974). In this case the available training data are split into K number of equal sized sets, where the model is then trained on data from $K - 1$ of the total sets, and then predictions are evaluated on the final set. This is repeated K number of times, where at each iteration the predictions are evaluated on a different subset (see Figure A.2). At each step the sum of the squared error between the predictions and the withheld set is evaluated E_{*i}, after which an average of all prediction errors is taken to represent the cross-validation score of that particular model. The most extreme case of this approach is where $K = n$, which is then referred to as leave-one-out cross-validation, however depending on the model this may be computationally prohibitive as the model must then be run n number of times. The Bayesian approach to model selection outlined in section 2.2.2 is an alternative framework for estimating the free (hyper)parameters of a model which utilises all of the available training data and can be evaluated without running multiple iterations of the model.

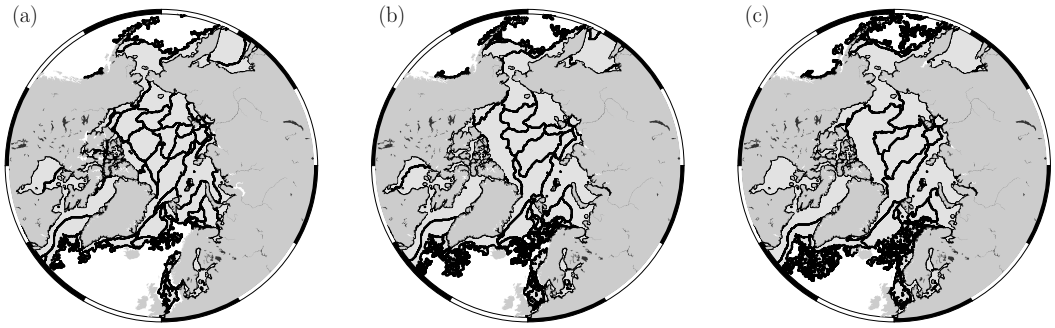


Figure A.3: Clusters generated from de-trended winter sea ice concentration time series data between 1979 and 2021, with significance levels (a) $\eta = 0.01$ ($N = 80$ clusters), (b) $\eta = 0.05$ ($N = 65$ clusters), and (c) $\eta = 0.10$ ($N = 61$ clusters).

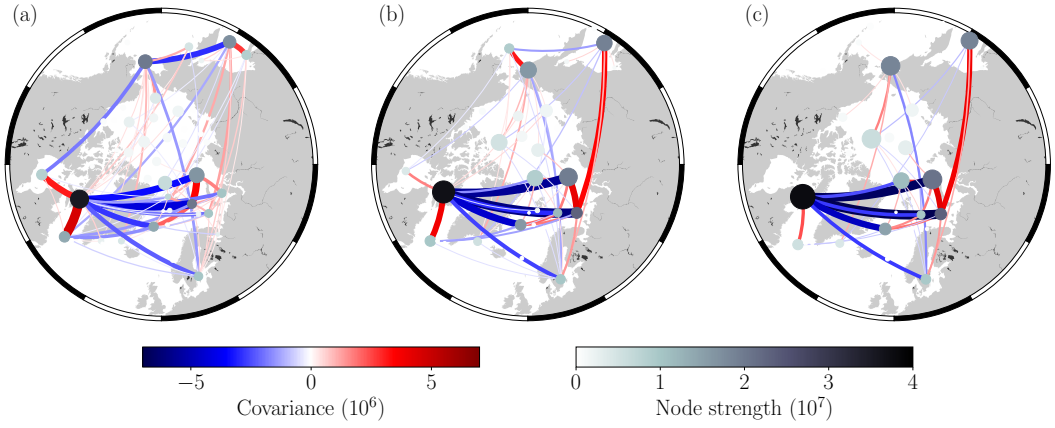


Figure A.4: Complex networks generated from de-trended winter sea ice concentration time series data between 1979 and 2021, based on clusters with significance levels (a) $\eta = 0.01$ ($N = 80$ clusters), (b) $\eta = 0.05$ ($N = 65$ clusters), and (c) $\eta = 0.10$ ($N = 61$ clusters). Only links which are significant at the 10% level are shown here to aid visualisation.

Appendix B

Chapter 4 supplementary figures

This appendix contains figures that were supplementary to the article "A Bayesian approach towards daily pan-Arctic sea ice freeboard estimates from combined CryoSat-2 and Sentinel-3 satellite observations", that forms chapter 4 of this thesis.

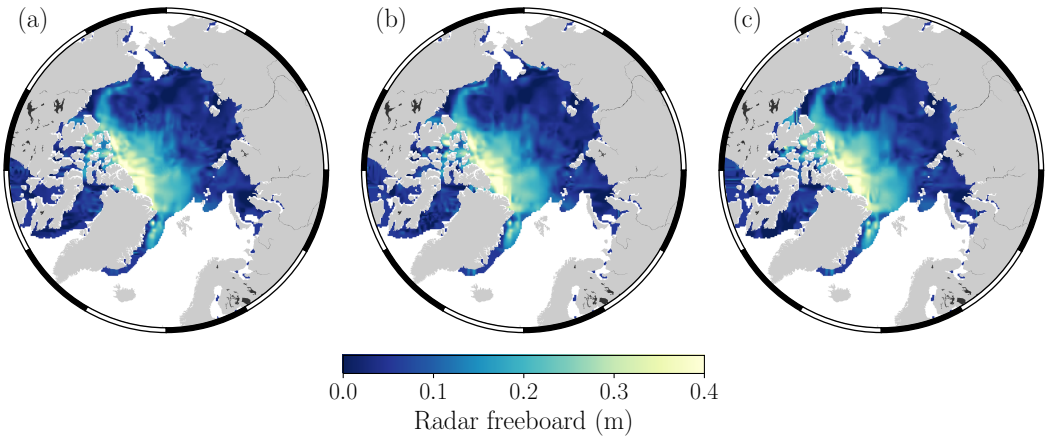


Figure B.1: Examples of 25×25 km gridded CS2S3 radar freeboard from Gaussian Process Regression, corresponding to the 1st of December 2018. Here we vary the number of days of observations used to train the model with (a) 9 days (as per our model in the main text), (b) 5 days (+/- 2 days around the prediction day), and (c) 3 days (+/- 1 day around the prediction day). With 3 days we see linear artefacts in some regions (Beaufort and Chukchi Seas), which are suppressed (but not entirely eliminated) by increasing to 5 days.

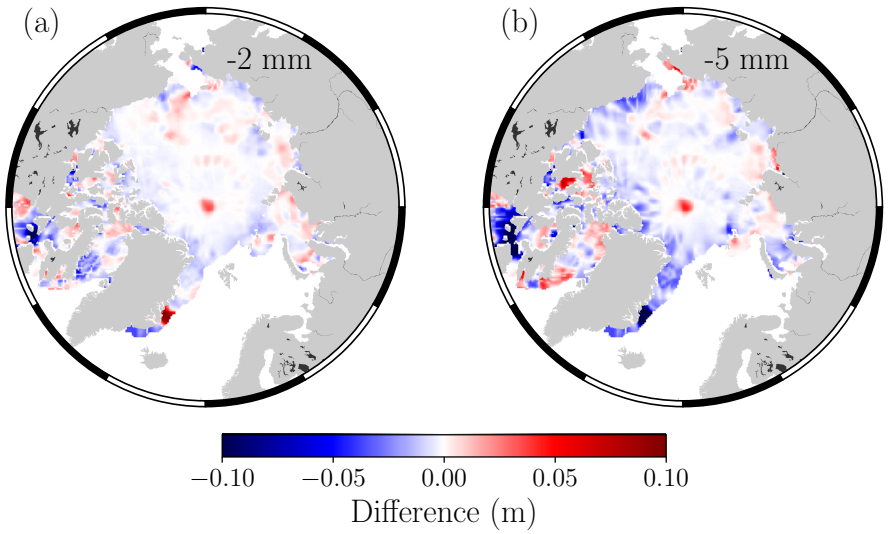


Figure B.2: The difference in prediction uncertainty between the model trained with 9 days of data and (a) the model trained with 5 days of data. (b) the model trained with 3 days of data. The average difference is given for each case (-2 mm and -5 mm respectively). Negative values indicate that the uncertainty of the 9-day model is lower than either the 5- or 3-day model. Note that uncertainty at the polar hole increases with more days of training data as no observations are ever recorded there.

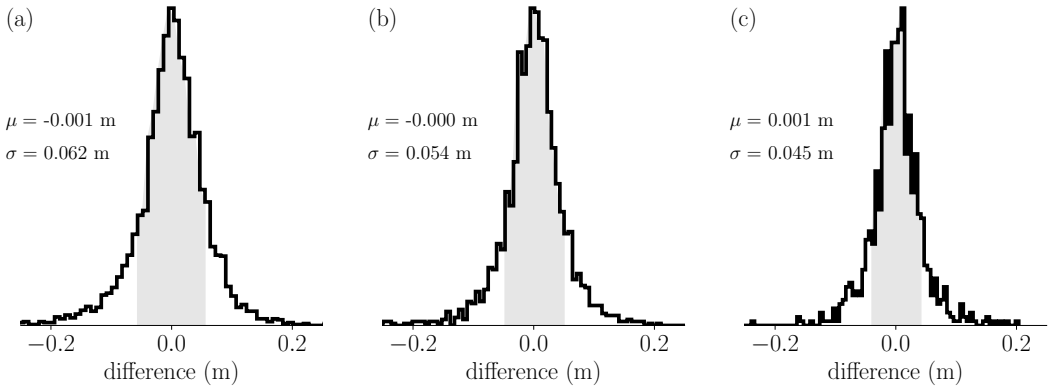


Figure B.3: Training error (observations – CS2S3) for one day (1st of December 2018), for interpolations run at (a) 25×25 km, (b) 50×50 km, (c) 100×100 km spatial resolution. Hence the along-track CS2 and S3 observations were first gridded to each respective resolution, and pan-Arctic predictions were subsequently generated at that same resolution. Notably here, increasing or decreasing the resolution does not result in a systematic increase or decrease in the average training error.

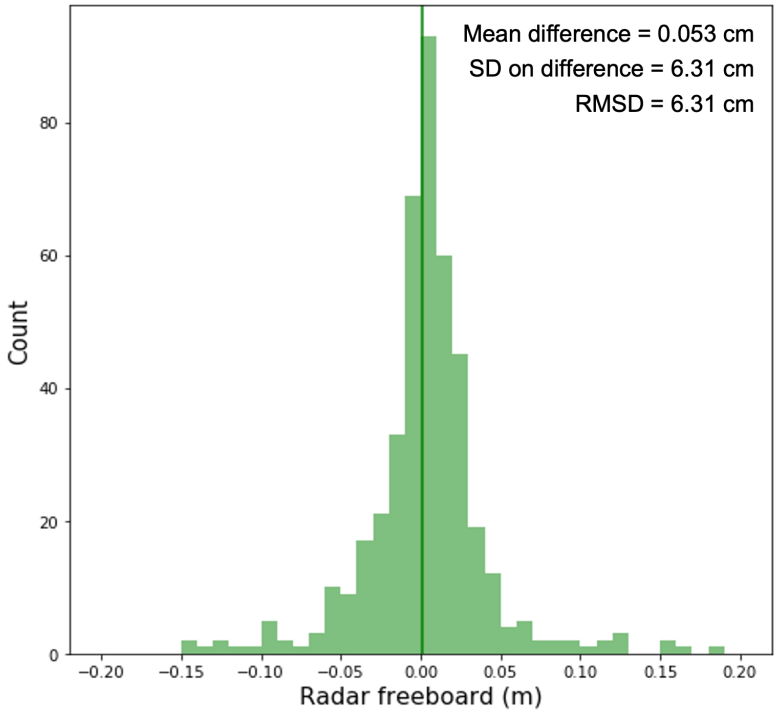


Figure B.4: Histogram of S3AS3B radar freeboard difference for tracks averaged on a 50×50 km grid. The green line shows the mean difference (=0.053 cm). A total of 35 tracks from the 14th and 15th October 2018 were gridded and compared. During this period S3A and S3B were operating in tandem mode, with S3A trailing S3B by $\sim$ 30 seconds in the same orbit. Sea ice drift during 30 seconds can be considered negligible, therefore we can assume that S3A and S3B are observing the same sea ice, and that differences are therefore the result of noise on individual freeboard measurements (see Wingham et al. (2006) for discussion of speckle noise on CS2 measurements). The standard deviation on the difference ($\sim$ 6 cm) provides an estimate of the uncertainty on 50 km grid-averaged S3 and CS2 radar freeboard measurements.

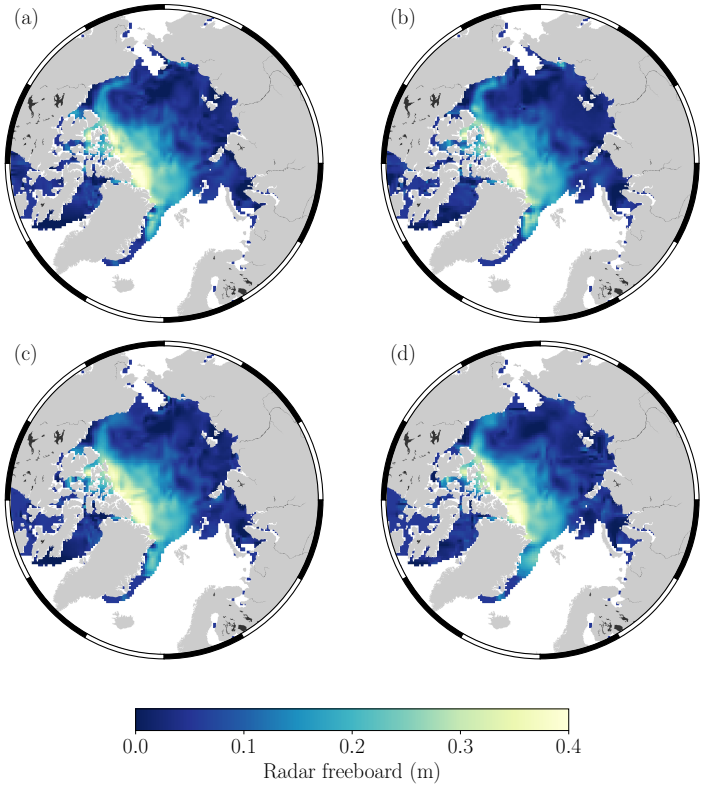


Figure B.5: 50×50 km gridded CS2S3 freeboard using 9 days of observations during training, but showing the freeboard predictions corresponding to each cross-validation experiment (section 4.4.2 main text). (a) Using CS2, S3A and S3B during training. (b) Using CS2 and S3B during training. (c) Using CS2 and S3A during training. (d) Using only CS2 during training. Notice how features such as the 'monkey tail' in the Beaufort Sea are less well defined in (b), (c) and (d), than in (a). Furthermore, without any S3 data in (d) we see linear interpolation artefacts.

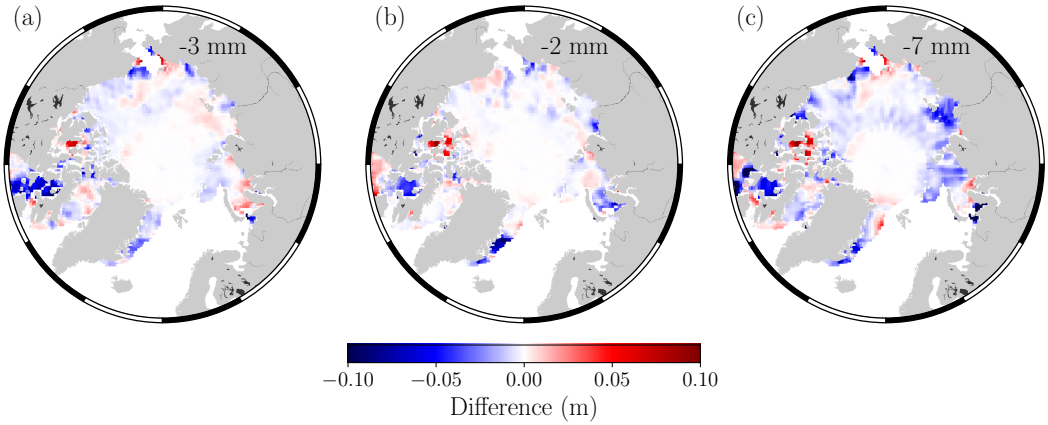


Figure B.6: The difference in prediction uncertainty between the model trained with CS2, S3A, and S3B (CS2S3), and (a) the model trained with CS2 and S3B (CS2S3(-S3A)). (b) the model trained with CS2 and S3A (CS2S3(-S3B)). (c) the model trained with CS2 only (CS2S3(-S3)). The average difference is given for each case. Negative values indicate that the uncertainty of the CS2S3 model is lower.

Appendix C

Chapter 5 supplementary figures

This appendix contains figures which are supplementary to chapter 5.

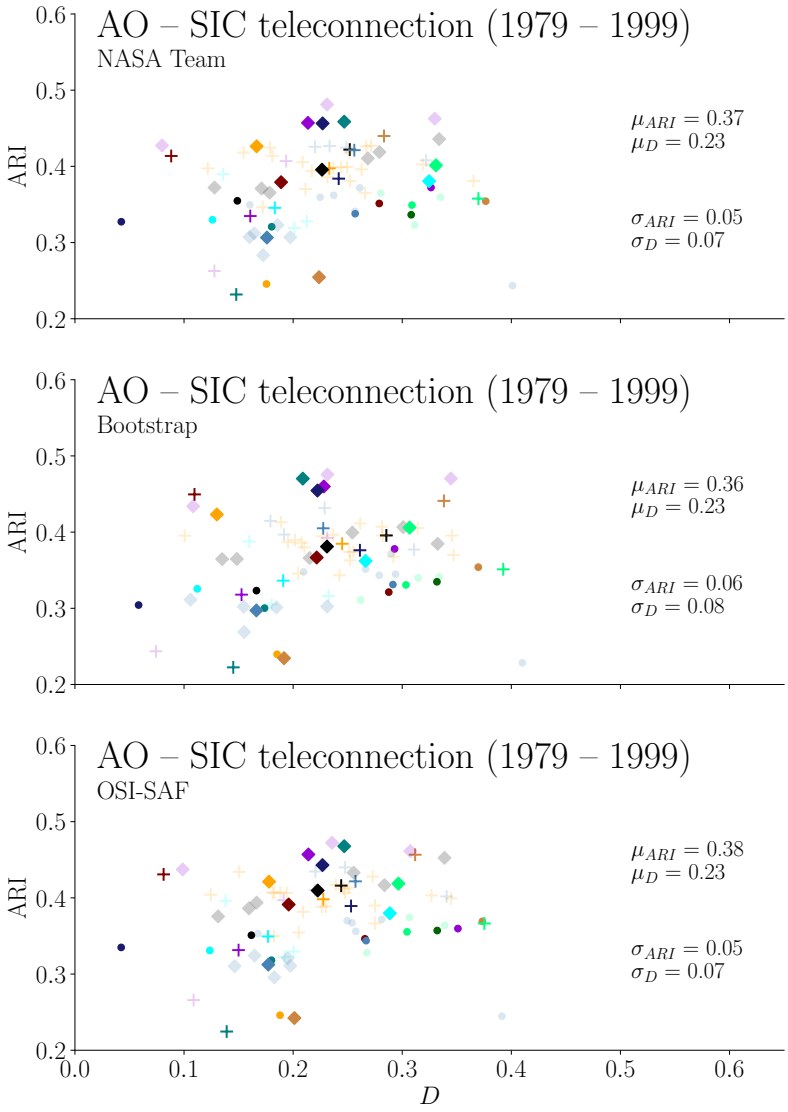


Figure C.1: ARI and D metrics for comparing observation and CMIP6 model summer sea ice concentration networks and the winter AO to summer sea ice teleconnection between 1979–1999, for every ensemble member for 31 different CMIP6 models (74 realisations). ARI and D are computed relative to NASA Team (top), Bootstrap (middle), and OSI-SAF (bottom) observational networks. Network distance values (D) are computed from observation and model 'link maps' as shown in Figure 5.9. The symbols and colours of each point are consistent with Figure 5.3.

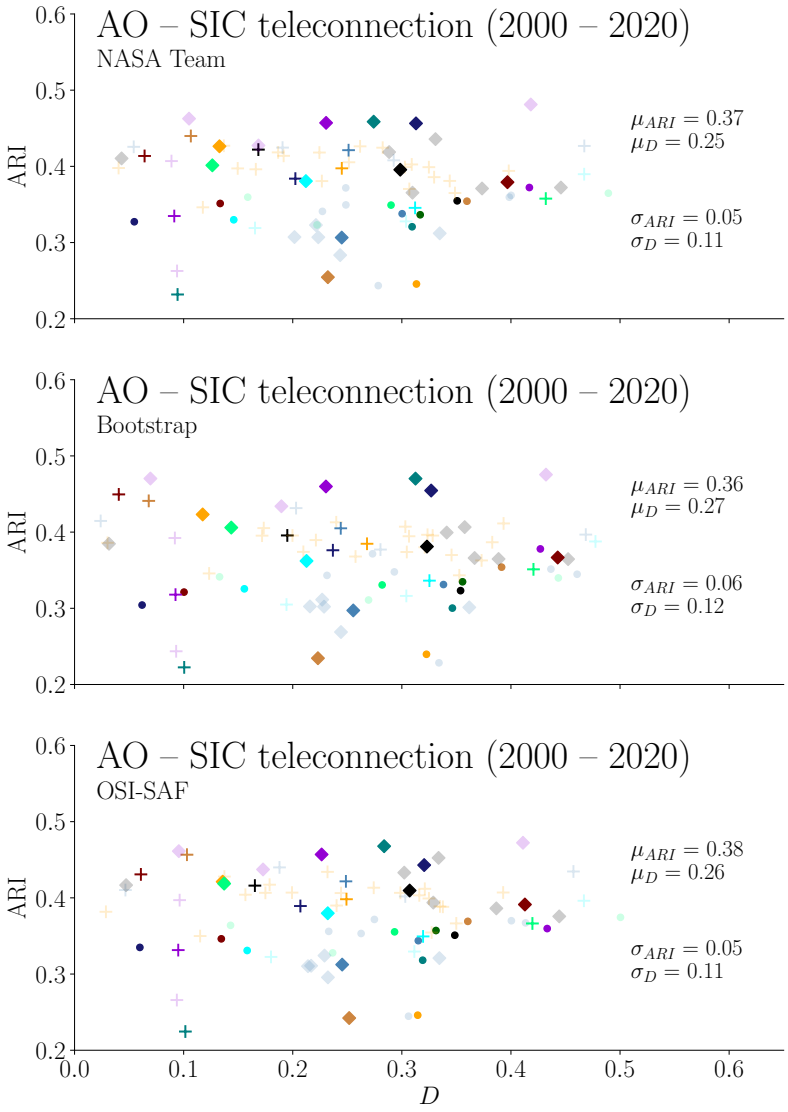


Figure C.2: ARI and D metrics for comparing observation and CMIP6 model summer sea ice concentration networks and the winter AO to summer sea ice teleconnection between 2000–2020, for every ensemble member for 31 different CMIP6 models (74 realisations). ARI and D are computed relative to NASA Team (top), Bootstrap (middle), and OSI-SAF (bottom) observational networks. Network distance values (D) are computed from observation and model 'link maps' as shown in Figure 5.9. The symbols and colours of each point are consistent with Figure 5.3.

Bibliography

Aaboe, S., Breivik, L.-A., Sørensen, A., Eastwood, S., & Lavergne, T. (2016). Global sea ice edge and type product user's manual. *OSI-403-c & EUMETSAT*.

Abe, S., & Suzuki, N. (2006). Complex-network description of seismicity. *Nonlinear Processes in Geophysics*, *13*(2), 145–150.

Aizerman, M. (1964). The probability problem of pattern recognition learning and the method of potential functions. *Automation and Remote Control*, *25*, 1175–1193.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Allard, R. A., Farrell, S. L., Hebert, D. A., Johnston, W. F., Li, L., Kurtz, N. T., Phelps, M. W., Posey, P. G., Tilling, R., Ridout, A., et al. (2018). Utilizing cryosat-2 sea ice thickness to initialize a coupled ice-ocean modeling system. *Advances in Space Research*, *62*(6), 1265–1280.

Ambaum, M. H., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic oscillation or north atlantic oscillation? *Journal of Climate*, *14*(16), 3495–3507.

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., et al. (2021). Seasonal arctic sea ice forecasting with probabilistic deep learning. *Nature communications*, *12*(1), 1–12.

Aporta, C. (2009). The trail as home: Inuit and their pan-arctic network of routes. *Human Ecology*, *37*(2), 131–146.

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P. W., B.Trewin, Rao, K. A., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J., C.Cassou, Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J., andA. Di Luca, C. D., Niang, A. D., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., B.Fox-Kemper, Fuglestvedt, J. S., Fyfe,

J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R., Hawkins, E., Hewitt, H., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Y.Kosaka, Kossin, J., Krakovska, S., Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S.-K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J.-B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson, A. A., Szopa, S., Takayabu, I., Treguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., & Zickfeld, K. (2021). Technical summary. In *Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Årthun, M., Onarheim, I. H., Dörr, J., & Eldevik, T. (2021). The seasonal and regional transition to an ice-free arctic. *Geophysical Research Letters*, *48*(1), e2020GL090825.

Ashton, S. R., & Sollich, P. (2012). Learning curves for multi-task gaussian process regression. *arXiv preprint arXiv:1211.0439*.

Balan-Sarojini, B., Tietsche, S., Mayer, M., Balmaseda, M., Zuo, H., de Rosnay, P., Stockdale, T., & Vitart, F. (2021). Year-round impact of winter sea ice thickness observations on seasonal forecasts. *The Cryosphere*, *15*(1), 325–344.

Balmaseda, M., Alves, O., Awaji, T., Behringer, D., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T., Stammer, D., & Others (2010). Initialization for seasonal and decadal forecasts. In *OceanObs 09*, (pp. 19–26). European Space Agency.

Bellman, R. E. (2015). *Adaptive control processes: a guided tour*. Princeton university press.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bernardo, J. M., & Smith, A. F. (2009). *Bayesian theory*, vol. 405. John Wiley & Sons.

Bintanja, R., & Van der Linden, E. (2013). The changing seasonal climate in the arctic. *Scientific reports*, *3*(1), 1–8.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Björnsson, H., & Venegas, S. (1997). A manual for eof and svd analyses of climatic data. *CCGCR Report*, *97*(1), 112–134.

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., & DeWeaver, E. (2011). Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *Journal of Climate*, *24*(1), 231–250.

Blanchard-Wrigglesworth, E., & Bitz, C. M. (2014). Characteristics of arctic sea-ice thickness variability in gcms. *Journal of Climate*, *27*(21), 8244–8258.

Bliss, A., & Anderson, M. (2014). Snowmelt onset over arctic sea ice from passive microwave satellite data: 1979–2012. *The Cryosphere*, *8*(6), 2089–2100.

Blockley, E. W., & Peterson, K. A. (2018). Improving met office seasonal predictions of arctic sea ice using assimilation of cryosat-2 thickness. *The Cryosphere*, *12*(11), 3419–3438.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*(1), 1–122.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4-5), 175–308.

Boers, N., Bookhagen, B., Barbosa, H. M. J., Marwan, N., Kurths, J., & Marengo, J. A. (2014). Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nature communications*, *5*, 5199.

Bonan, D., & Blanchard-Wrigglesworth, E. (2020). Nonstationary teleconnection between the pacific ocean and arctic sea ice. *Geophysical Research Letters*, *47*(2), e2019GL085666.

Bonan, D., Bushuk, M., & Winton, M. (2019). A spring barrier for regional predictions of summer arctic sea ice. *Geophysical Research Letters*.

Bonilla, E. V., Chai, K. M., & Williams, C. (2008). Multi-task gaussian process prediction. *Advances in neural information processing systems*, (pp. 153–160).

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, (pp. 144–152).

Bushuk, M., Giannakis, D., & Majda, A. J. (2014). Reemergence mechanisms for North Pacific sea ice revealed through nonlinear Laplacian spectral analysis. *Journal of Climate*, *27*(16), 6265–6287.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., & Gudgel, R. (2019). Regional arctic sea–ice prediction: Potential versus operational seasonal forecast skill. *Climate Dynamics*, *52*(5), 2721–2743.

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

Bushuk, M., Winton, M., Bonan, D. B., Blanchard-Wrigglesworth, E., & Delworth, T. L. (2020). A mechanism for the arctic sea ice spring predictability barrier. *Geophysical Research Letters*, *47*(13), e2020GL088335.

Candela, J. Q., & Hansen, L. K. (2004). Learning with uncertainty-gaussian processes and relevance vector machines. *Technical University of Denmark*, (pp. 1–152).

Cattiaux, J., & Cassou, C. (2013). Opposite cmip3/cmip5 trends in the wintertime northern annular mode explained by combined local sea ice and remote tropical influences. *Geophysical research letters*, *40*(14), 3682–3687.

Cavalieri, D. J., Parkinson, C. L., Gloersen, P., & Zwally, H. J. (1996). Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. NASA Natl. Snow and Ice Data Cent. *Distrib. Active Arch. Cent., Boulder, Colo.*.

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

Chevallier, M., Salas y Mélia, D., Voldoire, A., Déqué, M., & Garric, G. (2013). Seasonal forecasts of the pan-arctic sea ice extent using a gcm-based seasonal prediction system. *Journal of Climate*, *26*(16), 6092–6104.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., & Others (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature geoscience*, *7*(9), 627.

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., Ballinger, T., Bhatt, U., Chen, H., Coumou, D., et al. (2020). Divergent consensuses on arctic amplification influence on midlatitude severe winter weather. *Nature Climate Change*, *10*(1), 20–29.

Cohen, R., & Havlin, S. (2010). *Complex networks: structure, robustness and function*. Cambridge university press.

Collow, T. W., Wang, W., Kumar, A., & Zhang, J. (2015). Improving arctic sea ice prediction using piomas initial sea ice thickness in a coupled ocean–atmosphere model. *Monthly Weather Review*, *143*(11), 4618–4630.

Comeau, D., Giannakis, D., Zhao, Z., & Majda, A. J. (2019). Predicting regional and pan-arctic sea ice anomalies with kernel analog forecasting. *Climate Dynamics*, *52*(9-10), 5507–5525.

Comiso, J. (1983). Sea ice effective microwave emissivities from satellite passive microwave and infrared observations. *Journal of Geophysical Research: Oceans*, *88*(C12), 7686–7704.

Comiso, J. (2017). Bootstrap sea ice concentrations from nimbus-7 smmr and dmsp ssm/i-ssmis, version 3. *NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO*, *2017*, 575.

Comiso, J. C. (1986). Characteristics of arctic winter sea ice from satellite multispectral microwave observations. *Journal of Geophysical Research: Oceans*, *91*(C1), 975–994.

Comiso, J. C., Meier, W. N., & Gersten, R. (2017). Variability and trends in the arctic sea ice cover: Results from different techniques. *Journal of Geophysical Research: Oceans*, *122*(8), 6883–6900.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.

Crawford, A. D., Horvath, S., Stroeve, J., Balaji, R., & Serreze, M. C. (2018). Modulation of sea ice melt onset and retreat in the laptev sea by the timing of snow retreat in the west siberian plain. *Journal of Geophysical Research: Atmospheres*, *123*(16), 8691–8707.

Cressie, N., & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 209–226.

Curry, J. A., Schramm, J. L., & Ebert, E. E. (1995). Sea ice-albedo climate feedback mechanism. *Journal of Climate*, *8*(2), 240–247.

Dall, M., Beddows, D., Tunved, P., Krejci, R., Ström, J., Hansson, H.-C., Yoon, Y., Park, K.-T., Becagli, S., Udisti, R., et al. (2017). Arctic sea ice melt leads to atmospheric new particle formation. *Scientific reports*, *7*(1), 1–10.

Dall, M., Geels, C., Beddows, D., Boertmann, D., Lange, R., Nøjgaard, J., Harrison, R. M., Simo, R., Skov, H., Massling, A., et al. (2018). Regions of open water and melting sea ice drive new particle formation in north east greenland. *Scientific reports*, *8*(1), 1–10.

Dammann, D. O., Eicken, H., Mahoney, A. R., Meyer, F. J., & Betcher, S. (2018). Assessing sea ice trafficability in a changing arctic. *Arctic*, *71*(1), 59–75.

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(1), 1–15.

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

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

delEtoile, J., & Adeli, H. (2017). Graph theory and brain connectivity in alzheimer's disease. *The Neuroscientist*, *23*(6), 616–626.

Deser, C., Walsh, J. E., & Timlin, M. S. (2000). Arctic sea ice variability in the context of recent atmospheric circulation trends. *Journal of Climate*, *13*(3), 617–633.

Dijkstra, H. A., Hernández-García, E., Masoller, C., & Barreiro, M. (2019). *Networks in Climate*. Cambridge University Press.

Dinardo, S., Lucas, B., & Benveniste, J. (2014). Sar altimetry processing on demand service for cryosat-2 at esa g-pod. In *Proc. of 2014 Conference on Big Data from Space (BiDS'14)*, (p. 386).

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

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(4), 245–268.

Donges, J. F., Petrova, I., Loew, A., Marwan, N., & Kurths, J. (2015). How complex climate networks complement eigen techniques for the statistical analysis of climatological data. *Climate dynamics*, *45*(9-10), 2407–2424.

Donges, J. F., Zou, Y., Marwan, N., & Kurths, J. (2009). Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods. *European Physical Journal: Special Topics*, *174*(1), 157–179.

Drobot, S. D., Maslanik, J. A., & Fowler, C. (2006). A long-range forecast of Arctic summer sea-ice minimum extent. *Geophysical Research Letters*, *33*(10), 1–4.

Eguíluz, V. M., Fernández-Gracia, J., Irigoien, X., & Duarte, C. M. (2016). A quantitative assessment of arctic shipping in 2010–2014. *Scientific reports*, *6*(1), 1–6.

ERA5 (2017). Copernicus climate change service (cds): Fifth generation of ecmwf atmospheric reanalyses of the global climate. accessed 5th of august 2019. URL https://cds.climate.copernicus.eu/cdsapp#!/home

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

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

Fetterer, F., Savoie, M., Helfrich, S., & Clemente-Colón, P. (2010). Multisensor analyzed sea ice extent-northern hemisphere (masie-nh). Tech. rep., Technical report, National Snow and Ice Data Center. Boulder, Colorado USA.

Fountalis, I., Bracco, A., & Dovrolis, C. (2014). Spatio-temporal network analysis for studying climate patterns. *Climate Dynamics*, *42*(3-4), 879–899.

Fountalis, I., Bracco, A., & Dovrolis, C. (2015). Enso in cmip5 simulations: network connectivity from the recent past to the twenty-third century. *Climate Dynamics*, *45*(1-2), 511–538.

Fountalis, I., Dovrolis, C., Bracco, A., Dilkina, B., & Keilholz, S. (2018). δ-maps: from spatio-temporal data to a weighted and lagged network between functional domains. *Applied network science*, *3*(1), 21.

Giannakis, D., & Majda, A. J. (2012). Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, *109*(7), 2222–2227.

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., Vuuren, D. P. v., Berg, M. v. d., Feng, L., Klein, D., et al. (2019). Global emissions pathways under different socioeconomic scenarios for use in cmip6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific model development*, *12*(4), 1443–1475.

Giesse, C., Notz, D., & Baehr, J. (2021). On the origin of discrepancies between observed and simulated memory of arctic sea ice. *Geophysical Research Letters*, *48*(11), e2020GL091784.

Giles, K. A., Laxon, S. W., Ridout, A. L., Wingham, D. J., & Bacon, S. (2012). Western arctic ocean freshwater storage increased by wind-driven spin-up of the beaufort gyre. *Nature Geoscience*, *5*(3), 194–197.

Gillett, N. P., Kirchmeier-Young, M., Ribes, A., Shiogama, H., Hegerl, G. C., Knutti, R., Gastineau, G., John, J. G., Li, L., Nazarenko, L., et al. (2021). Constraining human contributions to observed warming since the pre-industrial period. *Nature Climate Change*, *11*(3), 207–212.

Girard, A., Rasmussen, C. E., Candela, J. Q., & Murray-Smith, R. (2003). Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. In *Advances in neural information processing systems*, (pp. 545–552).

Glantz, M. H., Katz, R. W., Nicholls, N., et al. (1991). *Teleconnections linking worldwide climate anomalies*, vol. 535. Cambridge University Press Cambridge.

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

Gong, H., Wang, L., Chen, W., Chen, X., & Nath, D. (2016). Biases of the wintertime arctic oscillation in cmip5 models. *Environmental Research Letters*, *12*(1), 014001.

Gong, H., Wang, L., Chen, W., Wu, R., Zhou, W., Liu, L., Nath, D., & Lan, X. (2019). Diversity of the wintertime arctic oscillation pattern among cmip5 models: Role of the stratospheric polar vortex. *Journal of Climate*, *32*(16), 5235–5250.

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., & Hudson, S. R. (2019). Evaluation of six atmospheric reanalyses over arctic sea ice from winter to early summer. *Journal of Climate*, *32*(14), 4121–4143.

Gregory, J. M., Stott, P., Cresswell, D., Rayner, N., Gordon, C., & Sexton, D. (2002). Recent and future changes in arctic sea ice simulated by the hadcm3 aogcm. *Geophysical Research Letters*, *29*(24), 28–1.

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

Hamilton, L. C., & Stroeve, J. (2016). 400 Predictions: the SEARCH sea ice outlook 2008–2015. *Polar Geography*, *39*(4), 274–287.

Hanna, E., Mernild, S. H., Cappelen, J., & Steffen, K. (2012). Recent warming in Greenland in a long-term instrumental (1881–2012) climatic context: I. Evaluation of surface air temperature records. *Environmental Research Letters*, *7*(4), 45404.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*, vol. 43. CRC press.

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Holland, M. M., Bailey, D. A., & Vavrus, S. (2011). Inherent sea ice predictability in the rapidly changing arctic environment of the community climate system model, version 3. *Climate dynamics*, *36*(7-8), 1239–1253.

Holland, M. M., Serreze, M. C., & Stroeve, J. (2010). The sea ice mass budget of the arctic and its future change as simulated by coupled climate models. *Climate Dynamics*, *34*(2-3), 185–200.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*(1), 193–218.

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003). An overview of the north atlantic oscillation. *Geophysical Monograph-American Geophysical Union*, *134*, 1–36.

Ionita, M., Grosfeld, K., Scholz, P., Treffeisen, R., & Lohmann, G. (2019). September arctic sea ice minimum prediction - a skillful new statistical approach. *Earth System Dynamics*, *10*(1), 189–203.

IPCC (2021). In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Jahn, A. (2018). Reduced probability of ice-free summers for 1.5 c compared to 2 c warming. *Nature Climate Change*, *8*(5), 409.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, *31*(3), 264–323.

James, F. C., & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or pandora's box? *Annual review of Ecology and Systematics*, *21*(1), 129–166.

Jin-Qing, Z., Wei-Jing, L., & Hong-Li, R. (2013). Representation of the arctic oscillation in the cmip5 models. *Advances in Climate Change Research*, *4*(4), 242–249.

Jun, S.-Y., Ho, C.-H., Jeong, J.-H., Choi, Y.-S., & Kim, B.-M. (2016). Recent changes in winter arctic clouds and their relationships with sea ice and atmospheric conditions. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 29130.

Kang, E. L., Cressie, N., & Shi, T. (2010). Using temporal variability to improve spatial mapping with application to satellite data. *Canadian Journal of Statistics*, *38*(2), 271–289.

Kapsch, M.-L., Graversen, R. G., & Tjernström, M. (2013). Springtime atmospheric energy transport and the control of arctic summer sea-ice extent. *Nature Climate Change*, *3*(8), 744–748.

Kay, J. E., & L'Ecuyer, T. (2013). Observational constraints on arctic ocean clouds and radiative fluxes during the early 21st century. *Journal of Geophysical Research: Atmospheres*, *118*(13), 7219–7236.

Kondor, R. I., & Lafferty, J. (2002). Diffusion kernels on graphs and other discrete structures. *Proceedings of the 19th international conference on machine learning*, *2002*, 315–322.

Kostopoulou, E. (2021). Applicability of ordinary kriging modeling techniques for filling satellite data gaps in support of coastal management. *Modeling Earth Systems and Environment*, *7*(2), 1145–1158.

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of Climate*, *29*(11), 4069–4081.

Kurtz, N., Markus, T., Farrell, S., Worthen, D., & Boisvert, L. (2011). Observations of recent arctic sea ice volume loss and its impact on ocean-atmosphere energy exchange and ice production. *Journal of Geophysical Research: Oceans*, *116*(C4).

Kutzbach, J. E. (1967). Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over north america. *Journal of Applied Meteorology*, *6*(5), 791–802.

Kwok, R. (2000). Recent changes in arctic ocean sea ice motion associated with the north atlantic oscillation. *Geophysical Research Letters*, *27*(6), 775–778.

Kwok, R. (2015). Sea ice convergence along the arctic coasts of greenland and the canadian arctic archipelago: Variability and extremes (1992–2014). *Geophysical Research Letters*, *42*(18), 7598–7605.

Kwok, R. (2018). Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018). *Environmental Research Letters*, *13*(10), 105005.

Kwok, R., Kacimi, S., Markus, T., Kurtz, N., Studinger, M., Sonntag, J., Manizade, S., Boisvert, L., & Harbeck, J. (2019). Icesat-2 surface height and sea ice freeboard

assessed with atm lidar acquisitions from operation icebridge. *Geophysical Research Letters*, *46*(20), 11228–11236.

Kwok, R., Kacimi, S., Webster, M., Kurtz, N., & Petty, A. (2020). Arctic snow depth and sea ice thickness from icesat-2 and cryosat-2 freeboards: a first examination. *Journal of Geophysical Research: Oceans*, *125*(3), e2019JC016008.

Kwok, R., Rignot, E., Holt, B., & Onstott, R. (1992). Identification of sea ice types in spaceborne synthetic aperture radar data. *Journal of Geophysical Research: Oceans*, *97*(C2), 2391–2402.

Kwok, R., Spreen, G., & Pang, S. (2013). Arctic sea ice circulation and drift speed: Decadal trends and ocean currents. *Journal of Geophysical Research: Oceans*, *118*(5), 2408–2425.

Landy, J. C., Bouffard, J., Wilson, C., Rynders, S., Aksenov, Y., & Tsamados, M. (2021). Mapping arctic sea ice thickness: A new method for improved ice freeboard retrieval from satellite altimetry. *Earth and Space Science Open Archive ESSOAr*.

Lange, R., Dall'Osto, M., Skov, H., Nøjgaard, J., Nielsen, I., Beddows, D., Simó, R., Harrison, R. M., & Massling, A. (2018). Characterization of distinct arctic aerosol accumulation modes and their sources. *Atmospheric Environment*, *183*, 1–10.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., et al. (2019). Version 2 of the eumetsat osi saf and esa cci sea-ice concentration climate data records. *The Cryosphere*, *13*(1), 49–78.

Lawrence, I. R. (2019). *Multi-satellite synergies for polar ocean altimetry*. Ph.D. thesis, UCL (University College London).

Lawrence, I. R., Armitage, T. W., Tsamados, M. C., Stroeve, J. C., Dinardo, S., Ridout, A. L., Muir, A., Tilling, R. L., & Shepherd, A. (2019). Extending the arctic sea ice freeboard and sea level record with the sentinel-3 radar altimeters. *Advances in Space Research*.

Lawrence, I. R., Tsamados, M. C., Stroeve, J. C., Armitage, T. W., & Ridout, A. L. (2018). Estimating snow depth over arctic sea ice from calibrated dual-frequency radar freeboards. *The Cryosphere*, *12*(11), 3551–3564.

Laxon, S., Peacock, N., & Smith, D. (2003). High interannual variability of sea ice thickness in the arctic region. *Nature*, *425*(6961), 947–950.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., et al. (2013). Cryosat-2 estimates of arctic sea ice thickness and volume. *Geophysical Research Letters*, *40*(4), 732–737.

Le Traon, P., Nadal, F., & Ducet, N. (1997). An improved mapping method of multisatellite altimeter data. *Journal of atmospheric and oceanic technology*, *15*(2), 522–534.

Lee, S.-W., & Song, J.-M. (2014). Economic possibilities of shipping though Northern Sea route. *The Asian Journal of Shipping and Logistics*, *30*(3), 415–430.

Light, B., Grenfell, T. C., & Perovich, D. K. (2008). Transmission and absorption of solar radiation by arctic sea ice during the melt season. *Journal of Geophysical Research: Oceans*, *113*(C3).

Lindsay, R., Zhang, J., Schweiger, A., & Steele, M. (2008). Seasonal Predictions of Ice Extent in the Arctic Ocean. *Journal of the Institution of Engineers (India): Environmental Engineering Division*, *113*.

Liston, G. E., Itkin, P., Stroeve, J., Tschudi, M., Stewart, J. S., Pedersen, S. H., Reinking, A. K., & Elder, K. (2020). A lagrangian snow-evolution system for sea-ice applications (snowmodel-lg): Part i—model description. *Journal of Geophysical Research: Oceans*, *125*(10), e2019JC015913.

Liu, Y., Key, J. R., Wang, X., & Tschudi, M. (2020). Multidecadal arctic sea ice thickness and volume derived from ice age. *The Cryosphere*, *14*(4), 1325–1345.

Lorenz, E. N. (1956). Empirical orthogonal functions and statistical weather prediction.

MacKay, D. J. (1998). Introduction to gaussian processes. *NATO ASI series F computer and systems sciences*, *168*, 133–166.

Malik, N., Bookhagen, B., Marwan, N., & Kurths, J. (2012). Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Climate dynamics*, *39*(3-4), 971–987.

Markus, T., Stroeve, J. C., & Miller, J. (2009). Recent changes in arctic sea ice melt onset, freezeup, and melt season length. *Journal of Geophysical Research: Oceans*, *114*(C12).

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

Maslanik, J., & Stroeve, J. (1999). Near-real-time dmsp ssm/i-ssmis daily polar gridded sea ice concentrations. *National Snow and Ice Data Center, Boulder, CO, digital media.[Available online at http://nsidc. org/data/docs/daac/nsidc0081_ssmi_nrt_seaice. gd. html.]*.

Maslanik, J., Stroeve, J., Fowler, C., & Emery, W. (2011). Distribution and trends in arctic sea ice age through spring 2011. *Geophysical Research Letters*, *38*(13).

Maslanik, J. A., Serreze, M. C., & Barry, R. G. (1996). Recent decreases in arctic summer ice cover and linkages to atmospheric circulation anomalies. *Geophysical Research Letters*, *23*(13), 1677–1680.

Massonnet, F., Fichefet, T., & Goosse, H. (2015). Prospects for improved seasonal arctic sea ice predictions from multivariate data assimilation. *Ocean Modelling*, *88*, 16–25.

Massonnet, F., Vancoppenolle, M., Goosse, H., Docquier, D., Fichefet, T., & Blanchard-Wrigglesworth, E. (2018). Arctic sea-ice change tied to its mean state through thermodynamic processes. *Nature Climate Change*, *8*(7), 599–603.

Matsumura, S., Zhang, X., & Yamazaki, K. (2014). Summer arctic atmospheric circulation response to spring eurasian snow cover and its possible linkage to accelerated sea ice decrease. *Journal of Climate*, *27*(17), 6551–6558.

Maykut, G. A. (1978). Energy exchange over young sea ice in the central arctic. *Journal of Geophysical Research: Oceans*, *83*(C7), 3646–3658.

McKenna Jr, J. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling & Software*, *18*(3), 205–220.

Mernild, S. H., Lipscomb, W. H., Bahr, D. B., Radić, V., & Zemp, M. (2013). Global glacier changes: A revised assessment of committed mass losses and sampling uncertainties. *Cryosphere*, *7*(5), 1565–1577.

Miller, R., Schmidt, G., & Shindell, D. (2006). Forced annular variations in the 20th century intergovernmental panel on climate change fourth assessment report models. *Journal of Geophysical Research: Atmospheres*, *111*(D18).

Mioduszewski, J. R., Vavrus, S., Wang, M., Holland, M., & Landrum, L. (2019). Past and future interannual variability in arctic sea ice in coupled climate models. *The Cryosphere*, *13*(1), 113–124.

Moore, S. E., & Huntington, H. P. (2008). Arctic marine mammals and climate change: impacts and resilience. *Ecological Applications*, *18*(sp2), S157–S165.

Morabito, F. C., Campolo, M., Labate, D., Morabito, G., Bonanno, L., Bramanti, A., De Salvo, S., Marra, A., & Bramanti, P. (2015). A longitudinal eeg study of alzheimer's disease progression based on a complex network approach. *International journal of neural systems*, *25*(02), 1550005.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P., & Simpson, I. (2021). An updated assessment of near-surface temperature change from 1850: The hadcrut5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361.

Msadek, R., Vecchi, G. A., Winton, M., & Gudgel, R. (2014). Importance of initial conditions in seasonal predictions of arctic sea ice extent. *Geophysical Research Letters*, *41*(14), 5208–5215.

Myers III, R. E., & Fouts, J. T. (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science teaching*, *29*(9), 929–937.

Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, Ontario, Canada.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256.

Niederdrenk, A. L., & Notz, D. (2018). Arctic sea ice in a 1.5 c warmer world. *Geophysical Research Letters*, *45*(4), 1963–1971.

Nolin, A. W., Fetterer, F. M., & Scambos, T. A. (2002). Surface roughness characterizations of sea ice and ice sheets: Case studies with misr data. *IEEE transactions on Geoscience and Remote Sensing*, *40*(7), 1605–1615.

North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. (1982). Sampling errors in the estimation of empirical orthogonal functions. *Monthly weather review*, *110*(7), 699–706.

Notz, D. (2015). How well must climate models agree with observations? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2052), 20140164.

Notz, D. (2017). Arctic sea ice seasonal-to-decadal variability and long-term change. *Past Global Changes Magazine*, *25*, 14–19.

Notz, D., & Bitz, C. M. (2017). Sea ice in earth system models. *Sea ice*, *3*, 304–325.

Notz, D., & Community, S. (2020). Arctic sea ice in cmip6. *Geophysical Research Letters*, *47*(10), e2019GL086749.

Notz, D., & Stroeve, J. (2016). Observed arctic sea-ice loss directly follows anthropogenic co2 emission. *Science*, *354*(6313), 747–750.

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

Onarheim, I. H., Eldevik, T., Smedsrud, L. H., & Stroeve, J. C. (2018). Seasonal and regional manifestation of arctic sea ice loss. *Journal of Climate*, *31*(12), 4917–4932.

Ono, J., Komuro, Y., & Tatebe, H. (2020). Impact of sea-ice thickness initialized in april on arctic sea-ice extent predictability with the miroc climate model. *Annals of Glaciology*, *61*, 97–105.

OSI-SAF (2017). Global sea ice concentration climate data record v2.0 - multimission, eumetsat saf on ocean and sea ice.

Overland, J. E., Francis, J. A., Hanna, E., & Wang, M. (2012). The recent shift in early summer Arctic atmospheric circulation. *Geophysical Research Letters*, *39*(19), 1–6.

Overland, J. E., & Wang, M. (2013). When will the summer arctic be nearly sea ice free? *Geophysical Research Letters*, *40*(10), 2097–2101.

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., & van Vuuren, D. P. (2014). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic change*, *122*(3), 387–400.

Paciorek, C. J., & Schervish, M. J. (2005). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, *17*(5), 483–506.

Panikkar, B., Lemmond, B., Else, B., & Murray, M. (2018). Ice over troubled waters: navigating the northwest passage using inuit knowledge and scientific information. *Climate Research*, *75*(1), 81–94.

Park, H.-S., Stewart, A. L., & Son, J.-H. (2018). Dynamic and thermodynamic impacts of the winter arctic oscillation on summer sea ice extent. *Journal of Climate*, *31*(4), 1483–1497.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

Peterson, K. A., Arribas, A., Hewitt, H., Keen, A., Lea, D., & McLaren, A. (2015). Assessing the forecast skill of arctic sea ice extent in the glosea4 seasonal prediction system. *Climate dynamics*, *44*(1-2), 147–162.

Petty, A. A., Hutchings, J. K., Richter-Menge, J. A., & Tschudi, M. A. (2016a). Sea ice circulation around the beaufort gyre: The changing role of wind forcing and the sea ice state. *Journal of Geophysical Research: Oceans*, *121*(5), 3278–3296.

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

Petty, A. A., Tsamados, M. C., Kurtz, N. T., Farrell, S. L., Newman, T., Harbeck, J. P., Feltham, D. L., & Richter-Menge, J. A. (2016b). Characterizing arctic sea ice topography using high-resolution icebridge data. *The Cryosphere*, *10*(3), 1161–1179.

Petty, A. A., Webster, M., Boisvert, L., & Markus, T. (2018). The nasa eulerian snow on sea ice model (nesosim) v1. 0: initial model development and analysis. *Geoscientific Model Development*, *11*(11), 4577–4602.

Ponsoni, L., Massonnet, F., Fichefet, T., Chevallier, M., & Docquier, D. (2019). On the timescales and length scales of the arctic sea ice thickness anomalies: a study based on 14 reanalyses. *The Cryosphere*, *13*(2), 521–543.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical recipes in c++. *The art of scientific computing*, *2*, 1002.

Proshutinsky, A., Krishfield, R., Timmermans, M.-L., Toole, J., Carmack, E., McLaughlin, F., Williams, W. J., Zimmermann, S., Itoh, M., & Shimada, K. (2009). Beaufort gyre freshwater reservoir: State and variability from observations. *Journal of Geophysical Research: Oceans*, *114*(C1).

Quartly, G. D., Rinne, E., Passaro, M., Andersen, O. B., Dinardo, S., Fleury, S., Guillot, A., Hendricks, S., Kurekin, A. A., Müller, F. L., et al. (2019). Retrieving sea level and freeboard in the arctic: a review of current radar altimetry methodologies and future perspectives. *Remote Sensing*, *11*(7), 881.

Raiffa, H., & Schlaifer, R. (1961). Applied statistical decision theory. Tech. rep.

Rakitsch, B., Lippert, C., Borgwardt, K., & Stegle, O. (2013). It is all in the noise: Efficient multi-task gaussian process inference with structured residuals. *Advances in neural information processing systems*, *26*, 1466–1474.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, *66*(336), 846–850.

Rasmussen, C., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*.

Rasmussen, C. E., & Candela, J. Q. n. (2005). Healing the relevance vector machine through augmentation. In *Proceedings of the 22nd international conference on Machine learning*, (pp. 689–696).

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global environmental change*, *42*, 153–168.

Ricker, R., Girard-Ardhuin, F., Krumpen, T., & Lique, C. (2018). Satellite-derived sea ice export and its impact on arctic ice mass balance. *The Cryosphere*, *12*(9), 3017–3032.

Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., & Haas, C. (2017). A weekly arctic sea-ice thickness data record from merged cryosat-2 and smos satellite data. *Cryosphere*, *11*(4), 1607–1623.

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

Romanovsky, V. E., Smith, S. L., & Christiansen, H. H. (2010). Permafrost thermal state in the polar northern hemisphere during the international polar year 2007–2009: a synthesis. *Permafrost and Periglacial processes*, *21*(2), 106–116.

Runge, C. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten. *Zeitschrift für Mathematik und Physik*, *46*(224-243), 20.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al. (2019b). Inferring causation from time series in earth system sciences. *Nature communications*, *10*(1), 1–13.

Runge, J., Heitzig, J., Petoukhov, V., & Kurths, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical review letters*, *108*(25), 258701.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019a). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11), eaau4996.

Runge, J., Petoukhov, V., & Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of climate*, *27*(2), 720–739.

Sakov, P., & Bertino, L. (2011). Relation between two common localisation methods for the enkf. *Computational Geosciences*, *15*(2), 225–237.

Sakshaug, E., Bjørge, A., Gulliksen, B., Loeng, H., & Mehlum, F. (1994). Structure, biomass distribution, and energetics of the pelagic ecosystem in the barents sea: a synopsis. *Polar Biology*, *14*(6), 405–411.

Sarkar, D., Osborne, M. A., & Adcock, T. A. (2019). Spatio-temporal prediction of tidal currents using gaussian processes. *Journal of Geophysical Research: Oceans*.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681.

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, *10*(5), 1299–1319.

Schröder, D., Feltham, D. L., Flocco, D., & Tsamados, M. (2014). September Arctic sea-ice minimum predicted by spring melt-pond fraction. *Nature Climate Change*, *4*(5), 353–357.

Schröder, D., Feltham, D. L., Tsamados, M., Ridout, A., & Tilling, R. (2019). New insight from cryosat-2 sea ice thickness for sea ice modelling. *The Cryosphere*, *13*, 125–139.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., & Kwok, R. (2011). Uncertainty in modeled arctic sea ice volume. *Journal of Geophysical Research: Oceans*, *116*(C8).

Screen, J. A., & Simmonds, I. (2010). The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, *464*(7293), 1334–1337.

Segal, R. A., Scharien, R. K., Duerden, F., & Tam, C.-L. (2020). The best of both worlds: Connecting remote sensing and arctic communities for safe sea ice travel. *Arctic*, *73*(4), 461–484.

Serreze, M. C., & Francis, J. A. (2006). The Arctic amplification debate. *Climatic change*, *76*(3-4), 241–264.

Serreze, M. C., Maslanik, J. A., Key, J. R., Kokaly, R. F., & Robinson, D. A. (1995). Diagnosis of the record minimum in Arctic sea ice area during 1990 and associated snow cover extremes. *Geophysical Research Letters*, *22*(16), 2183–2186.

Serreze, M. C., & Stroeve, J. (2015). Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Phil. Trans. R. Soc. A*, *373*(2045), 20140159.

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

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *The annals of Statistics*, (pp. 898–916).

Slater, T., Lawrence, I. R., Otosaka, I. N., Shepherd, A., Gourmelen, N., Jakob, L., Tepes, P., Gilbert, L., & Nienow, P. (2021). Earth's ice imbalance. *The Cryosphere*, *15*(1), 233–246.

Smedsrud, L. H., Halvorsen, M. H., Stroeve, J. C., Zhang, R., & Kloster, K. (2017). Fram strait sea ice export variability and september arctic sea ice extent over the last 80 years. *The Cryosphere*, *11*(1), 65–79.

Smola, A. J., & Kondor, R. (2003). Kernels and regularization on graphs. In *Learning theory and kernel machines*, (pp. 144–158). Springer.

Sollich, P. (2005). Can Gaussian process regression be made robust against model mismatch? In *Deterministic and Statistical Methods in Machine Learning*, (pp. 199–210). Springer.

Sollich, P., & Halees, A. (2002). Learning curves for gaussian process regression: Approximations and bounds. *Neural computation*, *14*(6), 1393–1428.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., & Miller, H. (2007). Ipcc fourth assessment report (ar4). *Climate change*, *374*.

Spaans, M., Barendregt, M., Muller, E., de Beurs, E., Nijman, H., & Rinne, T. (2009). Mmpi profiles of males accused of severe crimes: A cluster analysis. *Psychology, Crime & Law*, *15*(5), 441–450.

Spielman, D. A. (2010). Algorithms, graph theory, and linear equations in laplacian matrices. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, (pp. 2698–2722). World Scientific.

Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, *9*(1), 62–72.

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.

Steele, M., Bliss, A., Peng, G., Meier, W., & Dickinson, S. (2019). Arctic sea ice seasonal change and melt. *Natl. Snow and Ice Data Cent. Freeze Climate Indicators from Satellite Data*.

Stegle, O., Lippert, C., Mooij, J. M., Larence, N., & Borgwardt, K. (2011). Efficient inference in matrix-variate gaussian models with iid observation noise.

Steinhaeuser, K., & Chawla, N. V. (2010). Identifying and evaluating community structure in complex networks. *Pattern Recognition Letters*, *31*(5), 413–421.

Steinley, D. (2004). Properties of the hubert-arable adjusted rand index. *Psychological methods*, *9*(3), 386.

Stirling, I. (1997). The importance of polynyas, ice edges, and leads to marine mammals and birds. *Journal of Marine Systems*, *10*(1-4), 9–21.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., Bréon, F.-M., Church, J. A., Cubasch, U., Emori, S., & Others (2013). Technical summary. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, (pp. 33–115). Cambridge University Press.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133.

Stroeve, J., Barrett, A., Serreze, M., & Schweiger, A. (2014b). Using records from submarine, aircraft and satellites to evaluate climate model simulations of arctic sea ice thickness. *The Cryosphere*, *8*(5), 1839–1854.

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

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., & Serreze, M. (2007). Arctic sea ice decline: Faster than forecast. *Geophysical research letters*, *34*(9).

Stroeve, J., Liston, G. E., Buzzard, S., Zhou, L., Mallett, R., Barrett, A., Tschudi, M., Tsamados, M., Itkin, P., & Stewart, J. S. (2020a). A lagrangian snow evolution system for sea ice applications (snowmodel-lg): Part ii—analyses. *Journal of Geophysical Research: Oceans*, *125*(10), e2019JC015900.

Stroeve, J., Markus, T., Boisvert, L., Miller, J., & Barrett, A. (2014a). Changes in arctic melt season and implications for sea ice loss. *Geophysical Research Letters*, *41*(4), 1216–1225.

Stroeve, J., Nandan, V., Willatt, R., Tonboe, R., Hendricks, S., Ricker, R., Mead, J., Mallett, R., Huntemann, M., Itkin, P., et al. (2020b). Surface-based ku-and ka-band polarimetric radar for sea ice studies. *The Cryosphere*, *14*(12), 4405–4426.

Stroeve, J., & Notz, D. (2015). Insights on past and future sea-ice evolution from combining observations and models. *Global and Planetary Change*, *135*, 119–132.

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

Stroeve, J., Vancoppenolle, M., Veyssière, G., Lebrun, M., Castellani, G., Babin, M., Karcher, M., Landy, J., Liston, G. E., & Wilkinson, J. (2021). A multi-sensor and modeling approach for mapping light under sea ice during the ice-growth season. *Frontiers in Marine Science*, *7*, 1253.

Stroeve, J. C., Crawford, A. D., & Stammerjohn, S. (2016). Using timing of ice retreat to predict timing of fall freeze-up in the Arctic. *Geophysical Research Letters*, *43*(12), 6332–6340.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., & Jahn, A. (2015). Influence of internal variability on arctic sea-ice trends. *Nature Climate Change*, *5*(2), 86–89.

Tangang, F. T., Tang, B., Monahan, A. H., & Hsieh, W. W. (1998). Forecasting enso events: A neural network–extended eof approach. *Journal of Climate*, *11*(1), 29–41.

Thompson, D. W., & Wallace, J. M. (1998). The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters*, *25*(9), 1297–1300.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in medicine*, *16*(4), 385–395.

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

Tilling, R. L., Ridout, A., & Shepherd, A. (2016). Near-real-time arctic sea ice thickness and volume from cryosat-2. *The Cryosphere*, *10*(5), 2003–2012.

Tilling, R. L., Ridout, A., & Shepherd, A. (2018). Estimating arctic sea ice thickness and volume using cryosat-2 radar altimeter data. *Advances in Space Research*, *62*(6), 1203–1225.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, *1*(Jun), 211–244.

Tipping, M. E., & Faul, A. C. (2003). Fast marginal likelihood maximisation for sparse bayesian models. In *International workshop on artificial intelligence and statistics*, (pp. 276–283). PMLR.

Toudal, L., & Coon, M. D. (2001). Interannual variability of the sea-ice-induced salt flux in the greenland sea. *Annals of Glaciology*, *33*, 385–390.

Tschudi, M., Meier, W., Stewart, J., Fowler, C., & Maslanik, J. (2019). Ease-grid sea ice age, version 4.[1984–2020]. *Boulder, CL: NASA National Snow and Ice Data Center Distributed Active Archive Center)*, *10*.

Tschudi, M. A., Stroeve, J. C., & Stewart, J. S. (2016). Relating the age of arctic sea ice to its thickness, as measured during nasa's icesat and icebridge campaigns. *Remote Sensing*, *8*(6), 457.

Tsonis, A. A., & Roebber, P. J. (2004). The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, *333*(1-4), 497–504.

Tsonis, A. A., Swanson, K. L., & Roebber, P. J. (2006). What do networks have to do with climate? *Bulletin of the American Meteorological Society*, *87*(5), 585–596.

Urry, M. J., & Sollich, P. (2013). Random walk kernels and learning curves for gaussian process regression on random graphs. *The Journal of Machine Learning Research*, *14*(1), 1801–1835.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011). The representative concentration pathways: an overview. *Climatic change*, *109*(1), 5–31.

Van Vuuren, D. P., Kriegler, E., O'Neill, B. C., Ebi, K. L., Riahi, K., Carter, T. R., Edmonds, J., Hallegatte, S., Kram, T., Mathur, R., et al. (2014). A new scenario framework for climate change research: scenario matrix architecture. *Climatic Change*, *122*(3), 373–386.

Venegas, S. A., & Mysak, L. A. (2000). Is there a dominant timescale of natural climate variability in the arctic? *Journal of Climate*, *13*(19), 3412–3434.

Vinje, T. (2001). Anomalies and trends of sea-ice extent and atmospheric circulation in the nordic seas during the period 1864–1998. *Journal of Climate*, *14*(3), 255–267.

Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., & Borgwardt, K. M. (2010). Graph kernels. *Journal of Machine Learning Research*, *11*(Apr), 1201–1242.

Walker, G., & Bliss, E. W. (1932). World Weather. *V. Memoirs Royal Meteorological Society*, *4*, 53–84.

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, *109*(4), 784–812.

Walsh, J. E. (1980). Empirical orthogonal functions and the statistical predictability of sea ice extent. *Sea ice processes and models. University of Washington Press, Seattle*, (pp. 373–384).

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

Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I., & Colony, R. (1999). Snow depth on arctic sea ice. *Journal of Climate*, *12*(6), 1814–1829.

Watts, M., Maslowski, W., Lee, Y. J., Kinney, J. C., & Osinski, R. (2021). A spatial evaluation of arctic sea ice and regional limitations in cmip6 historical simulations. *Journal of Climate*, *34*(15), 6399–6420.

Willatt, R., Laxon, S., Giles, K., Cullen, R., Haas, C., & Helm, V. (2011). Ku-band radar penetration into snow cover on arctic sea ice using airborne data. *Annals of Glaciology*, *52*(57), 197–205.

Williams, J., Tremblay, B., Newton, R., & Allard, R. (2016). Dynamic preconditioning of the minimum september sea-ice extent. *Journal of Climate*, *29*(16), 5879–5891.

Wingham, D., Francis, C., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., Mallow, U., Mavrocordatos, C., et al. (2006). Cryosat: A mission to determine the fluctuations in earth's land and marine ice fields. *Advances in Space Research*, *37*(4), 841–871.

Yuan, X., Chen, D., Li, C., Wang, L., & Wang, W. (2016). Arctic sea ice seasonal prediction by a linear markov model. *Journal of Climate*, *29*(22), 8151–8173.

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

Zhang, R. (2015). Mechanisms for low-frequency variability of summer arctic sea ice extent. *Proceedings of the National Academy of Sciences*, *112*(15), 4570–4575.

Zhang, X., Sorteberg, A., Zhang, J., Gerdes, R., & Comiso, J. C. (2008). Recent radical shifts of atmospheric circulations and rapid changes in arctic climate system. *Geophysical Research Letters*, *35*(22).

Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). Enso-like interdecadal variability: 1900–93. *Journal of climate*, *10*(5), 1004–1020.

Zhang, Y.-F., Bitz, C. M., Anderson, J. L., Collins, N., Hendricks, J., Hoar, T., Raeder, K., & Massonnet, F. (2018). Insights on sea ice data assimilation from perfect model observing system simulation experiments. *Journal of Climate*, *31*(15), 5911–5926.

Zhang, Y.-F., Bushuk, M., Winton, M., Hurlin, B., Yang, X., Delworth, T., & Jia, L. (2021). Assimilation of satellite-retrieved sea ice concentration and prospects for september predictions of arctic sea ice. *Journal of Climate*, *34*(6), 2107–2126.

Zhou, C., Zemanová, L., Zamora-Lopez, G., Hilgetag, C. C., & Kurths, J. (2007). Structure–function relationship in complex brain networks expressed by hierarchical synchronization. *New Journal of Physics*, *9*(6), 178.