Ice Edge Verification

Measuring the skill in our Forecasts and disagreement in our Observations

Thesis

submitted in fulfilment of the requirements for the Degree of

Doctor rerum naturalium

By

Bimochan Niraula

FB-01, Universität Bremen
Bremen, Germany

Promotionskolloquium:	14 April, 2023
First reviewer:	Prof. Dr. Thomas Jung Alfred Wegener Institute Helmholtz-Zentrum für Polar- und Meeresforschung
Second reviewer:	Prof. Dr. Gunnar Spreen Universität Bremen Institut für Umweltphysik
Supervisor:	Dr. Helge F. Goessling Alfred Wegener Institute Helmholtz-Zentrum für Polar- und Meeresforschung

Abstract

Sea ice is one of the most crucial components of the polar system. Alongside its effects on the climate and weather, it also impacts human lives and operations in these regions. Climate scientists, as well as other stakeholders (such as marine industries, mining operations, governmental authorities as well as local/aboriginal population), have a high interest in getting accurate observation and prediction of sea ice presence, and its effective border – the ice edge. In this context, this thesis is focused on the verification of ice presence and ice edge, across different datasets. We do this by describing a new method of generating reference forecasts of the ice edge (as benchmark for predictions at sub-seasonal to seasonal timescales), analysing initial state errors in forecasts and analysis from ECMWF, and comparing the ice presence and ice edge difference between several observational and analysis datasets.

Operational systems generating sea ice forecasts at sub-seasonal timescales are mostly compared against simple references based solely on climatological or initial states, which can lead to a potential overestimation of their prediction skill. In chapter 1, we describe the Spatial Damped Anomaly Persistence (SDAP) method, which combines historical sea ice probability with the initial ice edge anomalies to generate probabilistic reference forecasts of the ice edge. The SDAP forecasts outperform both traditional references, as well as most dynamical forecast models from the Sub-seasonal to Seasonal (S2S) project at long lead times, establishing a more challenging benchmark for operational forecast systems.

Within the S2S database, forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) have the highest prediction skill compared to both traditional references and SDAP forecasts, but show significant initial error. In chapter 2, we analyse this initial state issue in the ECMWF forecasts, as well as corresponding analysis from ORAS5, by measuring the errors against observations from OSISAF. We find that the initial state errors are partly due to interpolation issues, and partly systematic underestimation, especially in the summer. We show the spatial distribution of the mean bias and provide evidence that between 10 to

20% of the error can be reduced simply by subtracting the mean bias.

While forecast verification methods assume observations to be 'the truth', observational records also have inconsistencies that have previously only been discussed in concentration or ice extent terms. In chapter 3, we analyse differences in ice presence between several observational and analysis datasets by measuring the Integrated Ice Edge Error and bias between each pair. We find significant mismatch between observations, particularly in the summer, and identify regions where certain observations disagrees with all other. Our results show that observational records from OSISAF potentially overestimate ice presence, while those from AMSR-E/2 potentially underestimate it.

In the last chapter, we discuss the results from the different studies in context of each other and mention how the inconsistencies in observations cast doubt on the forecast errors we initially measured, offering the possibility of a probabilistic observation and reemphasizing the need for accurate sea ice edge measurements and forecasts.

Contents

Man wants to know, and when he ceases to do so, he is no longer a man.

Fridtjof Nansen

Introduction

I would like to start this thesis by describing, in simple terms, the physical processes that typically take place every year on the surface of the polar oceans. Let us imagine ourselves somewhere north of Greenland at the end of the boreal summer season. With each day, the number of daylight hours are decreasing and it is getting colder over the entire Arctic. As the ambient temperature falls below freezing, water at the top of the oceanic layer slowly starts turning into ice. The ice clumps into floes that might initially be thin and isolated, but as the temperature gets colder, the floes increase in size and thickness, and coalesce with other floes to cover large areas of the ocean. By February, most of the Arctic Ocean is covered by a frozen layer.

This frozen ocean cover is sea ice.

FIGURE 1: Sea ice in the fjords of Svalbard during Spring 2022, showing different scales and stages of formation. Picture taken by the author.

0.1 Polar processes

The simplistic description given so far skips a lot of important details regarding the evolution of sea ice. While most of the Arctic does get covered with ice over winter, it is not a constant layer. It is perpetually evolving at the whims of the ocean and the atmosphere, expressed as both dynamic (movement/force related) and thermodynamic (temperature/heat related) processes (Petrich and Eicken, 2016). The floes drift, converge and diverge, leading to cracks, leads and ridges on the ice surface. There is a constant push and pull from both wind and ocean currents, while the topography of the land and ocean bottom also place barriers on how the ice can shift. There are various storms, temperature fluctuations, freshwater influx and precipitation events, each locally affecting the ice cover. On a bigger scale, the effects of global radiative forcing and the recent changes associated with it are also reflected by a changing sea ice cover, as we will discuss below.

Up to this point, I have only described the processes that take place from late summer to the end of winter, but what happens afterwards is equally important. By the end of April, the Arctic starts to be exposed to more and more sunlight, and the ambient temperature increases likewise. The ice cover starts to melt, the surface slowly covered by the melt-water, and the large floes break into smaller pieces, leading to increasingly larger gaps. The amount of ice keeps decreasing as we go further into the summer until it reaches a minimum right at the turn of the season.

However, not all of the ice melts away. Thicker ice floes, usually those that have already survived previous summers, can last through the summer into the next winter (Haas, 2016). Much of the ocean north of Canada and Greenland has historically been ice-covered throughout the summer, although the actual cover is changing fast due to climate change. How much sea ice lasts over the summer is a matter of immense importance and its connections to the climate are manifold, as we will learn further below.

Although the previous paragraphs mentioned only the Arctic, sea-ice is present at both poles (Meredith and Brandon, 2016). The fundamental physics governing the evolution of sea-ice is the same in Antarctica as it is in the Arctic - freezing over winter, thawing through spring and melting in the summer. However, the coastal topography at the two poles are very different, and this affects the sea ice distribution as well. The Arctic is an ocean surrounded by continents on most sides, with only a few exchange gates with the other oceans. Antarctica, on the other

hand, is a continent surrounded by the Southern Ocean. There is more exchange of heat and water from the subtropics and more of the ice can be carried away from the coast. Alongside, the thermal halocline isolating the ice from the warm deep waters in the Arctic, is not present in the southern hemisphere. Consequently, less of the ice survives through the austral summer and the overall ice cover in the southern hemisphere is younger and thinner compared to the northern hemisphere.

0.2 Sea ice in the climate

The presence of sea ice, in both hemispheres, is deeply intertwined with the global climate system. The forward relation is already quite evident – from freezing to melting, sea ice is modulated by the dynamic and thermodynamic shifts in the atmosphere and the ocean, which are affected by the polar and global climate at a large scale. Another world with different climate conditions would have drastically different sea ice (Yang et al., 2019), and a warming world has already brought extensive changes to the sea ice coverage (Landrum and Holland, 2020).

The relation is equally important in the other direction. The freezing of sea ice affects the salinity and vertical mixing of ocean layers (Liang and Losch, 2018), while the cover insulates the ocean surface from the atmosphere (Budikova, 2009). It affects both atmospheric and oceanic circulations in the polar regions (Aagaard and Carmack, 1989). Most importantly, the high albedo of sea ice enables a large portion of the incoming solar radiation to be reflected back to space, and a decrease in the sea ice cover causes an increase in the amount of sunlight trapped within the atmosphere (Perovich and Polashenski, 2012).

As it happens, the global climate is changing and the sea ice cover is indeed decreasing, particularly in the Arctic. This is seen in many of the sea ice related variables (which we will describe below) – Arctic ice extent and ice area is decreasing, mean thickness is lower and sea ice volume loss is quite high (Comiso et al., 2008; Parkinson and Cavalieri, 2008; Stroeve and Notz, 2018). The effects of climate change manifest more acutely in the Arctic than elsewhere on earth, a process termed "Arctic Amplification" (not just due to the albedo change, but also an effect of the local temperature Lapse rate; see Pithan and Mauritsen, 2014). As the Arctic warms, the sea ice melts and the albedo of the Arctic decreases, abosrbing more sunlight, in turn further warming the Arctic (and the planet as a whole), melting the ice more and further continuing this positive feedback cycle. Studies have already established that the current climate conditions have brought about a 'new Arctic'

(Landrum and Holland, 2020) and it is predicted that summer in the Arctic will be completely ice free by 2050 (Årthun et al., 2021; Wang and Overland, 2009).

0.3 Studying sea ice

While we have described sea ice and its role in the climate so far, let us now take a step back and examine how it is analyzed and represented in scientific contexts. There are two main methodologies driving most sea ice related studies: observations and numerical modeling. Observations themself can involve in situ research, or remote sensing. Furthermore, there are also some studies employing lab-based methods to better understand sea ice (e.g. using ice cores; see Abram et al., 2013). In this thesis however, we are only using data from remote sensing and modeling approaches, in order to have complete coverage.

Before going in depth about these methods however, it is important to acknowledge that trying to understand sea ice did not start with modern climate research. Indigenous people that have been living north of the Arctic Circle have been keen observers of the sea ice condition in their surroundings for centuries (Usher, 1971). They keep track of the growth and melt seasons, tidal movements and the thickness of ice floes, to be able to safely travel and hunt on the ice and fish in the polynyas (Huntington et al., 2016). Much of their research, and their methods, are only recently coming to light in wider discourse (Eicken, 2010). Similarly, the Vikings are known to have monitored sea ice conditions as it related to their voyages and settlements in the north Atlantic (Haine, 2008).

There were waves of voyages into the Arctic and Antarctic in the 18th and 19th centuries, bringing with them troves of knowledge about the presence and persistence of sea ice in the polar regions (MacLaren, 1994). Nansen's Fram expedition in the 1890s was a high point of scientific endeavour and discovery (Johnson, 1983). While wars raged around the world, scientists continued learning more about sea ice. Overhead military flights and under ice submarine voyages, especially in the cold war era, kept logs that would later be of great scientific value (Kwok and Rothrock, 2009). Several field campaigns have been instrumental in improving our knowledge of sea ice and in 2019-2020, the MOSAiC expedition spent a year in the Arctic (Shupe et al., 2020), representing what might be the pinnacle of polar research so far.

Remote sensing

In 1858, French artist and photographer Nadar took an aerial photograph of Paris from a balloon (Bann, 2009), effectively giving birth to the concept of remote sensing. It took several years before this name actually got used, but in the intermediate period the process got widely accepted and applied. Balloons, planes, rockets, and even pigeons took to the sky, capturing an aerial view of the planet (Verhoeven, 2009). Satellite based observations, initially limited to military use, began in the 50s. Then in 1972, Landsat-1 was launched, the first satellite actually made for scientific research. Now, there are several remote sensing instruments on different satellite platforms, providing high resolution records of various sea ice variables (Carsey, 1989; Kwok, 2010), some of which are used in this thesis.

Technological development has made satellite-based imagery a familiar concept to us all and it is possible to view remote locations via the "eye in the sky" almost in real-time (a visit to http://observer.farearth.com/observer/ can be recommended for some fascinating examples). Retrieving quantified measurements of sea ice condition is, however, not trivial as the sensors on a satellite platform do not observe the planetary surface the same way a person at the poles would. We refer the interested reader to Carsey (1989) and Spreen and Kern (2016) for detailed descriptions of the steps involved in remote sensing of sea ice and give a very simple overview here.

Various satellites carry passive microwave (PMW) sensors on board, which are capable of observing the thermal radiation emitted from the planetary surface at different frequencies. In the polar regions, radiation emitted from the sea ice surface has a very different signature (in terms of brightness temperature) compared to that from the ocean. Using openings in the ice ("tie-points") and the polarization of the sensor to calibrate this difference, different algorithms can convert it into a measurement of ice presence. Fortunately, the PMW sensors can see through the clouds very well, although atmospheric conditions do affect the measurements (Spreen and Kern (2016)), which are addressed by different correction schemes. There are different PMW sensors and algorithms in use, and the various combinations result in different datasets of sea ice measurement. We will see some of the more widely used datasets in chapter 3.

There are also other (non-PMW) sensors aboard satellites, as well as aircraft or ship based remote sensing platforms. Some of these instruments measure different sea ice variables than the fraction of ice cover. In this thesis however, we are focused solely on the edge and presence of sea ice, and are interested in long term pan-hemispheric comparisons. Ship, aircraft or ground based observations, which are

generally limited in their coverage, have therefore not been included in our studies. One final point that needs emphasis, is that all of these datasets are measurements of the true condition and just as any measurement, can have offsets from the 'truth' as we will discuss further in the thesis.

Numerical modeling

The second wheel driving forward our knowledge of sea ice is numerical modelling. Models fill in a lot of gaps in our understanding of sea ice, and climate research as a whole, as observations are limited in nature, do not generally allow for experimentation and give us a single state. In models we can add, omit or change various variables and processes, in order to simulate the past or present, or to make forecasts and future projections.

Since the first conception of a mature sea-ice model in 1971 (Maykut and Untersteiner, 1971), experiments could compare observational results with model output to further our understanding of sea-ice physics and processes. This led to further improvements in model development, increasing their use case. During the 80s, they started to be coupled with the ocean model (allowing for exchange of variables) and included within global climate models (Huntington et al., 2016). Over time, there has been a rapid expansion in the variety of their application, from 1D single column modeling to automated prediction systems that have high resolution, oceanic processes and coupling between several components (Lemieux et al., 2015; Johnson et al., 2019; Zuo et al., 2019). It is this last type of model that we are concerned with in this thesis.

Modern prediction systems

While the use of sea ice models for forecasting purposes is a recent development (Guemas et al., 2014), there are now several operational and research centers around the world producing sea ice prediction at various timescales (Vitart et al., 2017). Fig.2 (derived from Carrieres, 2017) gives a simple depiction of the various components that may be included in a modern prediction system, although the individual setup of the prediction system can differ depending on the institution.

Forecasts are often started from the initial conditions for a short cycle to find the background state. This is then corrected by incorporating observations by the Data Assimilation system (described below). The resulting output (called the analysis) is then used by the forecast system and the underlying processes are then

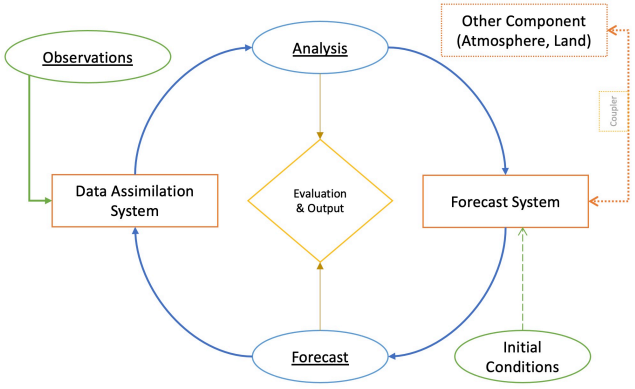

FIGURE 2: Schematic showing various components of a modern prediction system (derived from Carrieres, 2017). In this thesis, we are focusing on verification of Observation, Analysis and Forecast.

run for a given number of time steps (or iterations) to derive the actual forecasts (Carrieres, 2017). The (ice-ocean) forecast system can also be coupled with other components (such as the land or atmospheric models), in order to pass variables at chosen intervals of model run, depending on the system.

Data Assimilation (DA), as the name suggests, is the process of bringing together data from several sources, essentially integrating observations with model runs using minimization operations (Mu et al., 2020). Through DA, we can take incomplete and erroneous observations and fill the spatio-temporal gaps with the help of model physics (rather than simple interpolation). The resulting analysis is then a more complete and accurate picture of the initial state, which can be used by the forecast system to generate forecasts. Several efforts have made to use available observations over long periods, often retrospectively, to build a complete (and corrected) collection of atmospheric and oceanic variables (especially well established for atmospheric variables; see Hersbach et al., 2020). Such reanalysis (as they are called) are typically done with a fixed model version, so that it is more consistent over time and can also be used for climate-related studies. Other analysis are more "operational", designed with a particular target or to be fed into operational forecasts (Zuo et al., 2019).

In operational forecast centers, there is currently a push for improving prediction capabilities at sub-seasonal to seasonal (S2S) timescales (Vitart et al., 2017). This is especially pertinent for marine operations in the polar regions, as decreasing ice cover and increasing economic activities have brough an increase in marine traffic

to the polar regions (Palmer, 2013). The role of sea ice in modulating S2S predictability (including for atmospheric variables) has also highlighted the importance of improving the forecast skills in our ice-ocean models (a more detailed description, given by Chevallier et al., 2019, is recommended as an excellent overview). Perfect model studies have shown that sea ice extent is predictable at interannual timescales (Day et al., 2016), although a 'spring barrier' for predictability can lead to lower predictive skills in the summer (Bushuk et al., 2020). In this context, we will analyse forecasts made at S2S timescales in Chapter 1 of this thesis.

Sea ice variables

Data from both sources, observations or numerical models, are produced on a 2D representation of the planetary surface, divided into individual grid cells. The grid cells are the basic area units of the region of interest (most likely the poles in our case), akin to a pixel in graphical displays. The actual arrangement of such cells is determined by the grid and resolution of the observational data or the model and can range from regular lat-lon grid, to hexagonal (Pinori et al., 2008) or even an irregular mesh (Sein et al., 2017).

This short description of grid distribution can help us understand a very important concept in the sea ice community (and this thesis) – that of sea ice concentration (SIC). In simple words, sea ice concentration is the percentage of a grid cell that is covered by ice. This is a very important variable in both modeling and observational applications, allowing us to record or model cells that have both liquid water and ice within it. This formulation ignores the thickness or distribution of sea ice within the grid cell, but results in a standardized form of measuring ice cover.

Whether or not a grid cell is 'ice-covered' is generally determined by setting a threshold percentage and assigning every cell with concentration equal to or higher than that threshold to be "ice-covered". This transforms the measured or modeled concentration value into a binary field – ice or no ice. Any end user needing to decide whether or not a certain region has ice can then look at this binary field and clearly distinguish which grid cells are frozen. The threshold in use can also differ based on the user's needs – a hunting party traveling on top of the ice may prefer a low threshold, while an ice-breaker ship might accept a higher threshold. It is common practice in the polar science community to use 15% concentration as the standard threshold for ice presence, but this can differ, as we will see further in the thesis.

Following this, we can now learn about ice extent and ice area. Ice extent is

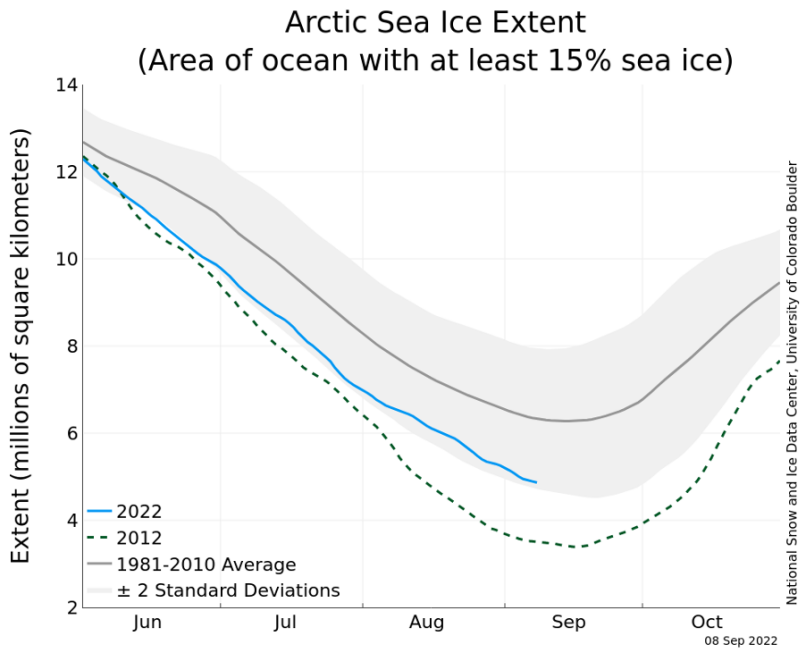

FIGURE 3: Arctic Sea Ice Extent measurements from NSIDC (2022), showing daily time-series for 2022 (as of 8 Sep 2022), alongside the record low year of 2012 and the 1981 to 2010 average.

the sum of area of all the grid cells that are considered to be ice covered (based on the threshold). This results in an area measurement, usually pan-hemispheric, which can be compared over time, or across different model systems. Sea ice extent is different from another commonly discussed value, sea ice area, which is measured by multiplying the concentration of each grid cell by its area and taking the overall sum. Both area and extent are very useful measurements for understanding the seasonal cycle over each year (as the poles freeze and thaw), as well as year to year changes and long-term trends in sea ice (as seen in Fig.3 from NSIDC, 2022).

Similarly, the contour of the threshold concentration gives us the ice edge, which is the effective border of ice presence. This is a very important variable, and in fact the focus of this entire thesis. The ice edge is of immense interest, not just to scientists, but also other stakeholders, as it allows for a clear demarcation of the ice-covered seas. The ice edge is used by several governmental bodies in legal and policy making contexts (Bay-Larsen et al., 2020; Veland and Lynch, 2016). Therefore, verifying that the ice edge is accurate (in both observation and forecast)

is very important.

0.4 Ice Edge verification

When we talk about forecasts, the first question that comes to mind is "How good are they?". To find the answer, we take the forecasts and the corresponding observations, and measure the accuracy with a verification metric. For ice-edge, two very useful metrics in use are the Integrated Ice Edge Error (IIEE; Goessling et al., 2016), and its probabilistic counterpart - the Spatial Probability Score (SPS; Goessling and Jung, 2018).We will describe both metrics fully in further chapters, but we can briefly mention here that the IIEE is the total area of mismatch between the forecast and observation of ice presence. This applies for deterministic (or binary) forecasts and observations. For probabilistic data, IIEE is replaced by SPS, which also considers the probability of ice presence (and can be considered the spatial integral of the Brier score). In practice, the observational ice edge is mostly a deterministic state, derived from the concentration threshold approach, which means the probability of ice presence can be assumed to always be either 0 (no ice) or 100% (ice). For an ensemble output, each ensemble member gives a similar deterministic state and simply averaging over them gives the probability (although other means of determining a probability of ice presence exist, as we will see in chapter 1). Similar to observations, a deterministic forecast can also be considered to be a probabilistic forecast with binary values (0 or 100%), making the SPS and IIEE equal in that context.

There are also other verification metrics that can be used to measure the accuracy of a forecast, the most important one being the Modified Haussdorff Distance (MHD; Dukhovskoy et al., 2015; Casati et al., 2022). MHD will also be described more in detail in chapter 1, but for a brief description - it is a distance measurement between two ice edges, used to find how far apart the forecasted and observed ice edges are. Measuring the MHD requires determining every point on the two ice edges and measuring the distance between them. However, this is not a trivial operation.

The ice cover is not uniform – it can have pockets of low concentration deep within the ice pack, or be broken up by islands or the coast. Consequently, the ice edge is generally not one continuous line, but consists of several smaller ice floes. Measuring the MHD means we must consider which edge points to include, how to measure the distance between each of them, how to handle the coast or openings in the ice, and how to keep the domain of comparison constant between different datasets and dates. In our study, these issues often resulted in us cutting out sections

of the data, either spatially or temporally. While we did measure and compare the MHD for the forecasts in chapter 1, we decided in the other chapters to prioritise having a more coherent timeseries of analysis and thus not use MHD. There have been promising developments in new methods for measuring the MHD (Casati et al., 2022) and while it has not been included here, we hope to follow up some of our work by including MHD alongside the other metrics.

For each of the metrics, knowing the verification score or distance gives an idea of how erroneous the forecast is, but does not by itself tell us if it is good enough for our needs. For this reason, verification metrics are more of a tool than the result by itself. We can judge the performance of a model forecast by comparing their resulting metric against that of another model, another version of the same model, or some other 'reference' data. The choice of this reference data in measuring forecast skill is quite important, as a weak reference can lead to overestimation of forecast skill. This search for a simple, reproducible and skillful reference forecast essentially gave birth to this entire thesis.

0.5 Scope of this dissertation

As we have established in the previous sections, this PhD dissertation is focused on the sea ice edge. This is a variable of crucial importance for operational needs and there is strong momentum to verify that both our observations (current and past), as well as forecasts (for the future) are as accurate as possible. This overarching theme has been explored here in three different but closely linked ice edge verification studies.

For operational forecasting centers, there has been an increase in user demand for skillful seasonal forecasts of sea ice condition, particularly in the Arctic. Most studies have measured this skill by comparing the forecast performance against two simple references- either historical patterns, or a fixed initial state. While there has been progress in advancing these references by incorporating prior trends or anomalies into them, their application with the ice edge has been limited and the use of probabilistic anomaly has not been explored so far. We describe a simple method of combining the historical probabilities with the anomalies of the initial ice state through Spatial Damped Anomaly Persistence (SDAP), resulting in an ice edge forecast that is skillful enough to be used as a challenging benchmark for sub-seasonal to seasonal (S2S) operational forecasts. It is a very simple method relying solely on historical and initial ice presence, yet the forecast skill is comparable to

those of fully coupled dynamical models. This study (published as Niraula and Goessling, 2021) is presented here as **Chapter 1**, with the scientific objectives as follows:

Q1 Can we create a simple reference forecast of sea-ice edge combining probabilistic information from both historical and current ice conditions?

Q2 Can such a forecast perform well in comparison to traditional references and S2S forecasts?

In chapter 1, we find that the ice edge forecasts from the European Centre for Medium- Range Weather Forecasts (ECMWF) are the only ones from the S2S database to perform better than the SDAP benchmark. However, they start with a high initial error, roughly half as high as climatological error. This has also been shown by other studies in the past and motivated us to systematically measure the ice edge error in the initial state of the forecast, alongside the analysis data that was used in the forecast setup (ORAS5, described more within the chapter), against observations from the Ocean and Sea Ice Satellite Application Facility (OSI SAF). We analyse the spatial and temporal patterns in the errors and show that the initial state error is a combination of interpolation error and seasonal bias in both forecast and analysis. This work is presented here as **Chapter 2**, with the following scientific objectives:

Q1 How well do the initial ice-edge states of ECMWF forecasts and ORAS5 analysis agree with observational data from OSI SAF?

Q2 Is there a consistent bias pattern that can explain the error of these ice states?

The results from chapter 1 show significant mismatch in ice edge between the ORAS5 analysis and OSI SAF observations. This has prompted us to ask whether some of the errors actually issue from OSI SAF, and how well does OSISAF itself agree with other observational datasets. What follows is an intercomparison study, where we analyse a collection of satellite derived observational and analysis

datasets of sea ice, measuring the pairwise IIEE against each other. These are well established, widely used datasets that represent different retrieval algorithms, measurement instruments, and production methods while still having pan hemispheric daily coverage over many years. We compare the bias between the datasets and find regions where certain datasets are substantially further from the consensus. This work is presented here as **Chapter 3**, with a simple scientific objective:

Q1 How well do different satellite based observational and analysis datasets of ice-edge agree with each other?

The chapters, as described above, were written as independent manuscripts to be submitted to scientific journals (at the time of submission, chapter 1 has already been published, while the other chapters are being prepared for publication). As a result, there is some overlap between them, particularly in the methods, and each chapter has its own Abstract, Introduction and Conclusion sections. Similarly, many important concepts have not been fully discussed in this introduction, but rather left to the chapters. The chapters are presented in a chronological manner, showing how our own ideas regarding ice edge verification have evolved over time. Some of the findings from the latter chapters will therefore challenge the results from earlier ones. We will briefly discuss this issue in the last chapter, concluding this thesis.

By three methods we may learn wisdom: First, by reflection. Second, by imitation. And third, by experience.

Confucius

1

Spatial Damped Anomaly Persistence of the sea-ice edge as a benchmark for Dynamical forecast systems

Abstract

Accelerated loss of the sea-ice cover and increased human activities in the Arctic emphasize the need for skillful prediction of sea-ice conditions at sub-seasonal to seasonal (S2S) time scales. To assess the quality of predictions, dynamical forecast systems can be benchmarked against reference forecasts based on present and past observations of the ice edge. However, the simplest types of reference forecasts – persistence of the present state and climatology – do not exploit the observations optimally and thus lead to an overestimation of forecast skill. For spatial objects such as the ice-edge location, the development of damped-persistence forecasts that combine persistence and climatology in a meaningful way poses a challenge. We

This chapter has been published in 'Journal of Geophysical Research: Oceans' under the same title as (Niraula and Goessling, 2021)

have developed a probabilistic reference forecast method that combines the climatologically derived probability of ice presence with initial anomalies of the ice-edge location, both derived from satellite sea-ice concentration data. No other observations, such as sea-surface temperature or sea-ice thickness, are used. We have tested and optimised the method based on minimization of the Spatial Probability Score. The resulting Spatial Damped Anomaly Persistence forecasts clearly outperform both simple persistence and climatology at sub-seasonal timescales. The benchmark is about as skilful as the best-performing dynamical forecast system in the S2S database. Despite using only sea-ice concentration observations, the method provides a challenging benchmark to assess the added value of dynamical forecast systems.

1.1 Introduction

Accelerated sea ice loss and the possibility of ice-free summers in the Arctic has increased the interest in potential human activities in the far North (Stephenson et al., 2011). To address the planning and safety concerns associated with this, government and private agencies need better predictions of sea-ice at sub-seasonal to seasonal time-scales (Jung et al., 2016). Over the past few years, many operational centers are already starting to provide such forecasts with longer lead times, although the skill of these forecasts – and how to assess the skill in the first place – is still under question (Smith et al., 2015).

There are numerous metrics to measure and quantify the accuracy of a forecast against observation, or 'true' conditions, depending on the variable in question (Wilks, 2019). Whether or not forecasts are considered skillful depends not only on the metric to measure the forecast error, but also what benchmark is used to measure skill against. The skill of the forecast produced by a particular forecast system can be compared against that of an earlier version of the same system (e.g., Balan-Sarojini et al., 2019), a different forecast system (e.g., Zampieri et al., 2018, 2019), and also against a simpler reference forecast for benchmark (e.g., Woert et al., 2004; Pohlmann et al., 2004) .

Model outputs are commonly compared against two observation-based reference forecasts: Climatology and Persistence. Climatology is based on the historical records for the given time of the year. Depending on the variable, it can be a simple mean, or a probabilistic estimate of a binary target, as described below. In the presence of a significant seasonal cycle and for lead times longer than just a

few days, a climatological forecast needs to change with lead time according to the evolving time of the year. Persistence, on the other hand, is maintaining the initial state of the variable - thus giving a constant output of the variable. A persistence forecast can also be constructed such that the seasonal evolution, estimated from previous years, is taken into account, resulting in an anomaly persistence forecast. For either of these benchmark approaches, a secular trend can be taken into account to derive a trend-adjusted variant of such a forecast (Van den Dool et al., 2006). By design, persistence has better forecast skill at shorter lead times, whereas climatology is more skillful at longer lead times when errors approach a saturation level due to chaotic error growth (Woert et al., 2004). Finally, a damped anomaly persistence forecast attempts to combine persistence and climatology in such a way that it gradually transitions from the persisted to the climatological state, thereby optimising the skill of the forecast at intermediate lead times (Van den Dool et al., 2006; Wayand et al., 2019).

Anomaly persistence and damped anomaly persistence can be applied easily to continuous quantities on a grid-cell per grid-cell basis and have been used as benchmark forecasts for quantities such as sea-surface temperature on decadal timescales (e.g., Pohlmann et al., 2004) and sea-ice concentration on seasonal timescales (e.g., Wayand et al., 2019). However, for spatial objects such as the ice-edge location (that corresponds to a binary, rather than a continuous, gridded field), it can be difficult to establish meaningful autocorrelations at longer time scales; anomalies in this case tend to migrate spatially with the seasonal cycle (e.g., Goessling et al., 2016) , which means that initial anomalies at one location may not be relevant at the same location after some time, but they might still hold information about anomalies at a different nearby location. This spatial migration is not taken into account by existing approaches of damped anomaly persistence for sea-ice concentration: The damped-anomaly benchmark in Wayand et al. (2019) applies an exponential decay of local sea-ice concentration anomalies, and in the 'ECMWFpres' system (detailed below) the initial sea-ice concentration is kept fixed for 15 days and then relaxed toward climatology. In contrast to these grid-cell by grid-cell methods, our approach utilises sea-ice anomalies for prediction in a more thorough way by transferring information across the domain and not just at increasing lead times.

In this study, we focus on sea-ice edge and sea-ice presence. For any daily dataset of sea-ice concentration (SIC), sea ice is considered to be present in all grid points with 15% or higher ice concentration. This gives a binary 'ice vs no-ice' map for each day, and the 15% concentration contour effectively gives the ice edge. For a model

output with multiple ensemble members, the ensemble mean of such binary maps results in a probabilistic measure of ice presence, termed sea-ice probability (SIP), for each day. Similarly, taking the mean of such binary ice presence maps for the same date over a number of years in the past results in the climatological SIP for the date. While it is possible to create a climatological forecast using the concentration output from models, we will use the terms climatology and climatological forecast (CLIM) in this manuscript to refer only to the climatological SIP derived from the satellite-based concentration records, which is described in section 1.2.1. Similarly, to derive a binary ice edge from a probabilistic forecast, the median contour (where SIP= 50%) is used to determine the boundary of ice presence, as used in section 1.5.

In the case of ice edge, an important variable for marine activities in the Arctic, the mismatch between predicted and 'true' ice edges can be quantified using the Spatial Probability Score (SPS; Goessling and Jung, 2018). The SPS (and its deterministic counterpart, the Integrated Ice Edge Error; see Goessling et al., 2016) determines the area where ice is either under-forecast or over-forecast in comparison to the true outcome. (Palerme et al., 2019) compared the SPS to the (modified) Hausdorff distance (MHD; Dukhovskoy et al., 2015), another commonly used verification metric, and determined that the SPS is more robust and less affected by isolated patches of ice. Therefore, we will primarily be using SPS as the comparison metric in this paper, but will also use MHD to test whether our results are robust with respect to the choice of the metric.

In terms of lead-time, Zampieri et al. (2018, 2019) have shown that some operational subseasonal-to-seasonal (S2S) forecast systems are more skillful than climatology in predicting the location of the ice edge several weeks ahead. After 1.5 months, however, even the most skilful dynamical forecast system is not performing better than climatology. This can be compared against perfect-model studies that suggest that the ice-edge position can be predictable up to 6 months ahead (Goessling et al., 2016), suggesting that forecast calibration (not applied in Zampieri et al., 2018, 2019) and/or forecast system improvements should in principle be possible. Moreover, simple initial-state persistence tends to outperform these uncalibrated dynamical systems for at least the first 3-4 days, leaving a temporal window between about 4 to 45 days where the best system can be considered skilful beyond the simple benchmark methods (Zampieri et al., 2018, 2019). However, given the simplicity of strict initial-state persistence and climatology, the term skilful might be complacent in the sense that a less naive, but still simple, benchmark forecast

method building on the concept of damped anomaly persistence may exhibit similar skill or even outperform existing dynamical forecast systems.

In this context, we have developed a method for predicting the location of the sea-ice edge that combines the climatologically derived probability of ice presence with initial (present) anomalies of the ice edge. In contrast to previous damped anomaly persistence methods that propagate local sea-ice anomalies forward in time on a per-grid-cell basis, our method propagates information also spatially from the location of the initial ice edge to nearby locations where corresponding anomalies may not be visible at the initial time but only later, as the seasonal cycle advances. Here we describe the method and an assessment of the forecasts produced. This paper is structured as follows: Data used for creating the forecasts are presented in section 1.2, followed by a description of the applied verification metrics in section 1.3. The Spatial Damped Anomaly Persistence (SDAP) method is described in detail in section 1.4. In section 1.5, we go through the results from verifying the forecasts produced with our method compared against other traditional references, as well as against the performance of model-based S2S forecast systems. The paper concludes with a short discussion in section 1.6.

1.2 Data

1.2.1 Sea-ice concentration observations

We have used the Global Sea Ice Concentration Climate Data Record from the Ocean and Sea Ice Satellite Application Facility (OSI SAF) to determine the climatological and initial sea-ice edge. The data is labelled as OSI-450 for years 1979 to 2015 and is based on passive microwave measurements. From 2016 onwards, OSI-450 was extended as OSI-430b and is available with a 16-day latency. Alongside the microwave measurements, these products also use operational analyses and forecasts from the European Centre for Medium- Range Weather Forecasts (ECMWF) for atmospheric corrections. A near real time version of this record without the additional corrections is also available, as OSI-430. All of the data is freely available on the OSI SAF website and further details regarding the processing are described by (Lavergne et al., 2019). OSI-450 (and OSI-430b) records are given on a Lambert Azimuthal Equal Area polar projection, also known as the EASE2 grid. The two hemispheres are separated and the horizontal grid spacing is 25km. Following this setup, our forecasts are also produced on the EASE2 grid at 25km resolution for each hemisphere.

1.2.2 Subseasonal-to-seasonal (S2S) forecast data

We compare the performance of our forecasts against the performance of the models from the Subseasonal to Seasonal (S2S) Prediction Project. The S2S database contains forecasts and reforecasts from several major operational centers. We have measured the uncalibrated concentration forecasts of 5 dynamical sea ice forecasting systems - the European Centre for Medium- Range Weather Forecasts (ECMWF), the Korea Meteorological Administration (KMA), Météo-France (MF), the National Centers for Environmental Prediction (NCEP), and the UK Met Office (UKMO). Alongside, we also used concentration forecasts from an older version of ECMWF (here named 'ECMWFpres') where the sea ice state is prescribed using initial persistence for the first 15 days and then relaxed toward climatology. Further description regarding the S2S project is given by (Vitart et al., 2017). For a consistent comparison, all forecasts have been interpolated to a common grid with 1.5 degree horizontal resolution and only the period between 1999 and 2010 are considered in the analysis, in line with the analysis provided in (Zampieri et al., 2018, 2019).We have, however, excluded the forecast from the Chinese Meteorological Administration (CMA) and increased the coverage of the sea mask used in this analysis.

1.3 Verification Metrics

1.3.1 Spatial Probability Score

We are primarily using the Spatial Probability Score (SPS; Goessling and Jung, 2018) for verification of the sea-ice forecasts, which is defined as the spatial integral of the squared probability difference (i.e., the half-Brier Score) as:

$$\text{SPS} = \int_A (P_f(x) - P_o(x))^2 dA$$

where where P_f and P_o are the forecasted and observed SIP. Since the S2S models provide ensemble forecasts of sea-ice concentration, we can derive a continuous (non-binary) probability of ice-presence. The resulting score, which is in area units, quantifies the area of mismatch between forecasts and observations. Measuring SPS of a deterministic, or binary, forecast results in the Integrated Ice Edge Error (IIEE; Goessling et al., 2016). We also use the SPS in our method for empirically optimising the weights by which our initial binary (anomaly persistence) forecast of the ice-edge location is damped towards climatology, resulting in a probabilistic (damped anomaly persistence) forecast, as detailed in section 1.4.2.

1.3.2 Modified Hausdorff Distance

Given that we have used the SPS not only for evaluation but also for the empirical estimation of optimal damping weights (detailed below), we have also used a second verification metric to validate the skill of our method. The Modified Hausdorff Distance (MHD; Dukhovskoy et al., 2015; Palerme et al., 2019) measures the distance between two contours and the resulting score is in distance units. For two contours A and B, with points a and b in each, MHD is defined as:

$$\mathrm{MHD}(A, B) = \max\{\mathrm{mean}_{a \in A} d(a, B); \mathrm{mean}_{b \in B} d(b, A)\}$$

Since MHD only considers the ice-edge (and not the probability of ice-presence), the ice edge has been derived from each probabilistic forecast using the SIP = 50% contour. As SPS and MHD measure forecast skill quite differently, using both metrics can reveal whether or not the forecasts from our method are skillful independent of the verification metric used.

1.4 Spatial Damped Anomaly Persistence (SDAP) Method

The steps towards creating the SDAP forecast of the ice-edge location can be divided into two main phases. In the initialisation phase, the initial anomaly in ice-edge location is first derived from the climatological Sea Ice Probability (SIP; the probability of sea-ice concentration exceeding 15%) at points constituting the initially observed ice edge and then inherited (projected using nearest-neighbour interpolation) from the initial ice edge to each point of the (quasi-)global grid. In the forecasting phase, the inherited anomaly and the climatology valid for the forecast target date are compared to determine, first, a deterministic ice-edge forecast by specifying "ice" versus "no ice" at each grid point, corresponding to a binary anomaly persistence forecast, and second, a probabilistic forecast by relaxing the deterministic forecast toward climatology, resulting in a damped anomaly persistence forecast. A schematic overview of the whole procedure is provided in Fig.1.2.

Our method, detailed in the following, propagates the initial anomalies in local sea-ice extent in space by spatial inheritance, and then uses the anomalies to predict SIP in time based on the seasonal evolution of the climatological sea-ice probability. The rationale of this approach is based on the assumption that anomalies, especially in the ice and ocean state associated with an ice-edge anomaly, have a certain spatial and temporal correlation length. For example, if the ice edge extends further than

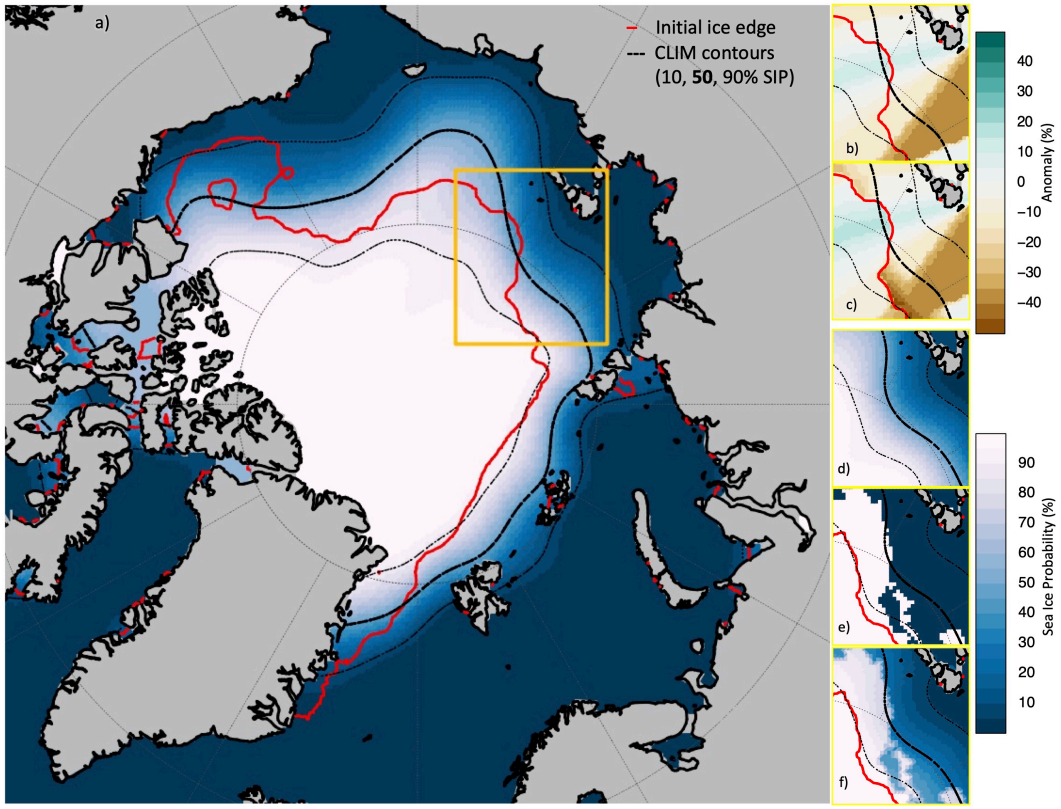

FIGURE 1.1: Map of the Arctic domain showing the climatological probability field for 1st of September 2020, and steps of the forecasting method (in insets) for the region selected. Insets 'b' and 'c' show the inherited and adjusted anomaly fields, 'd' shows the climatological probability field for the target date (30th of September, 2020), 'e' shows the Spatial Anomaly Persistence forecast and 'f' shows the Spatial Damped Anomaly Persistence forecast for the target date. In each panel, the red line denotes the 15% contour of the observed ice concentration for the relevant date while the dashed black lines show the 10%, 50% and 90% contour of the climatological probability field. Maps 'b' and 'c' use the Anomaly colour scale while all other maps use the Sea Ice Probability colour scale. Further details are in the text.

usual into the ocean, this might have been caused thermodynamically by colder-than-usual regional temperatures (e.g., due to anomalous atmospheric circulation) that would also have caused thicker-than-usual and/or denser floes in the adjacent pack ice and colder-than-usual sea-surface temperatures (SST) in the adjacent open ocean. If this situation occurred during the melt season, the thicker-than-average ice would hamper the ice-edge retreat; if it occurred during the freeze season, the colder-than-average SST would facilitate the ice-edge advance.

For both initialisation and forecasting phases of the method, we make use of the climatological sea-ice probability (SIP) at each gridpoint, derived as the fraction of years in the preceding 10 years with ice concentration > 15 % at the same date. For initialisation, we use the climatological SIP for the initial date (hereby referred to as CLIM_init) and for forecasting, the climatological SIP for the target date (hereby referred to as CLIM_target). The climatological SIP for each date is also one of the traditional reference forecasts used in the results section, where it is simply referred to as CLIM. For each date, a Gaussian filter with a radius of 220 km is applied to smoothen the climatological probability field. This results in smooth SIP contour lines which are advantageous for the subsequent steps.

1.4.1 Initialisation phase

The initial ice edge, defined as the 15% contour of ice concentration on the initialisation date, is overlaid on CLIM_init. For every point along the initial edge contour, the anomaly is first measured by subtracting the climatological SIP at the nearest grid cell from median probability (50%). If the ice edge is in a region of high (or low) climatological probability, the anomaly is negative (or positive). In other words, if the local ice extent (region with ice present) is smaller (or larger) than the climatological median, the ice edge should be in a region of high (or low) climatological probability and the anomaly is negative (or positive). Next, these anomalies are copied from the initial edge to the climatological median contour (corresponding to 50% ice-presence probability) using nearest-neighbor match to find the closest points. This intermediate step helps avoid biased patterns of spatial inheritance that otherwise occur due to geometrical effects (not shown). The anomaly is then passed (spatially "inherited") to the full grid, using a nearest-neighbour search to find the closest point on the climatological median contour from each grid cell, resulting in a map as shown in Fig.1.1b.

In some cases, a grid cell that has (does not have) ice in the initial observation might inherit a strong negative (positive) anomaly from the median, which causes

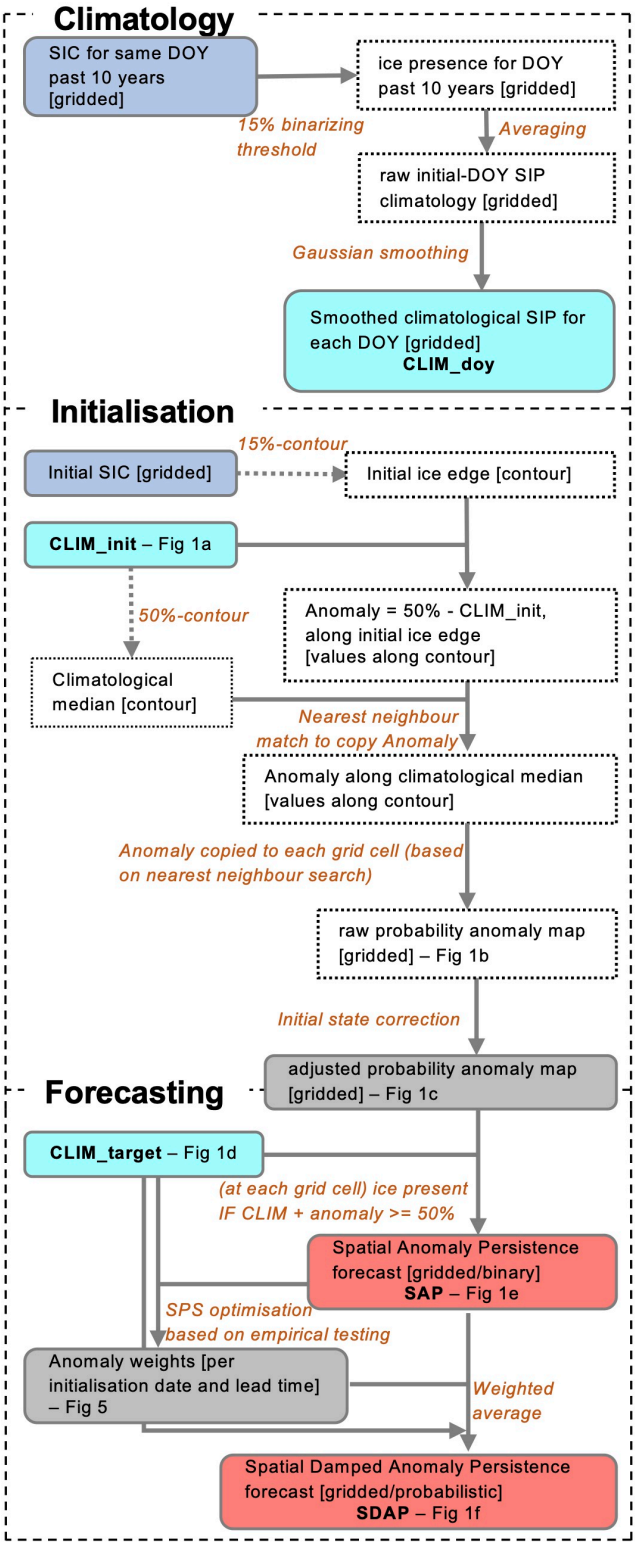

FIGURE 1.2: Schematic showing the steps taken to generate the Spatial Damped Anomaly Forecasts. Alongside the two phases described in the text (Initialisation and Forecasting), the steps for creating the climatological SIP (CLIM) for each day-of-year(doy) are also shown separately. The blue panels represent the input data, cyan panels represent CLIM, and red panels represent the forecast outputs. The grey panels show intermediate products that are used for several forecasts. The maps and contours mentioned in the steps of the schematic can be seen in various panels of Fig. 1. Further details are given in the text.

the initial forecast at day 0 to not have (have) ice in the grid-cell. This initial mismatch generally occurs in cases where the observational gradient of ice presence is locally reversed compared to the climatological gradient of ice presence probability. The problem is simply solved by checking the initial day forecast and replacing the anomaly inherited from the climatological median with the anomaly at the grid cell (i.e., 50% minus the climatological probability of ice presence) for grid cells with an initial mismatch. This step, labelled as initial state correction in the schematic, results in an adjusted probability anomaly map (Fig.1.1c) that can be persisted in time and used for forecasting.

1.4.2 Forecasting phase

Using the adjusted anomaly map from the initialisation phase (Fig.1.1c), a forecast for any lead time can be created by simply adding the anomaly to the climatological SIP (CLIM_target) at each grid cell for the target date, and then using a 50% threshold. For example, if we initialise our forecast on September 1 and want to forecast the sea-ice edge 29 days later, then we add the anomaly from September 1 to the climatological SIP for September 30 (Fig.1.1d). For every gridpoint, if the resulting probability is 50% or more, then we assign that point to be ice-present. This gives a map of 1 (ice) and 0 (no-ice) which is our deterministic (binary) Spatial Anomaly Persistence forecast (SAP) of ice-presence (Fig.1.1e).

The binary SAP forecast, corresponding to a single sharp ice edge, does not take into account that the spatial and temporal correlation scales of anomalies are limited. For example, an initial ice-edge anomaly in March obviously carries much less information about ice-edge anomalies in the following September, when (in most Arctic regions) the ice edge will be far away from the initial location, and sufficient time will have passed to turn initial anomalies into largely uncorrelated anomalies. Therefore, we create a probabilistic ice forecast by damping our deterministic forecast towards climatology. For each lead-time, a damped probabilistic forecast is obtained as the weighted average of the deterministic prediction and the climatological probability. The optimal anomaly weights are determined by using a training set of forecasts, where we compute the SPS for anomaly weights between 0 and 1 in steps of 0.05, for each combination of lead time and initial time of the year, and finding the weights that minimise the error. The optimised anomaly weights for each hemisphere (described in section 1.5.2) can then be used for other, independent, years to get the probabilistic Spatial Damped Anomaly Persistence (SDAP) forecast (Fig.1.1f).

In our study, we initialise the forecasts at the start of each month between 1989 and 2020 and make predictions of the ice edge for the following year. The first 10 years of the period are used for empirical training to determine the anomaly weights (see Fig.1.6) and are thus excluded from the analysis. The remaining 22 years (1999 to 2020) have been used for the evaluation shown in the results below. For comparison against the forecasts from dynamical models in the S2S dataset (section 1.5.4), the forecasts from our method were interpolated to the coarse 1.5° S2S grid, and only years 1999 to 2010 were used, as in (Zampieri et al., 2018, 2019).

1.5 Results

Here, we present the evaluation of the forecasts from our method compared against other traditional benchmarks for ice-edge forecast, as well as against forecasts from dynamical models in the S2S data set. Alongside climatological SIP (here referred to as CLIM) and initial-state persistence (PERS), the 50% contour of the climatological probability has also been used to generate a binary forecast - the climatological median (CLIM-MED). Similarly, the 50% contour of the spatial damped anomaly persistence (SDAP) has been used to generate a median damped anomaly persistence forecast (SDAP-MED), which is different from the binary Spatial Anomaly Persistence forecast (SAP) described in section 1.4.2.

1.5.1 Comparison against traditional benchmarks

The SPS result of all forecasts, averaged between 1999 and 2020 and across all seasons (Fig.1.3), reveals that the performance of the SAP forecast is better than CLIM up to day 15 and better than CLIM-MED up to day 39 in both hemispheres. The performance of SDAP is better than SAP, outperforming CLIM at 2 months of lead time by an average of 0.03 million km^2 in the Arctic (0.04 million km^2 in the Antarctic). While SAP forecasts show an improvement over simple persistence from day 6, SDAP is better from day 3. The method design enables this forecast to perform at least as well as CLIM even at long lead times, leading it to be the most skillful forecast in this comparison set at all lead times, except the first 2-3 days where PERS is marginally better.

As mentioned in section 1.3.1, SPS of binary forecasts is the same as their Integrated Ice Edge Error. Since SPS rewards the probabilistic information in a forecast (Goessling and Jung, 2018), both CLIM and SDAP have a clear advantage over their binary counterparts. CLIM-MED has a near constant skill loss of 0.3 million km^2 in

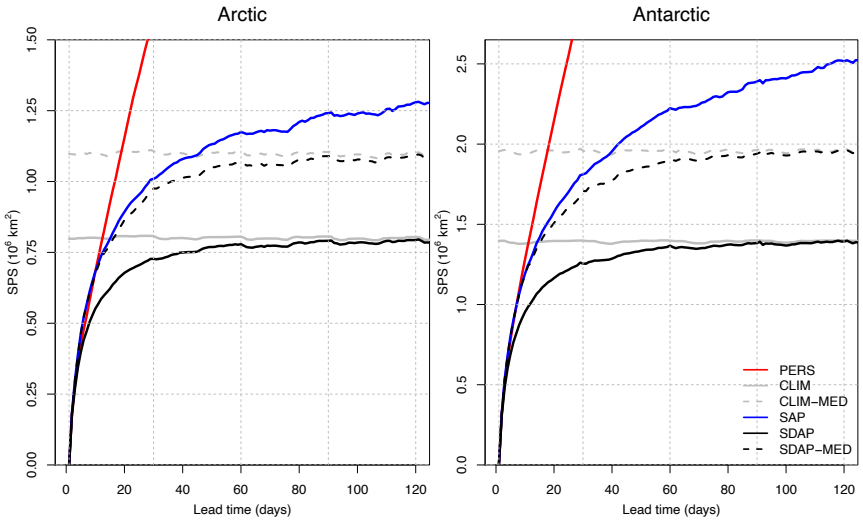

Figure 1.3: Spatial Probability Score (SPS) at increasing lead times (in days) for all forecasts averaged across all forecasts initialised between 1999 and 2020. The two dashed lines (CLIM-MED and SDAP-MED) are binary forecasts derived by assigning the ice edge at the 50% contour of the probabilistic forecasts (CLIM and SDAP). Further details are in the text.

the Arctic (0.5 million km^2 in the Antarctic) compared to CLIM. The error of SDAP-MED is low initially (similar to the other anomaly forecasts) and converges towards the error of CLIM-MED, similar to how the error of SDAP converges towards the error of CLIM.

Given that our methodology uses the SPS to optimise the anomaly weights, it is possible that using the SPS also as a verification metric leads to an overestimation of the skill. We thus also use the Modified Hausdorff Distance (MHD) as an independent verification metric, although the MHD can be applied only to the binary (median-based) forecast variants. While this verification method measures the forecast skill quite differently compared to SPS (Palerme et al., 2019), repeating the evaluation of the binary forecast variants with the MHD instead of the SPS provides overall a very similar picture: The SDAP-MED forecast outperforms the CLIM-MED forecast throughout the lead-time range, the undamped (SAP) forecast outperforms the CLIM-MED forecast up to 60 days lead time in the Arctic (43 days in the Antarctic), and simple persistence outperforms the other forecasts only slightly during the first few days (Fig.1.4). The confirmation of the overall results with an independent metric provides additional confidence in our reference forecast method.

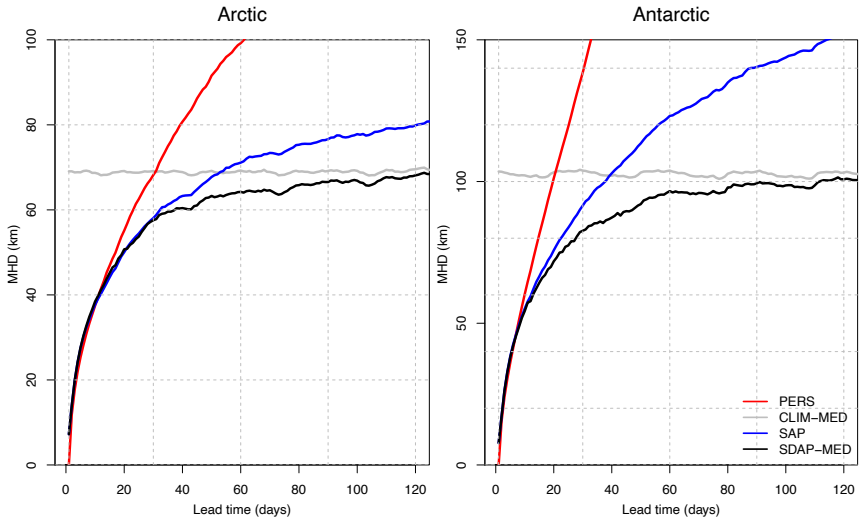

FIGURE 1.4: Same as figure 1.3, but for Modified Hausdorff Distance (MHD) of the binary forecasts compared to the observed ice edge.

The previous analysis was based on outputs averaged over the entire time period. Given the strong seasonal cycle of Arctic sea ice and the processes that dominate the sea-ice budget, we now assess the performance of our method as a function of season. The SPS for SDAP forecasts generally increase with lead time for all initialisation months, approaching and mimicking the SPS of the climatological forecast (Fig.1.5). The errors are on average higher for the summer period in both hemispheres, and also during the late winter in the Arctic, when the sea ice is at its largest extent. These seasonal cycles can be linked to corresponding variations of the ice-edge length, although the December/January SPS maximum in the Antarctic is more likely related to enhanced interannual lateral ice-edge variability (Goessling et al., 2016).

1.5.2 Anomaly weights

As mentioned above, the deterministic anomaly forecasts are damped by empirically determined weights and added to an inversely weighted climatology to derive the probabilistic anomaly forecast. The weights, derived independently for each hemisphere (see Fig.1.6) using SPS results from the optimisation period of 1989 to 1998, reveal the timescale at which information from the initial anomalies is lost and climatology becomes more informative. By day 30, the initial-anomaly weights decrease to 50% for most initialisations, but most forecasts have not been completely

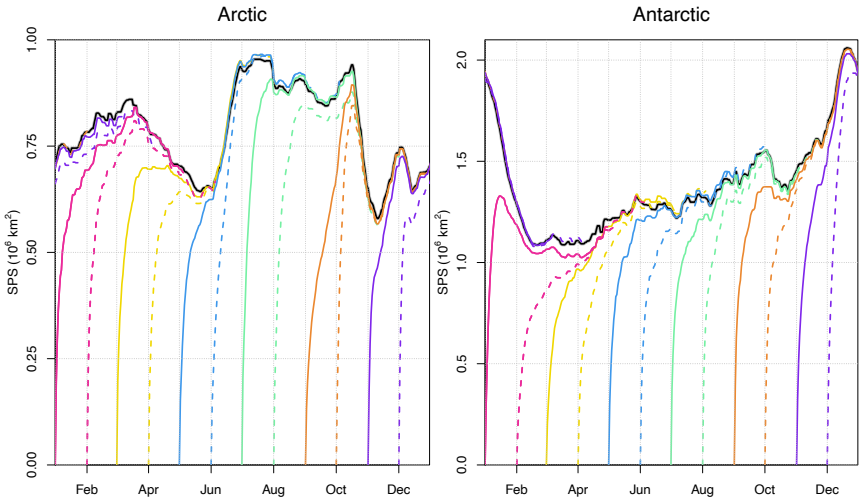

Figure 1.5: SPS for the first 120 days of SDAP forecasts initialised at the start of each month (as represented by the coloured lines), alongside CLIM for each date (black continuous line), with results averaged between 1999 and 2020.

damped to climatology even at 3 months of lead time. In both hemispheres, the anomaly weight stays high for a longer period at the end of the summer melt season - August/September in the Arctic and February in the Antarctic.

The weight of the anomaly is not always decreasing monotonically (Fig.1.6), although the influence of initial conditions should generally decrease with time. While some intermittent fluctuations can be caused by sampling uncertainty due to the limited optimisation period, we argue that some of this can be linked to what is known as reemergence of sea-ice anomalies (Blanchard-Wrigglesworth et al., 2011): When over the course of the seasonal cycle the (climatological) ice edge first migrates away from the initial location and then returns to a similar location later, the initial ice-edge anomalies may carry more information for that later state than for the earlier - but more remote - intermediate state. This seems to be the case in particular for the Arctic forecasts initialised at the beginning of August, which exhibit a local minimum of the anomaly weight around day 35-40 (around the sea-ice minimum) and higher weights again thereafter until around day 70 (Fig.1.6). The increased anomaly weight acts to keep the associated forecast error below the error of climatology for longer compared to other initialisation months.

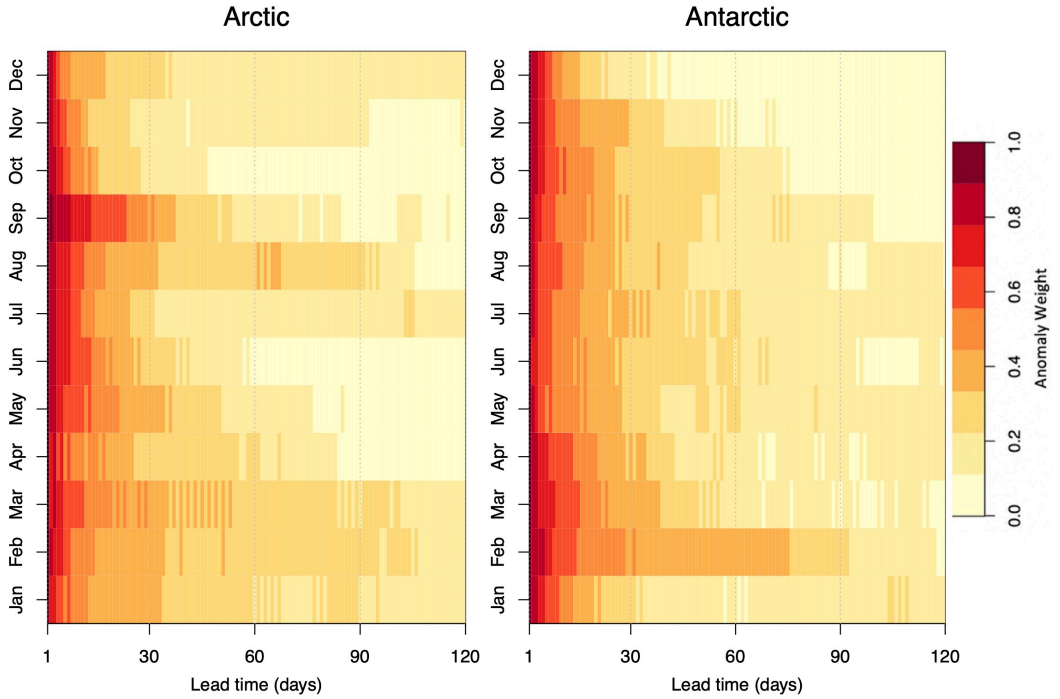

FIGURE 1.6: Anomaly weights for the SDAP forecast as a function of initialisation month and lead time (first 120 days). These weights determine the ratio of SAP forecast to climatology in the SDAP forecast and are measured using optimisation of SPS as described in the text.

1.5.3 September 2020 as an illustrative example

To illustrate our forecast method, we now consider the example of the Arctic ice edge with the initial condition on the 1st of September, 2020, as shown in Fig.1.1a. Sea-ice extent in the Arctic during September 2020 was the second lowest in satellite records and during October 2020 was the lowest October ice extent measured (NSIDC, 2020). Persistent offshore winds from the Siberian coast, associated with a positive phase of the Arctic Oscillation (AO) during the preceding winter, meant that the Eurasian parts of the Arctic were mostly ice-free and had strong negative anomalies (as seen in Fig.1.1b).

SDAP forecasts in this particular case are better in the American part of the Arctic than in the Eurasian part (Fig.1.7). The forecasts correctly suggest that positive anomalies would persist in the southeastern Beaufort Sea (on the left in each panel) until the end of the month, and even the persistence of the ice-free patch within the ice cover of the Beaufort Sea is correctly forecast, although the position is shifted. In the Greenland Sea and Fram Strait, the forecasts are fairly

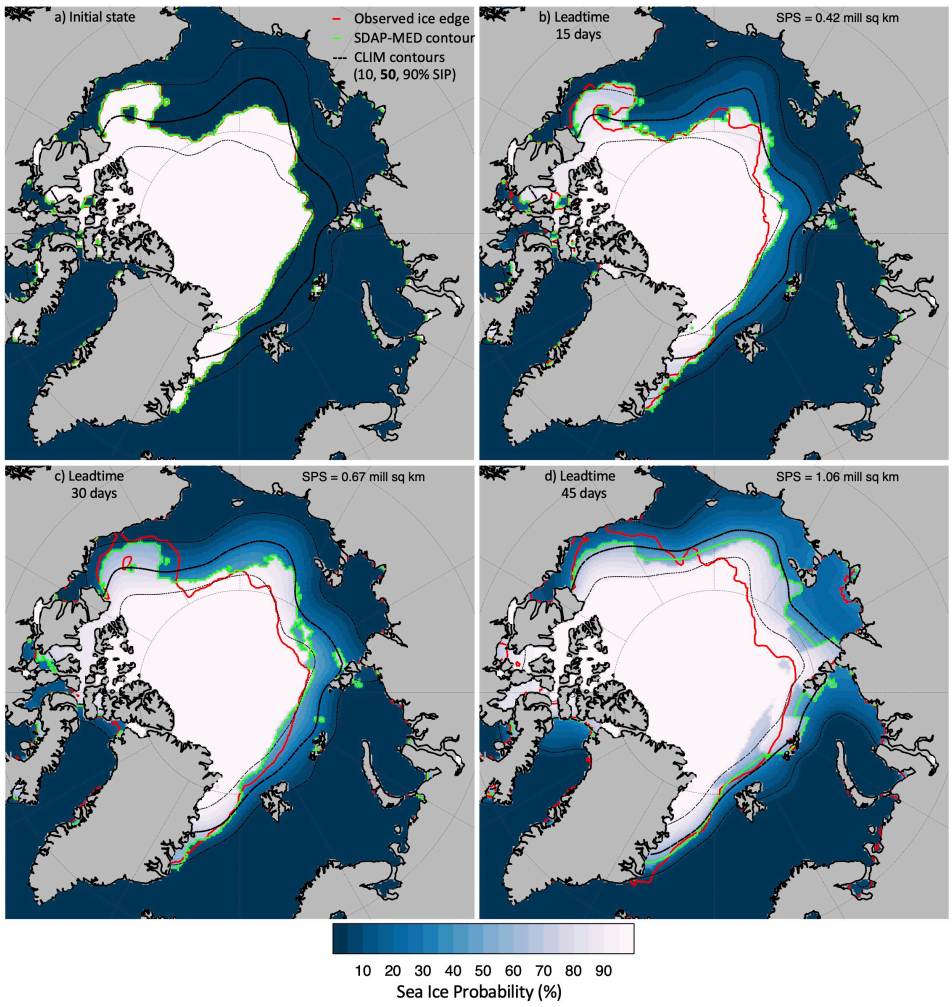

Figure 1.7: Observed sea ice conditions in the Arctic for 1st of September, 2020 and the corresponding SDAP forecasts at lead times of 15, 30 and 45 days. In each of the panels, the black lines show the contours of climatological probability (10, 50 and 90%) and the green line shows the median contour of SDAP forecast, while the red line shows the actual observed ice edge on the respective date.

accurate at day 15 and 30, and still better than the climatological median at day 45. North of the Laptev Sea, the forecast retains the initial negative anomalies - as dictated by the concept of anomaly persistence - and thus shows a similarity in shape to the initial ice edge. While a small patch remains accurately ice-free at day 30, the overall forecast in this region shows an expansion of the high-probability areas toward the coast following the seasonal evolution of the climatological probabilities. By contrast, the actual ice edge does not follow the usual seasonality but remains largely unchanged, thereby developing even stronger negative anomalies. The SDAP forecast remains better than climatology, but the anomaly intensification is not captured. This highlights the limitation of the SDAP approach, given that it is based on the persistence of anomalies.

North of Franz Joseph Land, the initial anomaly is strongly negative, yet there is ice present along the island coasts (in a positive anomaly region), allowing some parts of the median edge to inherit positive anomalies. Depending on the exact position of a grid cell in relation to the median and ice-edge contours, it might inherit different anomalies and predict different ice conditions. This can result in some discontinuity artefacts in the SDAP forecasts, as seen in this region at day 45, where some grid cells of the ocean in this region have a lower probability of ice-presence than the surrounding region, or show discontinuity in the SDAP-MED contour. We discuss this issue further in section 1.6.

1.5.4 Comparison against S2S dataset

The main motivation for this study is to use the damped anomaly forecast as a reference benchmark for evaluating dynamical sea-ice models. Therefore, here we compare the performance of the forecasts from this method against those from the S2S Prediction database (Vitart et al., 2017), in line with the analyses presented in (Zampieri et al., 2018, 2019). The results for the S2S models shown here are similar to those found by (Zampieri et al., 2018, 2019), although the scores have increased throughout due to the use of a larger sea-mask (larger area for potential error). The SDAP forecasts for years 1999 to 2010 were remapped to the common 1.5° S2S grid for this analysis and this has led to an increase in forecast error relative to climatology - in contrast to the evaluation on the OSI-SAF grid (Fig.1.3), the SDAP error now reaches and slightly surpasses the climatological error around 60-day lead time in the Arctic (Fig.1.8 left; see also Fig.A.2).

The undamped (thus binary) SAP forecast has a similar forecast skill as UKMO and KMA in both hemispheres (Fig.1.8), despite the fact that these forecast systems

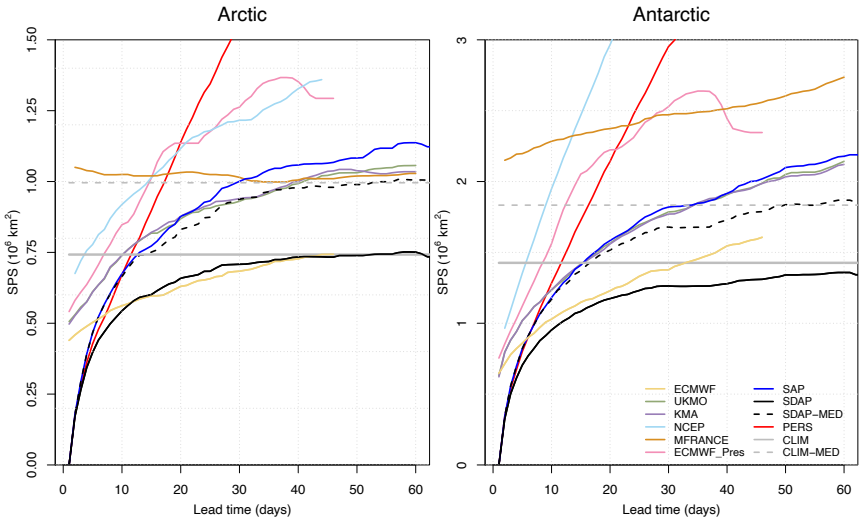

Figure 1.8: SPS for forecasts from the S2S dataset, alongside climatology (CLIM), initial state persistence (PERS), Spatial Anomaly Persistence (SAP) and Spatial Anomaly Damped Persistence (SDAP) forecasts, averaged across all seasons and years between 1999 and 2010.

have the advantage of providing ensemble-based probabilities rather than a binary ice edge. The damped (thus probabilistic) SDAP forecast in the Arctic clearly outperforms UKMO and KMA and is about as skillful as ECMWF. The latter is the best-performing model in the S2S set (Zampieri et al., 2018, 2019) and (without calibration) the only one that is more skillful than climatology beyond day 15 in the annual average. In the Antarctic, our SDAP method even outperforms the (uncalibrated) ECMWF ensemble, in particular beyond 20-day lead time when the ECMWF system appears to develop biases so that climatology provides a better forecast beyond 32-day lead time.

We also compare the performance of the SDAP forecasts against those from the S2S dataset using Modified Hausdorff Distance (MHD; Fig.A.1). Due to the differences in the method, including that MHD can be applied only to binary forecasts (based on the median ice edge where applicable), MHD measurements do not precisely mirror the results of the SPS measurements (for all models). Yet the ranking of the models is similar and the anomaly forecasts have a high skill compared to most other models or climatology. The average MHD of the ECMWF forecasts is again the lowest in the S2S set, except at short lead times below ~12 days. The SDAP forecasts outperform the ECMWF forecasts in the short range and roughly match the skill of ECMWF at longer lead times, particularly in the Antarctic.

1.6 Summary and Discussion

This paper describes a novel method for forecasting the presence and extent of sea ice in the Arctic or Antarctic based on persistence and damped persistence of probabilistic anomalies. The method requires only historical and initial ice-presence information to make the predictions, yet remains more skillful than climatological forecasts at month-long lead times. This is in contrast to most of the models from the S2S database, of which (without calibration) only one model performs better than climatology beyond 12 days of lead time.

The SDAP method uses the probabilistic anomaly from the initial date, which is spatially distributed ("inherited") by a nearest-neighbour search, and adds it to the climatology of the target date to generate a deterministic anomaly persistence forecast. With increasing lead time the deterministic anomaly is damped to generate a probabilistic forecast. At longer lead times, one can expect that dynamical and thermodynamical processes cause the initial anomalies to be less informative. Therefore, the method has been designed to increase the damping with time and converge to climatology at long lead times. The anomaly weights, determined empirically using the reforecasts between 1989 and 1998, show that initial anomalies remain highly informative (weight $> 30\%$) for about 20-30 days in most months, and even longer in late summer. It is likely that in the Arctic, as the climatological ice-edge shifts with the continued decrease in ice extent, the optimal damping should also change.

We note that there are instances of sharp spatial transitions in the anomalies inherited to the grid (as described in section 1.5.3) due to the spatial distribution of the initial ice edge. Spatially smoothing the anomaly before passing it to the grid or limiting the nearest-neighbor search to the main ice-pack could smoothen the transition. Explicitly adding a spatial component to the damping might also be better than using a pan-Arctic anomaly weight that evolves only with lead time. This might also have the potential to implicitly capture the re-emergence of anomalies when the ice edge returns to a location over the course of the seasonal cycle after the extent has reached its maximum or minimum. Nevertheless, the empirical approach used here, while simplistic, gives a good estimation of the overall decrease in the information content of the anomalies.

While our results show that the damped anomaly forecast outperforms most of the models in the S2S dataset, it must be noted that the skill of the dynamical sea-ice models would be higher than shown after bias correction or other forms of forecast

calibration, which is standard for forecasts of other predictands at subseasonal-to-seasonal timescales. Forecast calibration remains challenging for sea ice, although promising approaches have recently been suggested (e.g., Dirkson et al., 2019; Director et al., 2017). Some of the models only have a small number of ensemble members (for example KMA and UKMO both have 3 members each). This means a higher discretisation of SIP, which can also lead to an increase in the SPS measurement. Moreover, the resolution of the common S2S grid is low, and forecast skill was found to deteriorate after interpolation into this grid (Fig.A.2). It is possible that measuring the performance of the S2S models on their native grid would have resulted in a higher skill. The models output several variables, whereas our method is designed to only forecast ice-presence. Using another variable for forecast verification, or simply using a different concentration threshold, could also result in a higher prediction skill for the dynamical models, as shown by (Zampieri et al., 2019).

The SDAP method, applied here for predicting ice-presence equivalent to 15% or more sea-ice concentration, could also be used for predicting other binary fields. Considering sea-ice concentration, (Mizuta et al., 2008) proposed a different probabilistic method to estimate ice concentration by combining individual predictions for different concentration thresholds; While the method is quite different, a similar framework for our method can be used for estimating ice concentration, or thickness, by using several binary levels.

To conclude, the method proposed here is on average as skillful as the ECMWF forecast system, which is the most skillful one in the S2S database. Comparing only against persistence and climatology can give the impression that sea-ice forecasts from some of the S2S forecast systems can already be regarded as "skillful" and thus of potential value for users. However, using a more challenging benchmark such as the spatial damped anomaly persistence (SDAP) method introduced here reveals that dynamical forecast systems still have some way to go until they can generate substantial value beyond much simpler methods. We hope that, by including more challenging benchmark forecast methods such as ours in their evaluation workflow, other researchers and forecasting centers can build a better basis to improve their sea-ice forecast systems.

The thing that doesn't fit is the thing that's the most interesting: the part that doesn't go according to what you expected.

Richard P. Feynman

2

Ice edge error in initial state of ECMWF forecasts and ORAS5 analysis

Abstract

Previous research on sub-seasonal to seasonal (S2S) forecasts of sea ice have found that the ECMWF's ice edge forecasts exhibit a relatively high level of prediction skill when compared to other operational S2S systems. However, significant errors were observed in the initial day of the forecasts. Here, we trace the initial forecast errors back by scrutinising the underlying ocean analysis ORAS5, exploring the effects of regridding, and quantifying observational uncertainties. We show that interpolating the ice concentration data into a coarse grid (as used for the S2S project) notably increases ice edge errors, meaning that the initial forecast errors are lower in the native model grid than what has previously been reported. We also note that the analysis errors are comparable to observational uncertainty, although improvements in the data assimilation may still increase the accuracy of the reanalysis and forecasts.

Chapter 2 is being prepared for submission to 'Quarterly Journal of the Royal Meteorological Society'.

2.1 Introduction

With decreasing sea-ice cover in the Arctic (Comiso, 2012) and increasing economic activities, traffic in the polar seas is expected to increase substantially in the coming decades (Palmer, 2013). Anticipating the growth in user demand, several operational centers are now running fully coupled NWP systems in the polar regions, initialized using observations, with the aim of improving prediction capabilities at timescales ranging from weeks to months (Smith et al., 2015; Lawrence et al., 2019). Several such predictions have also been included in a database of sub-seasonal to seasonal forecasts (S2S; Vitart et al., 2017), which allows for the inter-comparison of sea ice forecasts, among other outputs, in a homogeneous framework.

In prior studies, Zampieri et al. (2018, 2019) compared the sea ice edge forecasts from the models in the S2S database (as well as climatology and initial state persistence) against global sea ice concentration observations from the Ocean and Sea Ice Satellite Application Facility (OSI SAF). By measuring the Spatial Probability Score (SPS; see Goessling and Jung, 2018) of the sea ice output, they showed that only one of the forecast systems in the database was able to produce skillful forecasts at lead times beyond a few weeks - the system of the European Centre for Medium Range Weather Forecasts (ECMWF). However, these forecasts started with substantial initial error, on average 50% of the climatological error (also shown in Niraula and Goessling, 2021). Understanding the origin of these large initial errors is an essential motivation for this study.

In order to examine the initial error in the ice edge forecasts from ECMWF, we first take a step back and briefly summarise how the prediction system is set up. The ECMWF numerical weather prediction system, also known as the Integrated Forecasting System (IFS), is a fully coupled ensemble setup consisting of several component models, as described in section 2.2 of this manuscript. We are focussing on the output of the ocean and sea-ice components, which compute the evolution of the sea-ice cover. In the current model setup, the ocean and sea ice conditions are initialised with data from ORAS5, the near real time component of the OCEAN5 reanalysis ensemble (Zuo et al., 2019). OCEAN5 assimilates sea ice concentration (and SST) from the Operational Sea Surface Temperature and Sea-Ice Analysis (OSTIA; Donlon et al., 2012)(OSTIA; Donlon et al., 2012) dataset, which itself derives the sea ice fields from OSISAF observations. Since the datasets are well linked, we can reasonably assume a strong agreement in the ice-edge between the observational data (OSI SAF), ocean reanalysis (ORAS5), and initial forecasts. Thus, significant

errors in the initial forecast are unexpected and we aim to understand the issue by comparing them against the errors in the intermediate reanalysis.

For the S2S project the re-forecasts were saved on a common grid, with a coarse horizontal resolution of 1.5 degrees (Vitart et al., 2017), while the native resolution of the ocean model (also used by ORAS5) is the ORCA 0.25 degrees grid (Owens et al., 2018). Prior research has suggested that interpolation of data into a coarser grid can introduce errors in the sea ice edge (Niraula and Goessling, 2021). Here, we examine the effect of interpolation on ice edge error by interpolating the ORAS5 (and OSISAF) ice concentration fields to the 1.5 degrees grid and back. Any additional error resulting from this process can be ascribed to interpolation, and may indicate whether the re-forecasts would have yielded better results in the native resolution.

The forecast system and the ocean reanalysis both use NEMO (for the ocean) and LIM (for ice) model components. In Zuo et al. (2019), the authors found that the average RMSE of Arctic SIC in ORAS5 compared against OSISAF observations over the 1993 to 2008 period is less than 5%. However, other studies have found evidence that the NEMO and LIM components can overestimate the sea ice extent in the winter and underestimate it in the summer (Rousset et al., 2015; Tietsche et al., 2015; Barthélemy et al., 2017). Comparing the ice edge from OCEAN5 and OSISAF against images from the Moderate Resolution Imaging Spectroradiometer (MODIS) during the winter 2017/18 period has also showed that the ice edge from OCEAN5 is more oceanwards (i.e. overestimating ice-presence) compared to OSISAF (Day et al., 2022). However, it was also seen that the ice-edge from OSISAF can notably differ from those observed in MODIS images. Given that pan-Arctic sea-ice extent can vary by up to 0.6 million square kilometers across different observational datasets (Ivanova et al., 2014), it is unclear whether the ice-edge errors in the ocean reanalysis products are significantly higher than observational uncertainties. To address this question, we are also assessing the discrepancy in ice-edge locations between OSISAF and another observational sea-ice concentration dataset, AMSR-E/2 (Kaleschke et al., 2001; Spreen et al., 2008). This dataset employs a different passive microwave data source and sea ice concentration retrieval algorithm than OSISAF, making it a suitable representative of the broader range of observational differences.

Based on these ideas, here we are measuring the daily ice edge error and biases in the analysis and initial forecast data against the satellite-based observations. We investigate the role of interpolation and put the errors in the context of observational uncertainty. The datasets are described in section 2.2 of the manuscript, while the main methods applied here and the verification metric used for comparison

are explained in section 2.3. Section 2.4 contains the comparison results, and the manuscript ends with a summary and discussion in section 2.5.

2.2 Data

OSI SAF observations

We primarily use the passive microwave sea ice concentration (SIC) measurements from the Ocean and Sea Ice Satellite Application Facility (OSI SAF) as our observational data. This is a global ice concentration climate data record, labeled 'OSI-450' for years 1979 to 2015, and extended from 2016 onwards as 'OSI-430' (with a 16-day latency). They are given on a Lambert Azimuthal Equal Area polar projection, also known as the EASE2 grid, separately for each hemisphere, and with a horizontal grid spacing of 25 km. Further details regarding the data can be found in Lavergne et al. (2019) and the data can be freely downloaded from the OSI SAF website. The OSI SAF concentration data is also used to create the climatological forecast (here referred to as CLIM) for each date by taking the observed state for the same date from the previous 10 years as a 10 member ensemble, resulting in a probabilistic forecast based on climatology.

S2S re-forecasts from ECMWF

The data referred to as ECMWF in the Sub-seasonal to Seasonal Prediction Project (S2S; Vitart et al., 2017) are the extended range re-forecasts from the European Centre for Medium Range Weather Forecasts (ECMWF). The ECMWF prediction system, also known as the Integrated Forecasting System (IFS), is a fully coupled global ensemble system that uses the NEMO 3.4 Ocean model and the Louvain-la-Neuve Sea Ice Model (LIM2), alongside other components. The ocean initial conditions for the re-forecasts come from the ORAS5 reanalysis (Owens et al., 2018).

The re-forecasts produced for the S2S project consist of 11 ensemble members, initialized twice weekly, and can be downloaded from the ECMWF data repository. We are using the first output time step available, given as '0 to 24 hour daily averaged forecast', which causes, on average, a 12h shift between this data and the exact initial time at which the corresponding ORAS5 and OSI SAF data are valid. Given the slow evolution of the sea-ice state, this slight inconsistency is considered negligible. The data was stored on a regular grid with a coarse horizontal resolution of 1.5 degrees. In the rest of this manuscript, this dataset is labeled as S2S_EC.

ORAS5 re-analysis ensemble

ORAS5 is the near real-time component of the OCEAN5 re-analysis system. It uses the NEMO 3.4 ocean model and LIM2 sea-ice model, and ice concentration from OSTIA is assimilated with NEMOVAR using a 5-day assimilation window. The system generates five ensemble members by perturbing the forcing and assimilated ocean data. The native grid of the ocean model (same as the forecast system above) has a horizontal resolution of 0.25 degrees. A detailed description of the analysis setup, forcing, perturbations and assimilation scheme is given by Zuo et al. (2019).

AMSR-E/2 observations

To estimate observational uncertainties, we are also using the daily sea ice concentration product from the University of Bremen - Institute of Environmental Physics. This data is derived by applying the Artist Sea Ice (ASI) algorithm (Kaleschke et al., 2001) to microwave radiometer data from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) onboard the NASA/Aqua satellite (from June 2002 to October 2011), and the Advanced Microwave Scanning Radiometer 2 (AMSR-2) on the JAXA/Shizuku/GCOM-W1 satellite (from July 2012 on). Further details regarding the algorithm, data processing and validation, including an evaluation against the NASA algorithms, can be found in Spreen et al. (2008). Here, we will refer to this dataset simply as AMSR.

2.3 Methodology

For this study, we are comparing daily sea-ice outputs for years 2001 to 2020. All datasets have been remapped to a common 25-km EASE2 grid for a consistent comparison. Similarly, a common conservative land-ocean mask is applied prior to area integrating operations, in order to ensure that only grid points considered as ocean by all the datasets are included. Interpolations are done directly to the concentration datafiles, prior to any other operation, using bilinear interpolation from the Climate Data Operators software (Schulzweida, 2020). This also applies to section 2.4.2, where we interpolate the observations from OSISAF, and the ORAS5 analysis, to the common 1.5 degrees S2S grid and then to the 25-km EASE2 grid. This is referred to as 'Round-robin interpolation' in the main text.

Ice presence is determined from a given sea-ice concentration map by using a threshold, such that grid-cells with concentration equal to or above that threshold

are considered to have ice present. Here, we use the 20% concentration threshold instead of the more common 15%, in line with the assimilation steps used in the ECMWF prediction system (ECMWF, 2020). In the case of ensemble outputs, we take the average of such ice-presence maps over all the ensemble members to derive the sea ice probability (SIP). We also generate a climatological SIP forecast (hereafter referred to as CLIM) for each date by using the preceding 10 years of observations for the same date of the year as the ensemble. For computations that require a deterministic (binary) ice edge, we use the control run of the ensemble or the median of the probabilistic outputs (i.e., the 50% SIP contour), depending on the context.

Spatial Probability Score (SPS)

We primarily use the Spatial Probability Score as the validation metric for comparing the forecasts and analysis against the observational data. SPS can be defined as the spatial integral of the half-Brier score, derived using the square of the probability difference between forecast/analysis and observation at each grid point (Goessling and Jung, 2018). In practice, the observational data is deterministic (i.e., has a single ice presence state as determined using the concentration threshold), which is interpreted as a SIP map with probabilities of only 0 and 100%. The SPS of a deterministic output is also equivalent to the Integrated Ice Edge Error (IIEE; Goessling et al., 2016). Both SPS and IIEE are computed on a per-grid-cell-basis and summed over the entire domain, either Arctic (labelled as 'nh') or Antarctic ('sh')

Bias measurement and corrections

In section 2.4.4, we find the SIP bias on each day by taking the difference between the analysis SIP and the observational ice-presence (0 or 1) at each grid cell. To understand the mean pattern in each hemisphere, we then average this over a rolling 2-week window and over all the years in consideration, leading to the mean bias patterns as shown.

2.4 Results

2.4.1 SPS and IIEE

We first consider the SPS (and IIEE) of S2S_EC and ORAS5 against observational records from OSISAF, and compare these errors to those of the climatological forecast (CLIM) as a benchmark. The daily SPS results (Fig. B.1) show significant errors in both the forecast and analysis results. CLIM has the highest SPS against OSISAF as expected, while the SPS of S2S_EC is half that of CLIM on average. Similarly, the SPS of ORAS5 is lower than that of S2S_EC, but it is still around 38% of the SPS of CLIM. All three datasets show strong seasonality in the error patterns, although there is some interannual variability, in particular in the error of CLIM. For example, in 2007 and 2012 the record-low Arctic late-summer sea-ice extent implied that a climatology based on previous years was a particularly bad benchmark.

The mean annual seasonality of the SPS for each dataset is shown in Fig. 2.1. We can note that the SPS of both analysis and forecast generally peaks in the summer months in both hemispheres. This is consistent with Goessling et al. (2016) who relate this to seasonal variations of the ice-edge length. While ORAS5 typically has a lower SPS than S2S_EC, they are similar during the summer peak, especially in the Arctic. Measuring the IIEE of the ensemble median, or the control run, against OSISAF also results in similar error patterns (Fig. 2.2).

2.4.2 Interpolation error

To investigate the effects of the interpolation on ice edge errors, we interpolate the observational data from OSISAF to the common 1.5° grid, and back to the EASE-2 grid and measure the IIEE of this round-robin interpolated data (here referred to as S2S_OSISAF) against the original observation. Similarly, we also interpolate the control run of the ORAS5 data to the 1.5° grid and then to the EASE-2 grid, and measure the IIEE of the round-trip interpolated data (S2S_ORAS).

In Fig. 2.2, we can observe the increase in error resulting from the round-trip interpolation. S2S_OSISAF has an IIEE as high as 0.65 sq kms in the Antarctic (0.24 sq kms in the Arctic) against the original version of the same data, while the IIEE of S2S_ORAS is also higher than the error of the original ORAS5 control run by 20 to 25%. While the increase in error is not present during the summer months (which we address further in the discussion), the errors of S2S_ORAS closely resemble the initial errors of the forecasts (S2S_EC). Based on this, we can conclude

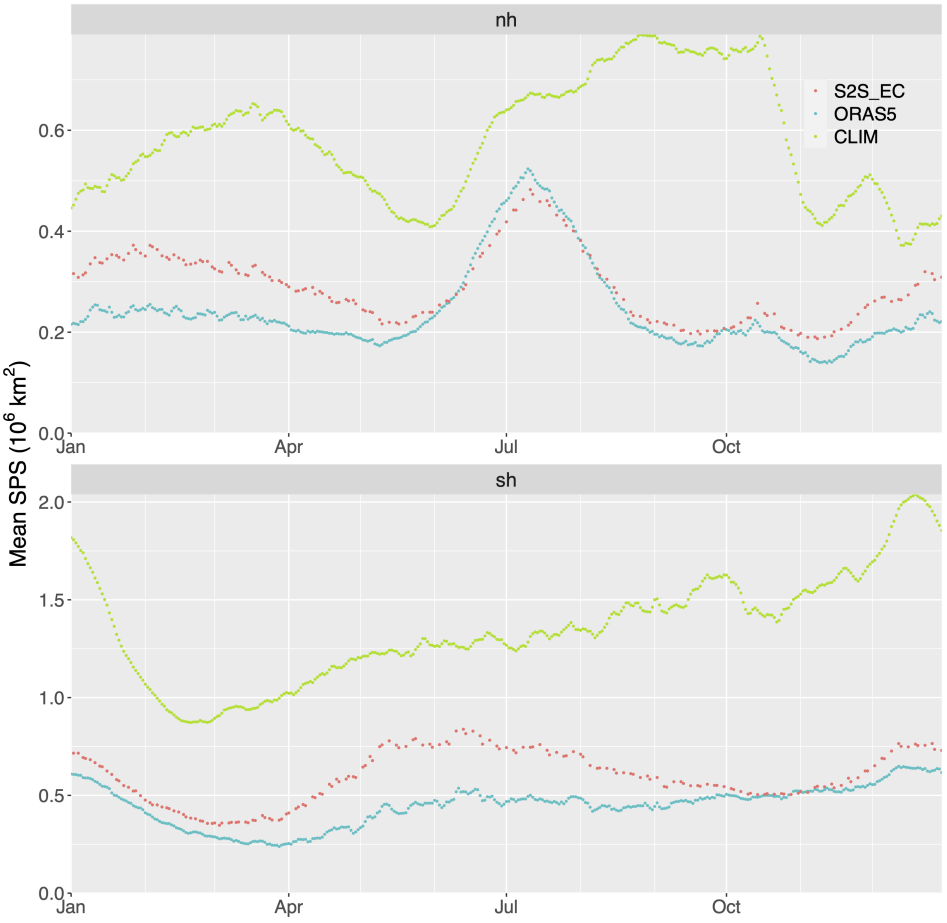

FIGURE 2.1: Mean seasonality of the Spatial Probability Score (SPS) in each hemisphere for S2S_EC, ORAS5 and CLIM, all measured against the observed ice edge from OSISAF.

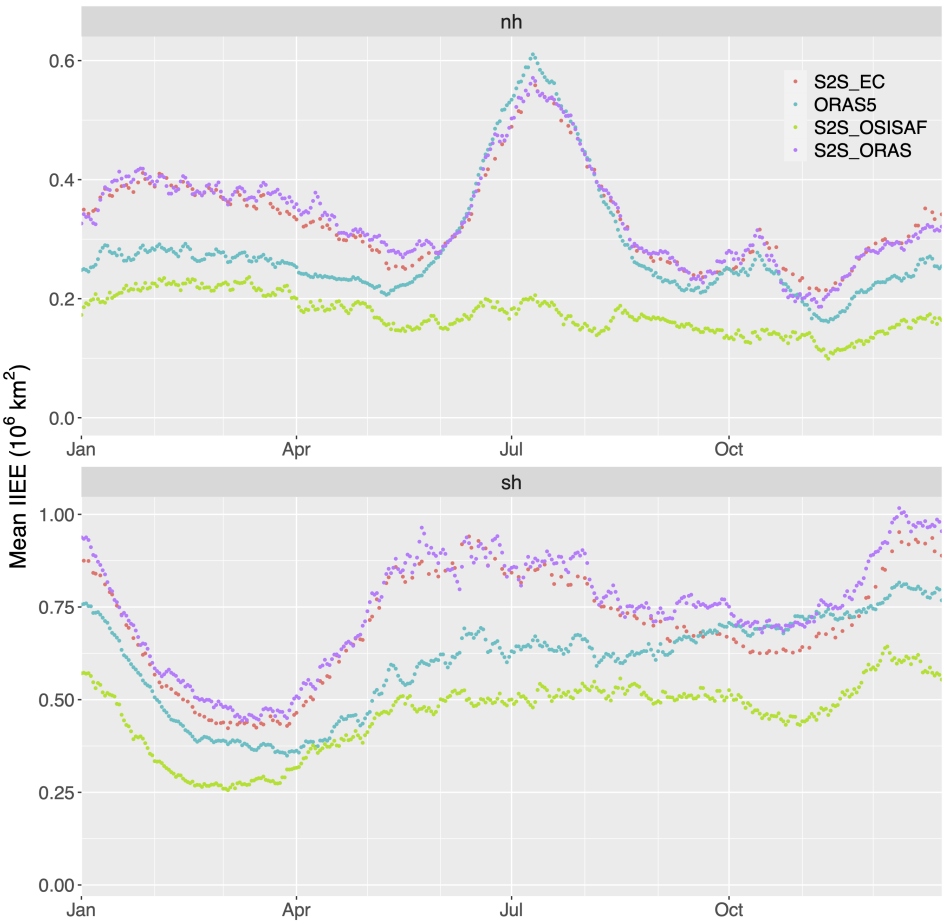

Figure 2.2: Mean seasonality of the Integrated Ice Edge Error (IIEE) in each hemisphere for S2S_EC, ORAS5, S2S_ORAS and S2S_OSISAF, all measured against the observed ice edge from OSISAF. CLIM is higher than all the other datasets and has not been shown here.

that the interpolation to the coarse 1.5° grid largely explains the difference between ORAS5 and S2S_EC, which also implies that the slight temporal mismatch between the day-1 forecasts and the initial states is not decisive. For the remainder of this study, we can thus focus solely on the analysis errors.

2.4.3 Comparison to observational uncertainty

In order to judge how far the ORAS5 ice-edge errors are larger than what should be considered acceptable due to uncertainties, we now examine observational uncertainties. To this end, we compare the mismatch between ORAS5 and OSISAF with the mismatch between OSISAF and another observational dataset, namely AMSR. We measure the pairwise area mismatch (or IIEE) between OSISAF, ORAS5 (ensemble median) and AMSR (Fig.B.2). While there are many outliers, the average mismatch between ORAS5 and OSISAF in the northern hemisphere is actually lower than the mismatch between OSISAF and AMSR. Overall, the analysis errors against both observations are within the range of the mismatch between the observations.

Comparing the pan-Arctic sea-ice extent between the three datasets, exemplary for 2018 and based on the 20% SIC threshold (Fig. 2.3), reveals that the ice extent of the ORAS5 ensemble median is generally within the range of the extent according to the two observational products. Throughout the year, the ice extent from OSISAF is higher than that from AMSR, and the ORAS5 extent is generally on the higher side of the distribution in winter and on the lower side in summer. The relationship is similar in other years, and the seasonal difference in over- or underestimation of the ice extent can also be seen in the bias maps discussed below.

2.4.4 Mean Bias

To better understand the spatial distribution of the analysis errors, we consider the daily SIP bias against the observations for each grid-cell. The mean SIP bias of ORAS5 relative to OSISAF in the Arctic (Fig. 2.4) is much larger in July than in January (in agreement with Fig. 2.2) and generally negative (i.e., ORAS5 estimates too little ice). In contrast, positive as well as negative biases occur in January. If we instead consider the bias against AMSR (Fig. B.3), we find that the July bias is less negative and the January bias is more positive, but the spatial pattern of the bias is similar as measured against OSISAF.

Similarly, when considering the mean SIP bias of ORAS5 relative to OSISAF in the Antarctic (Fig. 2.5), we again find a negative summer (January) bias and a

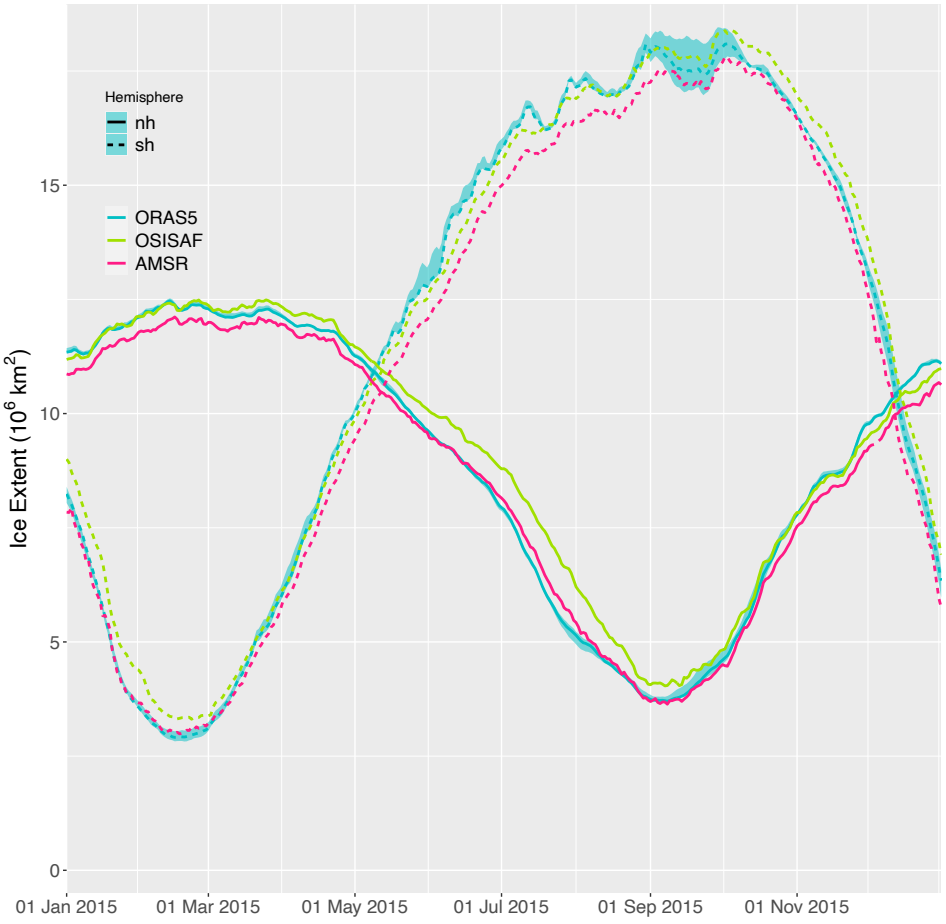

Figure 2.3: Ice Extent (in both hemispheres) computed using the 20% threshold of sea-ice concentration from OSISAF, AMSR(E) and ORAS5 (control run) for 2018. The ensemble spread of ORAS5 is represented by the shaded area.

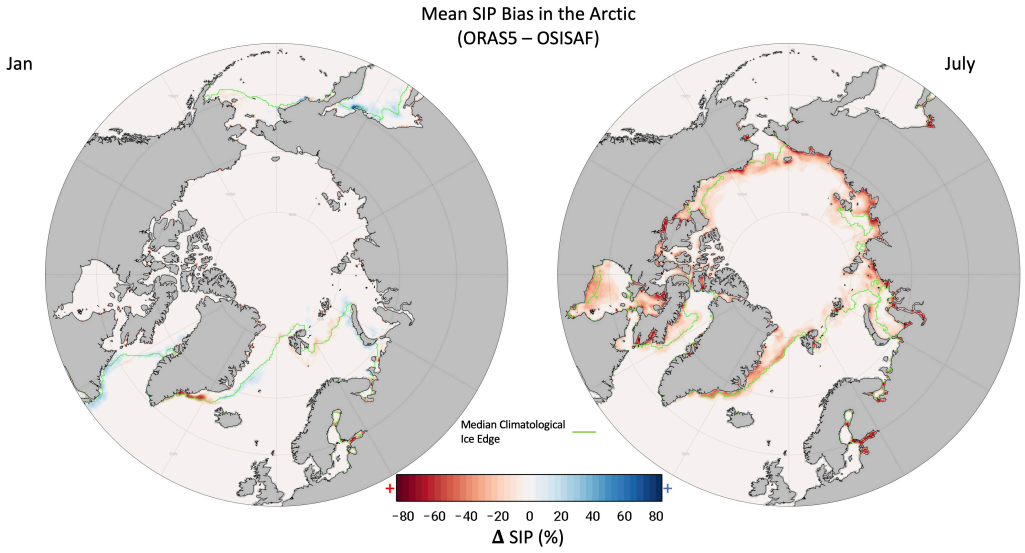

FIGURE 2.4: Mean Sea Ice Probability Bias of ORAS5 in comparison to OSISAF ice condition, for January (left) and July (right). Bias values larger than 0.8 or lower than -0.8 are marked with a plus sign (blue or red, for positive or negative bias) and the median climatological ice edge is shown as a green line in both maps.

weaker positive winter (July) bias. The January bias in the southern hemisphere is spread further away from the climatological median, compared to the north. We again see more positive bias AMSR (Fig. B.4) than against OSISAF. This aligns with what we learnt from comparing the ice-extent – that the ocean analysis estimates more ice than AMSR but less than OSISAF, and overestimates during the winter while underestimating during the summer.

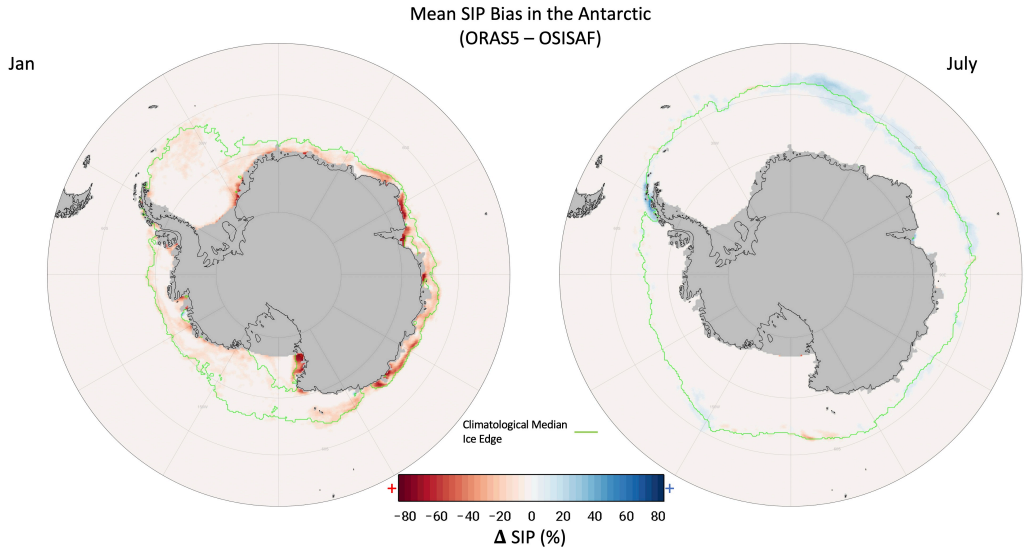

Figure 2.5: Same as Figure 4, but for the Southern Hemisphere.

2.5 Discussion and Conclusion

In this paper, we compare the sea ice edge in the initial day forecasts from ECMWF and the analysis from ORAS5 against observational records from OSI SAF by measuring the Spatial Probability Score (SPS) of the ensembles. We find that the SPS of the initial forecasts is roughly 50% of the Climatological error, while the analysis errors are around 38%. We show evidence that ice edge errors in the initial forecast have increased due to interpolation and would be as low as the analysis errors in the native grid. While the analysis errors are comparable to observational uncertainty, certain regions have strong biases, particularly related to underestimation near the ice-edge.

In Fig. 2.2, we see that the analysis errors do not increase due to the coarse grid interpolation during summer, unlike other months. This can be associated with the ice edge location in a given month and the area of the grid cells in this region. During summer, the ice edge is in the high Arctic, where the grid cells typically have a lower area and so the differences caused by the resolution are not as evident as the rest of the year. As a result of this, our conclusions differ from those in Zampieri et al. (2018) where the authors also considered the effects of resolution on SPS, but only compared summer forecasts.

While previous studies (Zampieri et al., 2018, 2019; Niraula and Goessling, 2021)

have measured the ice edge error of the S2S reforecasts by interpolating the ice-concentration observations to the S2S grid, here we use the higher-resolution EASE-2 grid as the common grid of reference. While interpolation of any kind is likely to introduce slight offsets in the data, interpolating to a lower resolution grid is likely to have bigger effects as the individual grid-cell areas are larger in a low-resolution grid. Furthermore, the interpolation algorithm used can also affect the resulting data – while we only discuss results of bilinear interpolation here, prior experiments with nearest neighbour interpolation resulted in even larger errors for S2S_OSISAF.

Independent of the interpolation, the analysis still has significant mismatch with OSISAF, which is likely inherited by the initial forecast. Computing the overestimation and underestimation components of the IIEE (as described in Goessling et al. (2016)) shows that the analysis errors in the summer months are mostly led by underestimation of ice presence, which corresponds with other studies (Rousset et al., 2015; Tietsche et al., 2015; Barthélemy et al., 2017; Day et al., 2022). The regions where this is most pronounced can be seen in the mean bias maps, as shown in Figs. 2.4 and 2.5.

We can do a simple bias correction experiment by subtracting this mean bias from the sea ice presence map on each date and then measuring the SPS. The resulting ice conditions have about 20% lower errors, with the largest improvements in the summer and particularly in the Arctic (Fig.B.5). However, we are not recommending any corrections to the analysis or attempting to perform some of the more recent and advanced bias correction procedures (e.g., Director et al., 2017; Dirkson et al., 2019), because we are not striving to make the model outputs more accurate but only to measure the difference caused by the mean bias. More detailed recommendations about possible improvements, while possible, are beyond the scope of this paper.

We have limited our study to the comparison of ice-edge location and Sea Ice Probability because of the central role the ice edge plays in atmosphere-ocean interactions and decision making, and also in order to align with prior work done by Zampieri et al. (2018, 2019). Future investigations should be carried out evaluating the ice concentration, and potentially also the sea surface temperature. A brief comparison of the SIP bias and SST bias (ORAS5 minus observed SSTs) resulted in similar regions of positive SIP bias (and warm SST bias) in the summer in the Arctic, but not in the Antarctic or the winter. Further research on the freezing conditions in the analysis may help point to the source of the bias. Similarly, an in-depth study scrutinizing the steps of the data assimilation process individually could be useful.

To conclude, we have evaluated the ice in initial day S2S forecasts from ECMWF alongside ocean reanalysis from ORAS5, and measured their errors against observational ice edge records from OSI SAF. We have demonstrated that a considerable part of the high initial forecast errors are due to the interpolation to a low-resolution grid, largely explaining the discrepancy between the initial forecast and analysis errors. The latter are within the range of observational uncertainties, but can still be reduced by simply subtracting a mean bias. Overall, there is still room for improvements in the ocean analysis and thus the sea-ice forecast initial states.

Quis custodiet ipsos custodes? (Who will watch the watchmen?).

Juvenal, Satire IV

3

Differences in sea ice edge representation in observational and analysis datasets

Abstract

There are currently several remote sensing sea ice concentration datasets and, from these observations, sea ice presence labels and model-based sea ice analysis can be produced. Prior studies have found that the use of distinct retrieval algorithms with the same radiometric measurements, or measurements from different instruments, results in differences in the sea ice concentration record among the datasets. Consequently, different sea ice edge position estimates can be expected from these products. Here, we analyse 5 observational and analysis datasets, and measure the mismatch between them pairwise in terms of Integrated Ice Edge Error (IIEE). We find significant mismatch between all pairs, with the highest IIEE between observations reaching 1.25 million square kms in the Arctic (for an ice extent of 7.58 million square kms) and 2.31 million square kms in the Antarctic (for an ice extent of 7.73

Chapter 3 is being prepared for submission to 'The Cryosphere'.

million square kms). Mismatch is highest during the summer in both hemispheres and mostly located along the Marginal Ice Zone. By comparing mean bias maps, we are also able to identify regions where certain datasets disagree with all others, suggesting an erroneous sea ice estimation in these regions. This study provides useful insights on the quality of single sea ice concentration estimates relative to the suite of products available, informs users on systematic shortcomings and drawbacks, and delivers valuable feedback to observation providers aiming to improve their sea ice products.

3.1 Introduction

Observational datasets of sea ice have an important role in shaping our knowledge of the polar system. At relatively short timescales, they are used for monitoring the evolution of the sea ice cover, understanding weather scale processes (Vihma, 2014; Chevallier et al., 2019), initializing and evaluating sub-seasonal to seasonal forecasts (Zampieri et al., 2018, 2019), or studying significant coupled events (Ludwig et al., 2019; Campbell et al., 2019). At longer time scales, the observed sea ice trends allow us to assess the effects of anthropogenic climate change in the Arctic and Antarctic (Stroeve et al., 2007). Consequently, errors and uncertainties in observational records can affect the quality of weather and climate predictions, as well as our ability to correctly verify these forecasts or even our understanding of the global sea ice trend.

Previous studies have compared the different algorithms used to retrieve sea ice concentration (SIC). Andersen et al. (2006) have shown Passive Microwave (PMW) retrieval algorithms to be particularly sensitive to atmospheric factors when measuring over new-ice or mixed ice-water conditions. Ivanova et al. (2014) compared 11 algorithms for PMW measurements of sea ice concentration and found that they differ in sea ice area by 0.0 to 1.3 million square kms. Similarly, by comparing results from the different algorithms against validation data from Ice charts and ENVISAT Advanced SAR data, Ivanova et al. (2015) found that none of the algorithms had a perfect estimate over both low-concentration and high-concentration regions. More recently, Meier and Stewart (2019) found the difference in sea ice extent between different observational products to be in the order of 0.5 to 1.0 million square kms.

Ice concentration (and consequently, ice presence) in the different datasets can disagree due to a variety of reasons. Mismatch can be brought about by differences in the underlying satellite measurements, generally passive microwave (PMW) data,

or the computation algorithms that the PMW measurements go through (Ivanova et al., 2014). Some processing chains also include other types of data for filtering or weather adjustments (Lavergne et al., 2019), allowing for additional post-production mismatch. On top of this, observation-based databases that use a combination of various sources (NIC, 2008) can have further sources of uncertainty.

Satellite-based observational records of sea ice are essentially derived by using the difference in brightness temperature, as measured by the PMW instruments, between open water and sea ice (Andersen et al., 2006; Spreen et al., 2008). Such retrievals are affected by the presence of water or wet snow on top of the sea ice in the observation window and can result in a lower or false sea ice concentration value (Kern et al., 2020). Uncertainty in brightness temperature is high in regions with a lower ice-cover, such as the Marginal Ice Zone (MIZ), and it is likely that this uncertainty differently affects the ice-presence generated from different classification algorithms (Andersen et al., 2006; Ivanova et al., 2014). Accordingly, it can be expected that the largest mismatch in sea ice presence also occurs within the MIZ, i.e. if one dataset estimates that there is no ice in a particular location, we would expect that the other dataset would estimate this location to be part of the MIZ, if not ice free.

Ice presence is commonly derived from SIC data by setting a threshold (generally 15%) and assigning all grid cells with SIC equal to or higher than the threshold to be 'ice-present'. The contour of the threshold concentration then becomes the effective ice edge for this dataset, while the area enclosed by this contour is the ice extent. End users can also set a higher concentration threshold as per needed, which changes the ice edge and ice extent. Some datasets, however, bypass the concept of sea ice concentration and binarize the formulation of its final product to ice presence/absence, thus reducing the adaptability, but also the ambiguity of the threshold method. The IMS Daily Northern Hemisphere Snow and Ice Analysis from the United States National Ice Center (NIC, 2008) is one such dataset that provides direct maps of ice presence, determined by human analysts using a variety of remote sensing and in situ data sources for the Arctic. Although it cannot be compared against other observational records in terms of concentration, we can compare the ice presence and ice edge directly in order to understand how well these datasets agree with each other.

In this study, we are deriving the ice-edge using PMW concentration records from three well established observational datasets. Two of these datasets, OSI SAF and NSIDC, are labeled as climate data records due to the long period of observations

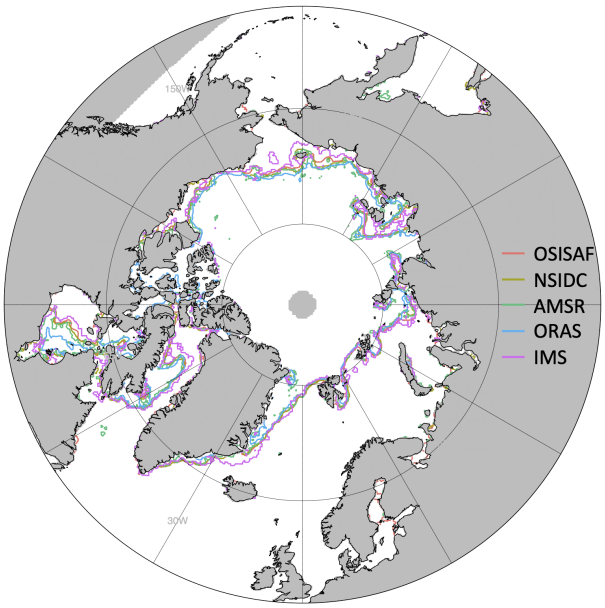

FIGURE 3.1: Ice Edge contours in the Arctic (for 15th of July, 2015) from the different datasets used in this study. The common mask, based on data availability of all the products, can be seen in grey.

for which they are available, and have been used for various climate and weather-related studies (Kohout et al., 2014; Hobbs et al., 2016; Roach et al., 2020; Notz and Marotzke, 2012; Mu et al., 2020). The third dataset, here labelled as AMSR, utilises a more recent retrieval method, but has already found wide acceptance and been used for several applications, such as operational forecast design (Zhao et al., 2020) and climate research (Frey et al., 2015; Campbell et al., 2019). All three datasets are described more in detail in section 3.2.

Alongside the primary observational data mentioned above, we are also including 2 secondary products that process the satellite observations further, either by using a numerical model for reanalysis (ORAS5) or by manual processing (IMS). ORAS5 is the most recent ocean reanalysis product from the European Centre for Medium Range Weather Forecasts (ECMWF; see Zuo et al., 2019). This reanalysis has been used for various applications including model initialisation (Johnson et al., 2019) and climate studies (Jackson et al., 2019). Similarly, the IMS (Interactive Multisensor

Snow and Ice Mapping System) dataset, mentioned before, is altogether different and involves human analysts using multiple sets of remote sensing and in situ data to manually generate ice-presence maps. Despite only existing for the northern hemisphere, IMS has been extensively used for ice edge verification (Lemieux et al., 2015; Wang et al., 2016; Barton et al., 2021). While both of these datasets are quite different from the PMW observational datasets, they are used in similar ways within the sea ice research community and therefore make an important contribution in our comparison.

Previous intercomparison studies have focused on differences in concentration, as well as on integrated quantities such as sea ice area and extent. Instead, we focus on the ice-edge location, considering a longer time period (2001 to 2020) and at pan-hemispheric scales. We have measured the mismatch between datasets using the Integrated Ice Edge Error (IIEE; Goessling et al., 2016), defined as the area integral of the regions where two ice-presence maps disagree. This is measured for each day, pairing each dataset against all others. We also analyse bias patterns between the pairs, in order to identify precisely the regions where the disagreement between the datasets is systematic. Comparing the bias between different pairs enables us to locate regions which are repetitively problematic for a dataset. Alongside, we also investigate the importance of the Marginal Ice Zone to the IIEE measurements, by measuring the cases where the disagreement in ice concentration is large enough to designate a region as ice free in one dataset while pack ice (or, ice outside of MIZ) in the other. The manuscript is presented as follows. In section 3.2, we briefly go over the datasets used in this study, and in section 3.3 we describe the Methodology. Section 3.4 presents the results from the pairwise comparisons. The manuscript ends with a discussion and short conclusion in section 3.5.

3.2 Data sources

This section provides a brief description of the 5 datasets considered in the study, and refers to the user guides and publications of each the datasets for more details. The reader should also note that not all of these sea ice products have an uncertainty estimate associated, which is why we have not considered this additional variable in our computations.

OSISAF

The first data we present is the global sea ice concentration record from the Ocean and Sea Ice Satellite Application Facility (OSI SAF). This data is based on brightness temperature from the SMMR, SSM/I and SSMIS instruments aboard various Defense Meteorological Satellite Program platforms, as well as ERA-Interim Data from ECMWF. Further details regarding the algorithm, processing and quality control are described by Lavergne et al. (2019). The data can be freely downloaded from the OSI SAF ftp servers, labeled as OSI-450 for years 1979–2015 and extended in near-real time as OSI-430 afterwards. The data is provided on a Lambert Azimuthal Equal Area polar projection, also known as the EASE2 grid, separately for each hemisphere and with a horizontal grid spacing of 25 km. For the rest of this manuscript, we will refer to this dataset as OSISAF.

NSIDC

We also use daily passive-microwave-derived sea ice concentration data from the National Snow and Ice Data Center (NSIDC) Climate Data Record. The input data used here comprises of brightness temperature from the same satellite sources as the OSI SAF data, but the retrieval algorithm is different and utilizes a combination of concentration estimates from the NASA Team-2 algorithm and NASA Bootstrap algorithm, both of which have been used widely in the sea ice community for many years. More details regarding these algorithms, data acquisition and processing can be found in the user guide Meier et al. (2021). The data can be freely downloaded from the NSIDC website and is available in 25 x 25 km polar stereographic projection for each hemisphere. Hereafter, this data is referred to as NSIDC.

AMSR-E/2

The next dataset included in this comparison is the daily sea ice concentration product from University of Bremen - Institute of Environmental Physics. This data is derived by applying the Artist Sea Ice (ASI) algorithm (Kaleschke et al., 2001) to microwave radiometer data from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) onboard the NASA/Aqua satellite (from June 2002 to October 2011), and the Advanced Microwave Scanning Radiometer 2 (AMSR-2) on the JAXA/Shizuku/GCOM-W1 satellite (from July 2012 onwards). Due to the advanced sensor and a new algorithm (allowing for retrievals at a higher frequency than previously done), this dataset has a high horizontal resolution. In Spreen et al.

(2008), the authors have given further details regarding the algorithm, data processing and validation, including an evaluation against the NASA algorithms. The data can be downloaded from the University of Bremen Sea Ice Remote Sensing Archive on 6.25 km polar stereographic hemispherical grids. We will refer to this dataset as AMSR henceforth in this manuscript.

ORAS5

ORAS5 is the near real-time component of the OCEAN5 reanalysis system, developed by ECMWF. It is based on the NEMO 3.4 ocean model coupled to the LIM2 sea ice model, and utilizes boundary conditions from ERA-Interim atmospheric reanalysis before 2015, and ECMWF forecasts from 2015 onwards. The sea ice concentration assimilated into ORAS5 comes from OSTIA (which itself is derived from OSISAF; see Donlon et al, 2007). The native ocean grid has a horizontal resolution of 0.5 degrees. A detailed description of the analysis setup, initial forcing and assimilation is given by Zuo et al. (2019). While this system is designed to generate 5 ensemble members by perturbing both initial forcing and assimilated data, we focus here only on the control member, henceforth labeled ORAS.

IMS

The final dataset in this comparison, the Daily Northern Hemisphere Snow and Ice Analysis from the US National Ice Center (NIC), reports the sea ice presence instead of the concentration. This dataset is produced by human analysts manually looking at all available images from satellites, ice charts or in-situ information, and judging "by eye if a cell or cells are more than 40% ice covered". It is also only available for the northern hemisphere for historical reasons. More details on the input data sources, processing steps and quality control is given by the user guide NIC (2008), which can be found alongside the analysis (in polar stereographic projection at 1 km, 4 km and 24 km resolutions) on the NSIDC data center. In our comparison, we are using the 4 km resolution data, available from February 2004 onwards, referred to as IMS in the rest of this manuscript.

3.3 Methodology

The datasets are all remapped to a common 25 km Lambert Azimuthal Equal Area grid using the bilinear remapping function in Climate Data Operator (Schulzweida,

2020) for a consistent comparison. Similarly, a common conservative land mask is applied to ensure that only grid points considered as ocean by all the datasets are used in the study (as can be seen in Fig.3.1 and Supplementary Fig.C.1). To determine the sea ice presence, we use the common 15% concentration threshold (for all except IMS, which is produced as a "ice or no ice" binary), and integrating the area over all grid cells where ice is present gives the ice extent measurement. Note that while the IMS user guide mentions 40% ice coverage within a grid cell as being the basis of 'ice-presence', similar prior studies, including those done by NSIDC / NIC researchers (Meier et al., 2015; Liu et al., 2020), have also used the 15% SIC threshold to determine corresponding ice edge in PMW data. Our initial tests using 40% SIC threshold for other observations did not lead to conclusive improvements in agreement (shown in Supplementary Fig.C.4).

We are using the Integrated Ice Edge Error (IIEE; Goessling et al., 2016) as the metric of comparison between each pair of data. IIEE is the integral sum of the areas of Overestimation and Underestimation (from the perspective of one dataset against the other). We measure these components individually on each date and take the sum to find the IIEE. The relative factor of the Overestimation or Underestimation components in the IIEE can also be analysed separately, as done by Zampieri et al. (2019). Note that these components are reversible (ie Overestimation from A to B equals Underestimation from B to A, and vice-versa).

To measure the bias patterns, we simply take the difference in ice-presence (1 for ice, 0 for no ice at each grid) between two datasets at each date. These are then averaged for each month over all the years to derive the mean bias maps shown in section 3.4.4. They have been presented here in 'Dataset 1 minus Dataset 2' format, such that positive (or negative) bias refers to cases where Dataset 1 has more (or less) ice present than Dataset 2 on average. Once again, the biases are reversible from the perspective of the other dataset (i.e. positive bias of A over B equals negative bias of B over A).

Further in the manuscript, we investigate the role of the Marginal Ice Zone on ice edge error. Following the definition used by Rolph et al. (2020), we consider the region of the ocean with ice-concentration between 15% and 80% to be the Marginal Ice Zone. To measure the IIEE outside the MIZ, we then measure the integrated area where one dataset predicts ice outside the MIZ (i.e. grid cells with SIC $> 80\%$), while the other dataset predicts no ice (i.e. SIC $< 15\%$).

3.4 Results

3.4.1 Ice Extent comparison

We start by simply comparing the ice extent given by each of the different datasets. The daily timeseries for the entire period (Supplementary Fig. C.2) shows that the datasets mostly agree on both the seasonality and the general trend of ice-presence in both hemispheres. Considerable differences in ice-extent in each hemisphere occur only around the maxima (generally in March for the Arctic and September for the Antarctic) and minima (vice-versa). Looking at the three observational records only, Arctic ice extent of OSISAF is larger by 1.20% (2.95% in the Antarctic) than NSIDC and by 6.14% (7.53% in the Antarctic) than AMSR on average, while NSIDC ice extent is larger by 4.91% (4.77% in the Antarctic) than AMSR.

To better understand the trend in ice-extent and the variation between the different datasets, we average the ice-extent for the first weeks of September and March for each hemisphere, as shown in Fig.3.2. In the Arctic, March ice-extent estimates are generally highest for NSIDC and OSISAF, and slightly lower for AMSR and ORAS. The ranking is similar in September, with OSISAF providing slightly higher extent estimates. IMS estimates are more variable, ranging from the lowest ice-extent (Mar 2006) to the highest (Sep/Mar 2014), possibly due to different data sources in use during these different periods. Similarly in the Antarctic, OSISAF ice extent is always highest, closely followed by NSIDC. During September ORAS generally comes next while AMSR has the lowest ice extent, but this is reversed during March. In September 2012, all datasets show a notable drop in ice-extent in the Arctic, agreeing well with the lowest record in all of satellite history (Smedsrud et al., 2017).

Using the March and September monthly average, we also computed the decadal trend of ice extent in the Arctic for each of the datasets, as shown in Supplementary Fig.C.3 (Antarctic ice-extent trends are not shown as they are not statistically significant). The trends are relatively similar for all datasets (except IMS), especially in March. On average, the mean September ice-extent in the Arctic is decreasing by 0.87 million square kms per decade. If the ice extent continues to decrease with this trend, we can expect the Arctic to be ice-free in September by 2070, if not earlier.

3.4.2 IIEE

We have measured the Integrated Ice Edge Error between each pair of datasets for the two hemispheres on each date. The daily IIEE timeseries between the pairs

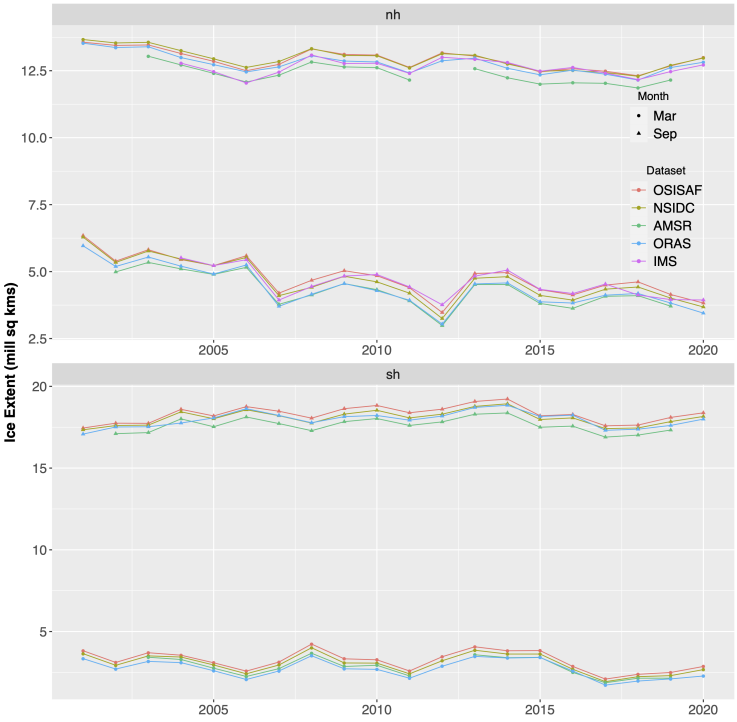

FIGURE 3.2: Monthly mean Ice extent according to the different datasets, for March and September in the Arctic (nh) and the Antarctic (sh).

(Supplementary Fig.3.3) show the range and seasonality of the mismatch. Most of the pairs have similar IIEE across years with no obvious trend, but there are exceptions such as OSISAF vs NSIDC (which shows a general increase in the Arctic summer IIEE from 2005 to 2011 and again from 2016 to 2019) and ORAS vs OSISAF or NSIDC (both of which show a slow decrease in mismatch from 2010 onwards), alongside occasional peaks. All datasets have a slight increase in IIEE against AMSR-2 (2012 onwards) compared to AMSR-E (until 2011), but the difference is not significant. Within the observations, the maximum IIEE in the Arctic is between OSISAF and AMSR at 1.25 million square kms (compared to the Ice Extent averaging at 7.58 million square kms between the two datasets) in summer 2011. Similarly in the Antarctic, the maximum was also between OSISAF and AMSR at 2.31 (compared to 7.78) million square kms, in austral summer 2018.

The mean seasonality of IIEE between each pair (Fig.3.3) shows that many of the peaks in mismatch are common among the different pairs. In the northern hemisphere, the maximum in mismatch occur in early July, with smaller peaks in October and December. The IIEE remains relatively constant in late winter and

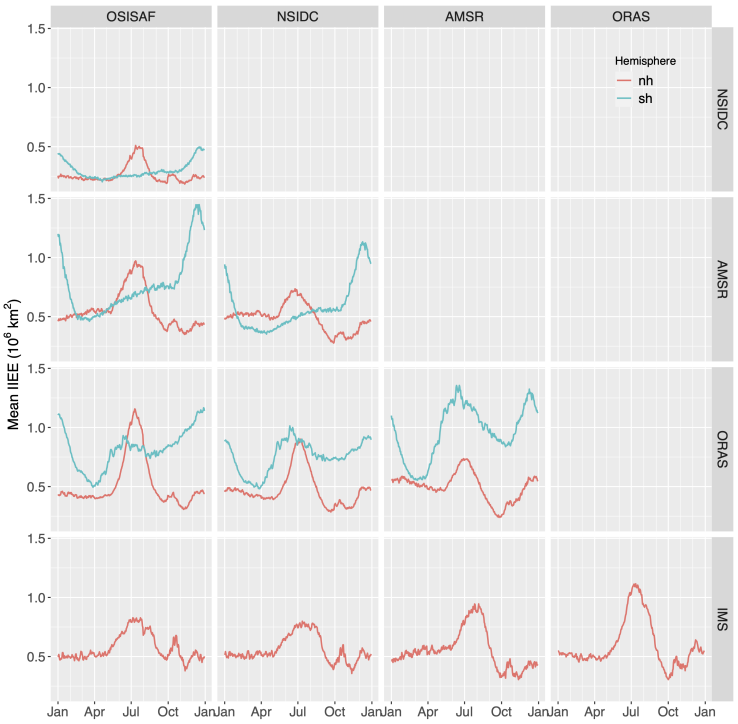

Figure 3.3: Mean annual seasonality of IIEE between the different dataset pairs. The pairs are marked by the row and column heads, while the hemispheres are differentiated by the colors.

Spring (i.e. between January and May) but is more variable during the freezing season. In the southern hemisphere, we again see a summer peak in mismatch in December. Most pairs have the lowest mismatch in March or April, a steady incline until November followed by a sharper growth in mismatch. All comparisons against ORAS in the southern hemisphere show a local maximum of IIEE in early June which is not present for the other pairs. This suggests that the Antarctic ice condition in ORAS is significantly misplaced this time of the year. We discuss these issues further in the manuscript.

As previously mentioned, IIEE is the sum of overestimation and underestimation. We can therefore also consider these components separately, or measure their contribution to the IIEE. The seasonality of Overestimation as a percentage of IIEE (shown in Supplementary Fig.C.6) essentially gives us an idea regarding how much of the error is dominated by additional ice presence in one of the datasets making each pair at the given time of the year. Here we see that overestimation by OSISAF makes a large contribution to the error through most of the year in pairs where it

is compared against another dataset. On the other hand, overestimation by AMSR is not contributing much to the error. ORAS has a more diverse behaviour, with a high overestimation contribution against AMSR in the winter, and lower contribution in the summer (shown reversed in the figure). It is important to note that a high percent contribution of Overestimation does not equal to a high error, but simply shows which dataset has more ice presence.

3.4.3 Mismatch outside of MIZ

Measuring IIEE in regions outside of MIZ as defined above, this area of mismatch can then be compared against the entire area of mismatch. For the most part, mismatch outside of MIZ is very low, especially between the observational datasets. In the case of OSISAF and NSIDC, for example, less than 1% of the mismatch was out of the MIZ, while for NSIDC and AMSR this was less than 5%. ORAS on the other hand, has a bigger ratio of mismatch outside of the MIZ, particularly in the southern hemisphere, and on average more often in comparison to AMSR. The high ratio in the southern hemisphere may be partly associated with the strong bias patterns we previously saw. Setting aside the analysis, our results support previous studies by confirming that disagreement between datasets is largely located within the MIZ.

3.4.4 Mean Bias

In order to better understand how the mismatch between the datasets is spatially distributed, here we evaluate the monthly mean bias between the datasets, focusing only on the months with the highest IIEE. In Fig.3.4, we can see the mean July bias for (a) OSISAF against NSIDC, (b) OSISAF over AMSR and (c) NSIDC over AMSR and identify the outliers in different regions. We see that OSISAF has a high positive bias against both NSIDC and AMSR in the Gulf of Finland and the Sakhalin bay, but AMSR and NSIDC agree well in this region. This suggests that OSISAF estimates of ice in this region are incorrect. Similarly, AMSR has substantial negative bias against OSISAF and NSIDC, most notably in the Lancaster and Barrow Straits of the Canadian Arctic Archipelago. However, OSISAF and NSIDC agree well in these regions, suggesting that the AMSR measurements are indeed too low here. These findings are also supported by the mean bias maps in Supplementary Fig.C.7, where we see very similar positive bias patterns for OSISAF minus IMS and negative bias for AMSR minus IMS.

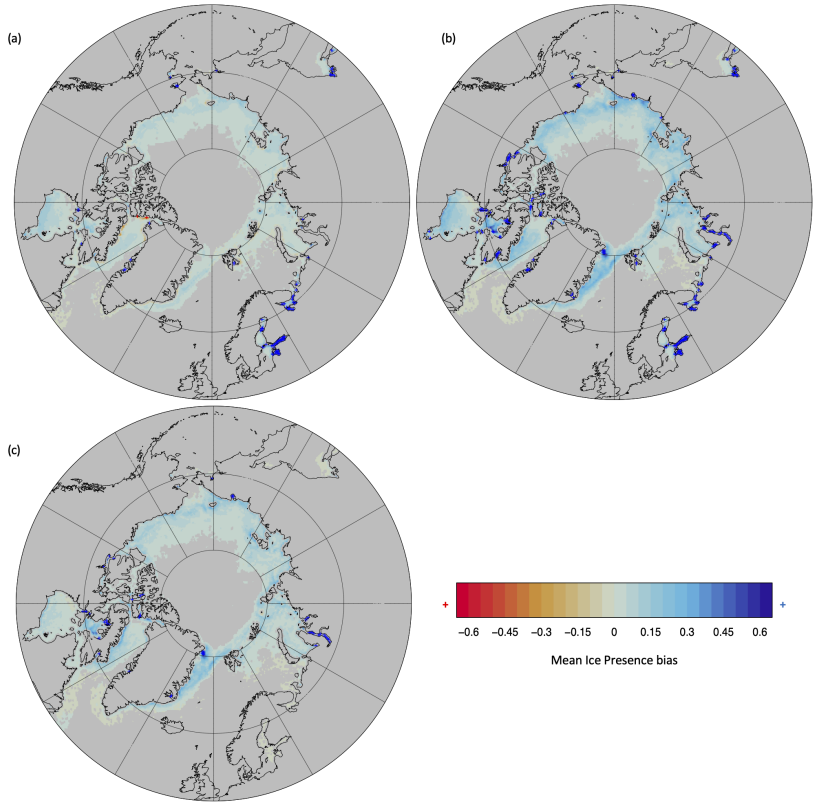

Figure 3.4: Mean ice-presence bias in the Arctic for the month of July, computed as (a) OSISAF minus NSIDC, (b) OSISAF minus AMSR and (c) NSIDC minus AMSR. Grid cells with bias higher than 0.6 (or lower than -0.6) have been marked with red (or blue) crosses, while grid cells that had 0 mismatch (or no ice at all) during this period have been left blank. Note that the bias patterns are reversible, i.e. positive bias for OSISAF minus NSIDC equals negative bias for NSIDC minus OSISAF.

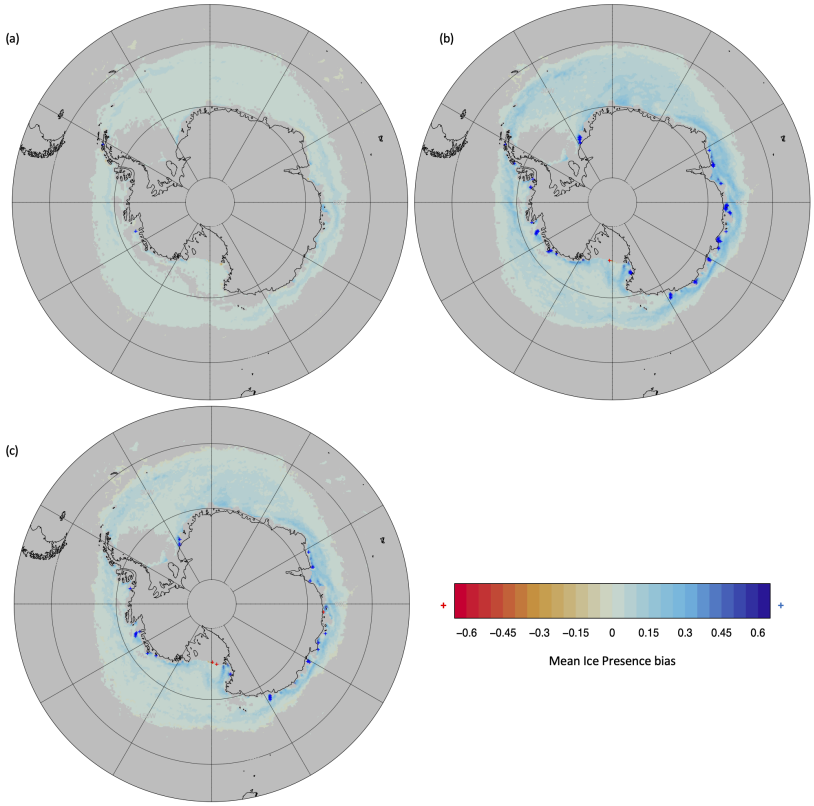

FIGURE 3.5: Same as Fig.3.4, but for the month of December in the southern hemisphere showing mean bias from (a) OSISAF minus NSIDC, (b) OSISAF minus AMSR and (c) NSIDC minus AMSR.

A similar comparison for the mean December bias in the southern hemisphere can also be seen in Fig.3.5. Once again, we can note that certain regions, mostly along the western coast, show a higher bias for OSISAF minus AMSR (inset b) and NSIDC minus AMSR (inset c). OSISAF and NSIDC agree relatively well in the southern hemisphere (or, at least there are no regions with a strong bias), suggesting that AMSR indeed underestimates ice presence in the aforementioned areas.

In Fig.3.6, we now compare the analysis from ORAS against the 3 observations. All three maps show a general negative bias (i.e. underestimation of ice by ORAS), although the distribution and strength of the negative bias decreases as we move from OSISAF to NSIDC to AMSR. Considering our analysis of OSISAF and AMSR ice estimates from the last two paragraphs, it is possible that some of the bias patterns (for example, those in the Gulf of Finland in inset a) are not accurate. Some of the bias we see in ORAS is present against all the observations, while

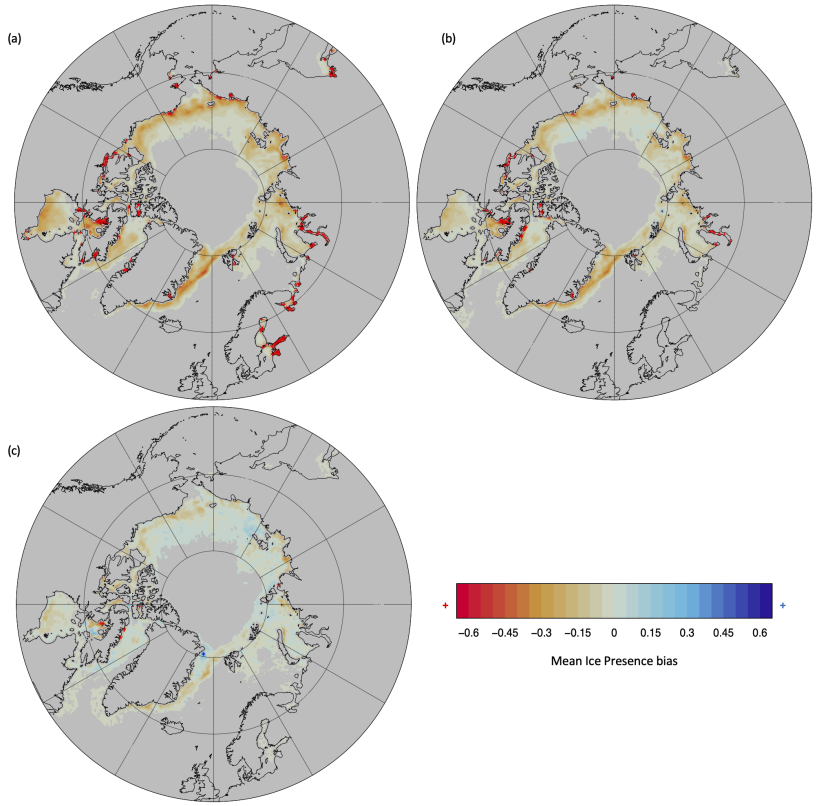

Figure 3.6: Same as Fig.3.4 (showing mean bias in July), but now comparing the ORAS analysis against the observations as (a) ORAS minus OSISAF, (b) ORAS minus NSIDC and (c) ORAS minus AMSR.

others that are not seen against AMSR can still be assumed to be true based on our understanding of AMSR underestimation. We see a very similar pattern in the southern hemisphere as well (Supplementary Fig.C.8), with some strong negative bias patterns against all observations (particularly in the Ross and Wedell Sea), while others seen against OSISAF and NSIDC, but not AMSR (such as those along the western coast). Based on these comparisons, we can reaffirm that ice estimates from ORAS during the summer are generally too low in both hemisphere.

3.5 Summary and Discussion

In this manuscript, we have analyzed the ice presence and ice edge data from several satellite based observational and analysis sources. By measuring the mismatch between the different datasets on each day using the Integrated Ice Edge Error, we

found significant differences in the estimation of ice presence, reaching up to 1 million square km in the northern hemisphere and 1.5 million square km in the south. For the observational datasets, the mismatch is primarily located within the MIZ such that less than 5% of the cases had one of the datasets estimating SIC more than 80% while the other one estimated no ice (i.e. SIC <15%), as suggested by prior research. For the ORAS reanalysis however, this ratio was much higher with up to 15% of cases falling out of the MIZ.

Among the observational datasets discussed here, OSISAF and NSIDC are very close. They use the same satellite sources for lower level data, similar algorithms for converting this data to SIC measurements and their resolutions are the same. However, they have a relatively constant minimum mismatch of 0.25 million square km, and a similar range of IIEE at both hemispheres unlike the other pairs. The validation report for OSISAF (M, 2019) specifically mentions that difference with NSIDC can arise due to the use of data from overlapping satellite measurements and more gap filling in OSISAF (It also mentions that the presence of a pole hole in the NSIDC data can be a cause of difference, but this is not the case in our study as our common mask omits the pole hole). Beitsch et al. (2015) also mention that the weather filter used in the 2 NSIDC algorithm can limit detection of low SIC conditions. These issues potentially lead to the mismatch measured here, alongside the positive bias patterns as seen in Fig.3.4(and C.7).

By comparison, the mismatch between AMSR and the other two observations, particularly OSISAF, is much higher. We saw that the ice extent from AMSR is generally lower than the others, the IIEE is more often led by underestimation from AMSR, and the mean bias maps also suggest more negative bias patterns. In Spreen et al. (2008), the authors have already noted that the ASI algorithm seems to underestimate the ice concentration compared to the Bootstrap and NASA-Team 2 algorithms (a combination of which results in the NSIDC data used here), particularly in the Antarctic. The additional difference with OSISAF could then be explained by the positive bias in OSISAF mentioned above.

The two other datasets that use different sets of data in the input also have significant mismatch against the purely observational datasets. Through most of the comparison, ORAS is underestimating ice presence in the summer. In the southern hemisphere, it has strong bias patterns in certain coastal regions, as we saw in Fig.3.6(and C.8), leading to the high IIEE. We have already discussed the differences in the bias patterns, and how some of them (particularly those against OSISAF) may represent error in the observation rather than the analysis. A similar assessment can

also be made for IIEE: ORAS has a lower IIEE against NSIDC than OSISAF (and even lower with AMSR). If OSISAF is overestimating ice presence as we suggested above, it could mean that the analysis actually corrects some of the overestimation issues and improves on the observations that it is based on.

Considering the mean bias of the 3 observational datasets against IMS (Fig.C.7), we saw that some of the regions where OSISAF has strong positive bias against the other observations also matches against IMS. However, there are other regions where both NSIDC and OSISAF have generally negative bias compared to IMS, particularly in Hudson Bay, Baffin Bay and off the southeastern coast of Greenland. IMS data is primarily an operational product tailored towards marine navigational needs. Since it is manually created by human agents, it suffers from human inconsistencies (Meier, 2017). It is possible for agents to make higher estimate of ice presence in regions (or time periods) of particular interest to emphasize safety concerns. Consequently, comparing IMS bias patterns on a pan Arctic basis requires more careful scrutiny.

We can see from the discussion so far that the mean bias maps are very useful for identifying regions where the datasets are likely to make a wrong estimate of ice. Here, we have been focused on the two months with the highest IIEE measurements in each hemisphere - July for the Arctic and December for the Antarctic (summer in both case). Looking at such mean bias maps for other months is also useful and can give more insights. We have seen that some of the regions discussed above (for example, the Ross Sea for ORAS and the Gulf of Finland for OSISAF) are problematic for all months, while some other regions are only problematic during certain times of the year (for example the Sea of Okhotsk during summer). This last issue can be intuitively associated with the seasonality of ice presence - a region that is certain to be completely covered with (or free of) ice during a specific time of the year will have very little mismatch during that period. For regions that are more likely to have consistent mismatch, these bias maps can be a good starting point to understand and correct such differences.

Our measurement of peak IIEE during summer months re-confirms (in ice edge terms) what previous studies have suggested in terms of concentration bias (Andersen et al., 2007; Tonboe et al., 2016; Kern et al., 2020). SIC retrieval algorithms can have difficulties identifying thin ice (Kwok, 2010) or differentiating between open water vs melt ponds on the surface (Ivanova et al., 2015). Rösel et al. (2012) have found that the peak in melt pond fraction in the Arctic occurs between the last week of June and early July (depending on the latitude), roughly corresponding with the

timing of peak IIEE in our results. However, the sensitivity to atmospheric conditions differs between the algorithms (Andersen et al., 2007; Beitsch et al., 2015) and the translation from concentration bias to IIEE is not linear due to the threshold approach (small changes near the 15% concentration will have a bigger impact on ice-edge error rather than a bigger change at higher concentrations).

For most research purposes, the observational datasets discussed here are used as 'truth', and forecast error is measured by verifying the prediction against these datasets. In our case, however, they are verified against each other, using a forecast verification metric (IIEE) for measuring mismatch. This brings forth two issues: Firstly, if we establish that the observations have errors in them (as we have suggested above), then certain forecasts measured against these observations might be more accurate than initially determined. This has already been discussed above for ORAS, where some of the underestimation in comparison to OSISAF is potentially more accurate (assuming OSISAF is overestimating ice in certain regions compared to reality).

Secondly, this highlights the issue of establishing 'truth'. Among our 3 (or 4, if we include IMS) observational datasets, we do not assume any to be the 'truth' and thus cannot state the precise error of each dataset. Further comparisons can include some of the in-situ ice observation data, from ship or airborne campaigns, and assume them to be 'truth'. However, such datasets do not fully cover the Arctic or Antarctic and are limited both temporally and spatially (Beitsch et al., 2015). Similarly, data from other non-PMW sources are limited to clear weather observations (Kern et al., 2003). As a result, using such datasets would not allow us to establish the long term pan-hemispheric records as we have done here.

Still round the corner there may wait
A new road or a secret gate.

J.R.R. Tolkien

4

Conclusions and Outlook

Verification of the sea ice edge has been the overarching theme of this thesis. We started by designing a new method of creating reference forecasts based on Spatially Damped Anomaly Persistence (SDAP). The forecasts from this method were then verified against observations (from OSISAF), and the results compared to the verification results of operational model forecasts from the S2S project. In this process we showed that forecasts from ECMWF were the only ones outperforming SDAP forecasts (and Climatology, confirming prior results), but we also noted a large initial error in these forecasts. The next part of the study was then focused on verification of the initial ice edge of the ECMWF forecasts, further including the reanalysis (ORAS5) that was used to initialise these forecasts. Here we showed that the errors are present in both forecast and analysis, considered their spatial and temporal distribution and also attempted a simple correction method to correct for mean biases. We then followed this in the next part of the study by measuring the mismatch between several different observational datasets in terms of ice edge. This essentially serves as verification of the observational data from OSISAF that had been used as 'truth' so far in the thesis (although we do not qualify any of the other observational datasets as 'truth' here). Overall, we start the thesis by measuring the skill in the forecasts, and end it by measuring the disagreements in the observations,

as stated in the title.

In **Chapter 1**, we were successfully able to create a reference forecast of sea-ice edge by simply measuring the historical probability of ice presence, measuring the anomaly to this probability along the current ice edge (at the time of initialisation) and spatially passing a damped version of this anomaly to other grid cells. The method has a relatively low demand, needing only historical and current ice edge information, yet skillful enough to create forecasts that outperform both Climatology and several fully coupled dynamic models at considerable lead times. In terms of the initial objectives (page xix), both questions can be answered – yes.

In **Chapter 2**, our focus was on the initial ice edge of the ECMWF forecasts and the ORAS5 analysis. While the presence of the initial error (in the case of the forecast) had already been shown in chapter 1, here we analysed the errors (against OSISAF) more in depth, focusing on when and where they are arising. To answer the initial questions (page xix), we showed that errors are highest in the summer, mostly close to the ice edge and often dominated by low ice in the case of the forecast and the analysis. We showed that interpolation errors explain the gap between the analysis and forecast errors, and also considered the mean bias patterns for the summer months. Subtracting the mean bias from the sea ice probability led to improvement (in terms of SPS and IIEE) for both datasets, indicating that there are systematic errors that can and should be addressed.

In **Chapter 3**, we took a step backwards and asked whether the observations from OSISAF (assumed to be the 'truth' so far) itself agreed well with other observational and analysis data. Our initial question (page xx) had a surprising answer – while some pairs of observational data agreed well with each other, other pairs had a strong disagreement in their estimate of the ice edge, most notably between OSISAF and AMSR. Mismatch between observations is also highest during the summer, and is mostly located within the Marginal Ice Zone. Comparing the bias maps between different pairs also allowed us to identify outliers within the datasets. Based on this approach, we infer that OSISAF is estimating too much ice compared to the other datasets, while AMSR is estimating too little.

The findings from chapter 3 actually challenge some of the results that we saw in chapter 1 and 2. In the first chapter, verifying the S2S model forecasts against other observational data might have a different error compared to the verification against OSISAF. Since the SDAP forecasts are generated using the initial conditions and climatology from OSISAF, they might be carrying forward some of the issues and might have a higher error against other observations. Similarly in chapter

2, we initially found that the initial forecasts and analysis both had too little ice compared to the observation. Yet, now we suspect that OSISAF may be estimating too much ice (in comparison to reality), meaning that the initial forecast and analysis errors could potentially be lower than what we previously found. Interestingly, both ECMWF forecasts and ORAS5 analysis use the ice conditions from OSTIA, which is itself derived from OSISAF ice concentration. Any intermediate changes (going from the near real time data used by OSTIA and hence the analysis vs the gap filled observations used here) should not be significant, which means the disagreement comes largely from how the ocean model handles the initial conditions.

We can reframe these contradictions as one of the two major challenges in our study – that of establishing a truth. Chronologically, the earlier chapters of the thesis assumed OSISAF to be the truth, while chapter 3 did not have a specific truth. Both cases brought their own issues, whether it is the likelihood of measuring error wrongly or the inability to measure error objectively. Observational datasets themselves have verification processes, which is generally done within a limited time or place, often relying on in situ or radar based data. However, the coverage of such datasets is quite limited and we would not be able to use them at pan-hemispheric or longer timescales.

The second major challenge in our study is caused by the different grids and resolution of the datasets being compared. In each of these studies, we needed to interpolate some of the data from their native grid to a different one for common comparison. The S2S forecasts used in chapter 1 and 2 had already been interpolated to the common lat/lon grid with a horizontal resolution of 1.5 degrees as part of the S2S data storage . While the comparison of SDAP forecast performance against the S2S forecasts was done in this common grid, we decided that the 25 km Lambert Azimuthal Equal Area polar projection (which is the OSISAF native EASE-2 grid) was a more appropriate choice as the common grid of comparison for the rest of the thesis. All the interpolations done in this thesis used the bilinear interpolation feature of CDO as tests with other interpolation schemes (nearest neighbour or distance weighted) showed higher error growths.

In chapter 2, we interpolated OSISAF observations to the S2S common grid and then back to the OSISAF grid. This double-interpolated data had a mean IIEE of 0.17 million sq kms in the northern hemisphere and 0.46 million sq kms in the southern hemisphere against the original OSISAF data. However, similarly interpolating the OSISAF observations to a higher resolution 6.25 km polar stereographic grid and back led to a much lower error growth (0.02 million sq kms in the Arctic

and 0.07 million sq kms in the Antarctic). This suggests that interpolation to a lower resolution grid is more likely to cause issues than that to a higher resolution grid. Since the EASE2 grid resolution is sufficiently high, we found it acceptable to interpolate the observation and analysis datasets in chapter 3 to this grid. Ideally, each forecast and observation could be verified in their original grid, minimising the uncertainties introduced by interpolation. Given that the datasets are all in different grids, however, we need to continue using this technique for comparisons.

One potential solution for the interpolation issue could be to perform the computations in both grids (the observation and the forecast or the two different observations) and take an average of the resulting error score. As such, care should be applied to ensure that the area of comparison and the land-ocean mask are constant across both comparison grids. If we are comparing the error score of two different forecasts (or different pairs of observations), it might even be necessary to perform the computations in all of the different grids to ensure that the results are not influenced by the differences in grid size. This can get complicated and we have not fully explored this method yet, but it bears consideration.

Given the variety of observations that we have assembled on the same grid (and assuming that interpolation has not introduced significant errors), we can create a very useful dataset by combining them into a probabilistic record of observed ice presence. This would find particular use in ice edge verification. Going back to the definition of SPS, note that we take the difference in probability of ice presence between the forecast and the observation. While we currently use a binary probability for the observation, using a probabilistic observation would be more robust. We could also address the issue in chapter 3 with establishing truth by using such a probabilistic observation and measuring the SPS of each dataset against it. Further issues need to be address prior to the creation and use of a combined probabilistic observation (such as the uncertainty in certain datasets, repetition of observation in the verification ensemble, and the choice of concentration threshold). This work is still ongoing and has not been included as part of this thesis, but we hope to follow up some of our results with this new approach.

One other piece that has mostly been absent in this thesis, is the use of forecast calibration. There have been promising developments in this field over the last few years (e.g., Director et al., 2017; Dirkson et al., 2019) but the application of such methods is not always straightforward or computationally efficient. The simple bias correction method we used in chapter 2 resulted in lowered errors, although the aim of this exercise was to identify systematic errors rather than generate well calibrated

sea ice outputs. The probabilistic anomalies used in the SDAP method could also possibly be adapted for calibration of S2S forecasts. We have shown that anomalies have considerable memory and further research in using this information to correct forecast bias can potentially take calibration methods down a new road.

I'd like to end this thesis by repeating what was mentioned in the Introduction – the ice edge is of immense interest to multiple stakeholders, and getting it right is very important, in both forecasts and observations. The sea ice community has made immense progress in the past few years, and while our current forecasts lose skill within a month (or two) and our observations disagree amongst each other, there have been constant improvements. This thesis is also a contribution to that effort and I hope that our findings can be put to use in closing the errors between forecasts, anlayses, observations, and reality.

Acknowledgments

I would like to use this opportunity to express my gratitude for all the people who have helped me in different ways along this PhD journey, starting with Helge Goessling. Helge has been the most amazing supervisor throughout my PhD - kind in his words, supportive in his actions – and I am extremely grateful for our time working together. I am also indebted to Lorenzo Zampieri, whose shoulders I have stood on for a lot of my work. Grazie Lorenzo! Thanks also to Thomas Jung for his guidance and motivation, and to Barbara Casati for her warm support and encouragement. Thank you to Martin Losch, Lars Kaleschke, J-F Lemieux, Thomas Rackow, Longjiang Mu, Bruno Tremblay, Steffen Tietsche and Jonny Day, for all their help and contribution to my Phd progress. Thank you to Katharina and Ke, for making things work. You are / have been a lifeline for us in the section. Thank you also to Claudia and POLMAR, for all the opportunities you've made possible. Finally, thanks to Dr. Gunnar Spreen and Dr. Thomas Jung for kindly agreeing to review my thesis.

I also want to remember all the friends and family who have been there for me through these past few years. Danke zuerst an meine schönen Mitbewohni – Lea, Waleed, Sadi, Lucca, Flo, Ayman, Wiebke, Laura and Frida for putting up with me. Gracias a mis hermanos, Kardeşlerim – Mariano, Onur, for Al Ways Inspiring. Thanks to my crew/mates – Jenna, JC, Margarita, Rémi, Diego, Olga, Deniz, Miri, Franzi, Thommy, Nattu for bringing shots of joy in my life. Shkran Abed, for your alshakhsia habibi. Thanks also to my colleagues in SSIP/Team-Helge – Lorenzo, Longjiang, Marylou, Valentin, Martina, Sveta, Antonio, Simon, as well as others in CliDyn, for all their support. The list is endless, so I will simply thank all the friends and PhD students in AWI and Bremen for these 3+ years.

The journey started long before Germany. So, I want to thank my friends in Canada – Ian, Anna, Julia, Lana, Michael, Priscilla, Ken, Xime, Andy, Cody, Carl, Mikaila + Geisterfamily, Tara, everyone in MOC, team Bruno (esp. Charles and Mathieu), CMC (esp J-F), Jon, Brian, and everyone at McGill/Montréal for those formative years. I also want to thank my people in Nepal – particularly Rénasha and Gokul - and elsewhere in the world - Djibril, Ken, Anu, Badi + family, Kishor Mama + Maiju, Kayam, Kavyam, Rohit, Aviral/Joondroid, Asin, Ranav Dai, Rishi Dai, Basu Mama. Sabailai Muri Muri Dhanyabad!

Merci en particulier à Chloé, pour tout l'amour. Très heureux de t'avoir dans ma vie et trop hâte de vivre pleins d'aventures de la vie ensemble.

And finally, thank you to my family – Baba (Mohan), Mamu (Bimala), Nanu (Bibechana) and Muwa (Subhadra). They have been a constant source of support and encouragement for me, before and during my PhD. I have come a long way in life because of my parents and how strongly they valued education. And so, I dedicate all my work, and this thesis, to them.

Binochan बिनोचन

References

Aagaard, K. and Carmack, E. C. (1989). The role of sea ice and other fresh water in the Arctic circulation. *Journal of Geophysical Research*, 94:14485. x

Abram, N. J., Wolff, E. W., and Curran, M. A. (2013). A review of sea ice proxy information from polar ice cores. *Quaternary Science Reviews*, 79:168–183. xi

Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G., and Pedersen, L. T. (2007). Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic sea ice. *Journal of Geophysical Research*, 112. 55, 56

Andersen, S., Tonboe, R., Kern, S., and Schyberg, H. (2006). Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using numerical weather prediction model fields: An intercomparison of nine algorithms. *Remote Sensing of Environment*, 104:374–392. 40, 41

Balan-Sarojini, B., Tietsche, S., Mayer, M., Alonso-Balmaseda, M., and Zuo, H. (2019). Towards Improved Sea Ice Initialization and Forecasting with the IFS. 2

Bann, S. (2009). "When I was a Photographer": Nadar and History. *History and Theory*, 48:95–111. xii

Barthélemy, A., Goosse, H., Fichefet, T., and Lecomte, O. (2017). On the sensitivity of Antarctic sea ice model biases to atmospheric forcing uncertainties. *Climate Dynamics*, 51:1585–1603. 25, 36

Barton, N., Metzger, E. J., Reynolds, C. A., Ruston, B., Rowley, C., Smedstad, O. M., Ridout, J. A., Wallcraft, A., Frolov, S., Hogan, P., Janiga, M. A., Shriver, J. F., McLay, J., Thoppil, P., Huang, A., Crawford, W., Whitcomb, T., Bishop, C. H., Zamudio, L., and Phelps, M. (2021). The Navy's Earth System Prediction Capability: A New Global Coupled Atmosphere-Ocean-Sea Ice Prediction System Designed for Daily to Subseasonal Forecasting. *Earth and Space Science*, 8. 43

Bay-Larsen, I., Bjørndal, T. G., and Hermansen, E. A. T. (2020). Mapping ice in the Norwegian Arctic – on the edge between science and policy. *Landscape Research*, pages 1–15. xvi

Beitsch, A., Kern, S., and Kaleschke, L. (2015). Comparison of SSM/I and AMSR-E Sea Ice Concentrations With ASPeCt Ship Observations Around Antarctica. *IEEE Transactions on Geoscience and Remote Sensing*, 53:1985–1996. 54, 56

Blanchard-Wrigglesworth, E., Bitz, C. M., and Holland, M. M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, 38:n/a–n/a. 15

Budikova, D. (2009). Role of Arctic sea ice in global atmospheric circulation: A review. *Global and Planetary Change*, 68:149–163. x

Bushuk, M., Winton, M., Bonan, D. B., Blanchard-Wrigglesworth, E., and Delworth, T. L. (2020). A Mechanism for the Arctic Sea Ice Spring Predictability Barrier. *Geophysical Research Letters*, 47. xv

Campbell, E. C., Wilson, E. A., Moore, G. W. K., Riser, S. C., Brayton, C. E., Mazloff, M. R., and Talley, L. D. (2019). Antarctic offshore polynyas linked to Southern Hemisphere climate anomalies. *Nature*, 570:319–325. 40, 42

Carrieres, T. (2017). Sea Ice Analysis and Forecasting Towards an Increased Reliance on Automated Prediction Systems. *Sea Ice Analysis and Forecasting*, 2. xiii, xiv

Carsey, F. (1989). Review and status of remote sensing of sea ice. *IEEE Journal of Oceanic Engineering*, 14:127–138. xii

Casati, B., Dorninger, M., Coelho, C. A. S., Ebert, E. E., Marsigli, C., Mittermaier, M. P., and Gilleland, E. (2022). The 2020 International Verification Methods Workshop Online: Major Outcomes and Way Forward. *Bulletin of the American Meteorological Society*, 103:E899–E910. xvii, xviii

Chevallier, M., Massonnet, F., Goessling, H., Guémas, V., and Jung, T. (2019). Chapter 10 - The Role of Sea Ice in Sub-seasonal Predictability. xv, 40

Comiso, J. C. (2012). Large Decadal Decline of the Arctic Multiyear Ice Cover. *Journal of Climate*, 25:1176–1193. 24

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L. (2008). Accelerated decline in the Arctic sea ice cover. *Geophysical Research Letters*, 35. x

Day, J. J., Keeley, S., Arduini, G., Magnusson, L., Mogensen, K., Rodwell, M., Sandu, I., and Tietsche, S. (2022). Benefits and challenges of dynamic sea ice for weather forecasts. *Weather and Climate Dynamics*, 3:713–731. 25, 36

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

Director, H. M., Raftery, A. E., and Bitz, C. M. (2017). Improved Sea Ice Forecasting through Spatiotemporal Bias Correction. *Journal of Climate*, 30:9493–9510. 21, 36, 60

Dirkson, A., Merryfield, W. J., and Monahan, A. H. (2019). Calibrated Probabilistic Forecasts of Arctic Sea Ice Concentration. *Journal of Climate*, 32:1251–1271. 21, 36, 60

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, 116:140–158. 24

Dukhovskoy, D. S., Ubnoske, J., Blanchard-Wrigglesworth, E., Hiester, H. R., and Proshutinsky, A. (2015). Skill metrics for evaluation and comparison of sea ice models. *Journal of Geophysical Research: Oceans*, 120:5910–5931. xvii, 4, 7

ECMWF (2020). IFS Documentation CY47R1 - Part II: Data Assimilation. 28

Eicken, H. (2010). Indigenous Knowledge and Sea Ice Science: What Can We Learn from Indigenous Ice Users? *SIKU: Knowing Our Ice*, pages 357–376. xi

Frey, K. E., Moore, G. W. K., Cooper, L. W., and Grebmeier, J. M. (2015). Divergent patterns of recent sea ice cover across the Bering, Chukchi, and Beaufort seas of the Pacific Arctic Region. *Progress in Oceanography*, 136:32–49. 42

Goessling, H. F. and Jung, T. (2018). A probabilistic verification score for contours: Methodology and application to Arctic ice-edge forecasts. *Quarterly Journal of the Royal Meteorological Society*, 144:735–743. xvii, 4, 6, 12, 24, 28

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T. (2016). Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, 43:1642–1650. xvii, 3, 4, 6, 14, 28, 29, 36, 43, 46

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., Salas y Mélia, D., and Tietsche, S. (2014). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142:546–561. xiii

Haas, C. (2016). Sea ice thickness distribution. *Sea Ice*, pages 42–64. ix

Haine, T. (2008). What did the Viking discoverers of America know of the North Atlantic Environment? *Weather*, 63:60–65. xi

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

Hobbs, W. R., Massom, R., Stammerjohn, S., Reid, P., Williams, G., and Meier, W. (2016). A review of recent changes in Southern Ocean sea ice, their drivers and forcings. *Global and Planetary Change*, 143:228–250. 42

Huntington, H. P., Gearheard, S., Holm, L. K., Noongwook, G., Opie, M., and Sanguya, J. (2016). Sea ice is our beautiful garden: indigenous perspectives on sea ice in the Arctic. *Sea Ice*, pages 583–599. xi, xiii

Ivanova, N., Johannessen, O. M., Pedersen, L. T., and Tonboe, R. T. (2014). Retrieval of Arctic Sea Ice Parameters by Satellite Passive Microwave Sensors: A Comparison of Eleven Sea Ice Concentration Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 52:7233–7246. 25, 40, 41

Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M. (2015). Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations. *The Cryosphere*, 9:1797–1817. 40, 55

Jackson, L. C., Dubois, C., Forget, G., Haines, K., Harrison, M., Iovino, D., Köhl, A., Mignac, D., Masina, S., Peterson, K. A., Piecuch, C. G., Roberts, C. D., Robson, J., Storto, A., Toyoda, T., Valdivieso, M., Wilson, C., Wang, Y., and Zuo, H. (2019). The Mean State and Variability of the North Atlantic Circulation: A Perspective From Ocean Reanalyses. *Journal of Geophysical Research: Oceans*, 124:9141–9170. 42

Johnson, L. (1983). The FRAM expeditions: Arctic ocean studies from floating ice, 1979–82. *Polar Record*, 21:583–589. xi

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H., and Monge-Sanz, B. M. (2019). SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12:1087–1117. xiii, 42

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., and Yang, Q. (2016). Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. *Bulletin of the American Meteorological Society*, 97:1631–1647. 2

Kaleschke, L., Lüpkes, C., Vihma, T., Haarpaintner, J., Bochert, A., Hartmann, J., and Heygster, G. (2001). SSM/I Sea Ice Remote Sensing for Mesoscale Ocean-Atmosphere Interaction Analysis. *Canadian Journal of Remote Sensing*, 27:526–537. 25, 27, 44

Kern, S., Kaleschke, L., and Clausi, D. (2003). A Comparison of Two 85-GHz SSM/I Ice Concentration Algorithms With AVHRR and ERS-2 SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 41:2294–2306. 56

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., and Tonboe, R. (2020). Satellite passive microwave sea-ice concentration data set inter-comparison for Arctic summer conditions. *The Cryosphere*, 14:2469–2493. 41, 55

Kohout, A. L., Williams, M. J. M., Dean, S. M., and Meylan, M. H. (2014). Storm-induced sea-ice breakup and the implications for ice extent. *Nature*, 509:604–607. 42

Kwok, R. (2010). Satellite remote sensing of sea-ice thickness and kinematics: a review. *Journal of Glaciology*, 56:1129–1140. xii, 55

Kwok, R. and Rothrock, D. A. (2009). Decline in Arctic sea ice thickness from submarine and ICESat records: 1958-2008. *Geophysical Research Letters*, 36:n/a–n/a. xi

Landrum, L. and Holland, M. M. (2020). Extremes become routine in an emerging new Arctic. *Nature Climate Change*. x, xi

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T. (2019). Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. *The Cryosphere*, 13:49–78. 5, 26, 41, 44

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., and Bauer, P. (2019). Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system. *Quarterly Journal of the Royal Meteorological Society*, 145:3432–3454. 24

Lemieux, J., Beaudoin, C., Dupont, F., Roy, F., Smith, G. C., Shlyaeva, A., Buehner, M., Caya, A., Chen, J., Carrieres, T., Pogson, L., DeRepentigny, P., Plante, A., Pestieau, P., Pellerin, P., Ritchie, H., Garric, G., and Ferry, N. (2015). The Regional Ice Prediction System (RIPS): verification of forecast sea ice concentration. *Quarterly Journal of the Royal Meteorological Society*, 142:632–643. xiii, 43

Liang, X. and Losch, M. (2018). On the Effects of Increased Vertical Mixing on the Arctic Ocean and Sea Ice. *Journal of Geophysical Research: Oceans*, 123:9266–9282. x

Liu, Y., Helfrich, S., Meier, W. N., and Dworak, R. (2020). Assessment of AMSR2 Ice Extent and Ice Edge in the Arctic using IMS. *Remote Sensing*, 12:1582. 46

Ludwig, V., Spreen, G., Haas, C., Istomina, L., Kauker, F., and Murashkin, D. (2019). Observation of the 2018 North Greenland polynya with a new merged

optical and passive microwave sea ice concentration dataset. *The Cryosphere Discussions*. 40

M, B.-K. (2019). Global Sea Ice Concentration Climate Data Record Validation Report. OSI-450 and OSI-430. *online*. 54

MacLaren, I. S. (1994). From Exploration to Publication: The Evolution of a 19th-Century Arctic Narrative. *Arctic*, 47:43–53. xi

Maykut, G. A. and Untersteiner, N. (1971). Some results from a time-dependent thermodynamic model of sea ice. *Journal of Geophysical Research*, 76:1550–1575. xiii

Meier, W., Fetterer, F., Windnagell, A., and Stewart, S. (2021). NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration User Guide. *NSIDC*, 4. 44

Meier, W. N., Fetterer, F., Stewart, J. S., and Helfrich, S. (2015). How do sea-ice concentrations from operational data compare with passive microwave estimates? Implications for improved model evaluations and forecasting. *Annals of Glaciology*, 56:332–340. 46

Meier, W. N. and Stewart, J. S. (2019). Assessing uncertainties in sea ice extent climate indicators. *Environmental Research Letters*, 14:035005. 40

Meredith, M. P. and Brandon, M. A. (2016). Oceanography and sea ice in the Southern Ocean. *Sea Ice*, pages 216–238. ix

Mizuta, R., Adachi, Y., and Kusunoki, S. (2008). Estimation of the future distribution of sea surface temperature and sea ice using the CMIP3 multi-model ensemble mean. 21

Mu, L., Nerger, L., Tang, Q., Loza, S. N., Sidorenko, D., Wang, Q., Semmler, T., Zampieri, L., Losch, M., and Goessling, H. F. (2020). Toward a Data Assimilation System for Seamless Sea Ice Prediction Based on the AWI Climate Model. *Journal of Advances in Modeling Earth Systems*, 12. xiv, 42

NIC, U. (2008). IMS Daily Northern Hemisphere Snow and Ice Analysis User Guide. *U.S. National Ice Center*. 41, 45

Niraula, B. and Goessling, H. F. (2021). Spatial damped anomaly persistence of the sea ice edge as a benchmark for dynamical forecast systems. *Journal of Geophysical Research: Oceans*, 126. xix, 1, 24, 25, 35

Notz, D. and Marotzke, J. (2012). Observations reveal external driver for Arctic sea-ice retreat. *Geophysical Research Letters*, 39:n/a–n/a. 42

NSIDC (2020). October — 2020 — Arctic Sea Ice News and Analysis. 16

NSIDC (2022). Charctic Interactive Sea Ice Graph — Arctic Sea Ice News and Analysis. xvi

Owens, R., Hewson, T., Hewson, T., and Hewson, T. (2018). ECMWF Forecast User Guide. 25, 26

Palerme, C., Müller, M., and Melsom, A. (2019). An Intercomparison of Verification Scores for Evaluating the Sea Ice Edge Position in Seasonal Forecasts. *Geophysical Research Letters*, 46:4757–4763. 4, 7, 13

Palmer, L. (2013). Melting Arctic ice will make way for more ships and more species invasions. *Nature*. xv

Parkinson, C. L. and Cavalieri, D. J. (2008). Arctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research*, 113. x

Perovich, D. K. and Polashenski, C. (2012). Albedo evolution of seasonal Arctic sea ice. *Geophysical Research Letters*, 39:n/a–n/a. x

Petrich, C. and Eicken, H. (2016). Overview of sea ice growth and properties. *Sea Ice*, pages 1–41. ix

Pinori, S., Crapolicchio, R., and Mecklenburg, S. (2008). Preparing the ESA-SMOS (Soil Moisture and Ocean Salinity) mission - Overview of the User Data Products and Data Distribution Strategy. xv

Pithan, F. and Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7:181–184. x

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., and Tschuck, P. (2004). Estimating the Decadal Predictability of a Coupled AOGCM. *Journal of Climate*, 17:4463–4472. 2, 3

Roach, L. A., Dörr, J., Holmes, C. R., Massonnet, F., Blockley, E. W., Notz, D., Rackow, T., Raphael, M. N., O'Farrell, S. P., Bailey, D. A., and Bitz, C. M. (2020). Antarctic sea ice area in CMIP6. *Geophysical Research Letters*, 47. 42

Rolph, R. J., Feltham, D. L., and Schröder, D. (2020). Changes of the Arctic marginal ice zone during the satellite era. *The Cryosphere*, 14:1971–1984. 46

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F. (2015). The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities. *Geoscientific Model Development*, 8:2991–3005. 25, 36

Rösel, A., Kaleschke, L., and Birnbaum, G. (2012). Melt ponds on Arctic sea ice determined from MODIS satellite data using an artificial neural network. *The Cryosphere*, 6:431–446. 55

Schulzweida, U. (2020). CDO User Guide. *Zenodo*. 27, 45

Sein, D. V., Koldunov, N. V., Danilov, S., Wang, Q., Sidorenko, D., Fast, I., Rackow, T., Cabos, W., and Jung, T. (2017). Ocean Modeling on a Mesh With Resolution Following the Local Rossby Radius. *Journal of Advances in Modeling Earth Systems*, 9:2601–2614. xv

Shupe, M., Rex, M., Dethloff, K., Damm, E., Fong, A., Gradinger, R., Heuzé, C., Loose, B., Makarov, A., Maslowski, W., et al. (2020). Arctic Report Card 2020: The MOSAiC Expedition: A Year Drifting with the Arctic Sea Ice. *Arctic Report Card*, 2020. xi

Smedsrud, L. H., Halvorsen, M. H., Stroeve, J. C., Zhang, R., and Kloster, K. (2017). Fram Strait sea ice export variability and september Arctic sea ice extent over the last 80 years. *The Cryosphere*, 11:65–79. 47

Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger, J., Skachko, S., Liu, Y., Dupont, F., Lemieux, J., Beaudoin, C., Tranchant, B., Drévillon, M., Garric, G., Testut, C., Lellouche, J., Pellerin, P., Ritchie, H., Lu, Y., Davidson, F., Buehner, M., Caya, A., and Lajoie, M. (2015). Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 142:659–671. 2, 24

Spreen, G., Kaleschke, L., and Heygster, G. (2008). Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research*, 113. 25, 27, 41, 44, 54

Spreen, G. and Kern, S. (2016). Methods of satellite remote sensing of sea ice. *Sea Ice*, 9:239–260. xii

Stephenson, S. R., Smith, L. C., and Agnew, J. A. (2011). Divergent long-term trajectories of human access to the Arctic. *Nature Climate Change*, 1:156–160. 2

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. (2007). Arctic sea ice decline: Faster than forecast. *Geophysical Research Letters*, 34. 40

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

Tietsche, S., Balmaseda, M. A., Zuo, H., and Mogensen, K. (2015). Arctic sea ice in the global eddy-permitting ocean reanalysis ORAP5. *Climate Dynamics*, 49:775–789. 25, 36

Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L., and Kern, S. (2016). The EUMETSAT sea ice concentration climate data record. *The Cryosphere*, 10:2275–2290. 55

Usher, P. (1971). The Canadian Western Arctic: A Century of Change. *Anthropologica*, 13:169. xi

Van den Dool, H. M., Peng, P., Johansson, A., Chelliah, M., Shabbar, A., and Saha, S. (2006). Seasonal-to-Decadal Predictability and Prediction of North American Climate—The Atlantic Influence. *Journal of Climate*, 19:6005–6024. 3

Veland, S. and Lynch, A. H. (2016). Arctic ice edge narratives: scale, discourse and ontological security. *Area*, 49:9–17. xvi

Verhoeven, G. J. J. (2009). Providing an archaeological bird's-eye view - an overall picture of ground-based means to execute low-altitude aerial photography (LAAP) in Archaeology. *Archaeological Prospection*, 16:233–249. xii

Vihma, T. (2014). Effects of Arctic Sea Ice Decline on Weather and Climate: A Review. *Surveys in Geophysics*, 35:1175–1214. 40

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.-J., Xiao, H., Zaripov, R., and Zhang, L. (2017). The Subseasonal to Seasonal (S2S) Prediction Project Database. *Bulletin of the American Meteorological Society*, 98:163–173. xiii, xiv, 6, 18, 24, 25, 26

Wang, L., Scott, K. A., and Clausi, D. A. (2016). Improved Sea Ice Concentration Estimation Through Fusing Classified SAR Imagery and AMSR-E Data. *Canadian Journal of Remote Sensing*, 42:41–52. 43

Wang, M. and Overland, J. E. (2009). A sea ice free summer Arctic within 30 years? *Geophysical Research Letters*, 36:n/a–n/a. xi

Wayand, N. E., Bitz, C. M., and Blanchard-Wrigglesworth, E. (2019). A Year-Round Subseasonal-to-Seasonal Sea Ice Prediction Portal. *Geophysical Research Letters*, 46:3298–3307. 3

Wilks, D. S. (2019). Forecast Verification. *Statistical Methods in the Atmospheric Sciences*, pages 369–483. 2

Woert, M. L. V., Zou, C.-Z., Meier, W. N., Hovey, P. D., Preller, R. H., and Posey, P. G. (2004). Forecast Verification of the Polar Ice Prediction System (PIPS) Sea Ice Concentration Fields. *Journal of Atmospheric and Oceanic Technology*, 21:944–957. 2, 3

Yang, J., Ji, W., and Zeng, Y. (2019). Transition from eyeball to snowball driven by sea-ice drift on tidally locked terrestrial planets. *Nature Astronomy*, 4:58–66. x

Zampieri, L., Goessling, H. F., and Jung, T. (2018). Bright Prospects for Arctic Sea Ice Prediction on Subseasonal Time Scales. *Geophysical Research Letters*, 45:9731–9738. 2, 4, 6, 12, 18, 19, 24, 35, 36, 40

Zampieri, L., Goessling, H. F., and Jung, T. (2019). Predictability of Antarctic Sea Ice Edge on Subseasonal Time Scales. *Geophysical Research Letters*, 46:9719–9727. 2, 4, 6, 12, 18, 19, 21, 24, 35, 36, 40, 46

Zhao, J., Cheng, B., Vihma, T., Heil, P., Hui, F., Shu, Q., Zhang, L., and Yang, Q. (2020). Fast Ice Prediction System (FIPS) for land-fast sea ice at Prydz Bay, East Antarctica: an operational service for CHINARE. *Annals of Glaciology*, 61:271–283. 42

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. *Ocean Science*, 15:779–808. xiii, xiv, 24, 25, 27, 42, 45

Årthun, M., Onarheim, I. H., Dörr, J., and Eldevik, T. (2021). The Seasonal and Regional Transition to an Ice-Free Arctic. *Geophysical Research Letters*, 48. xi

A

Supplementary Information for Chapter 1

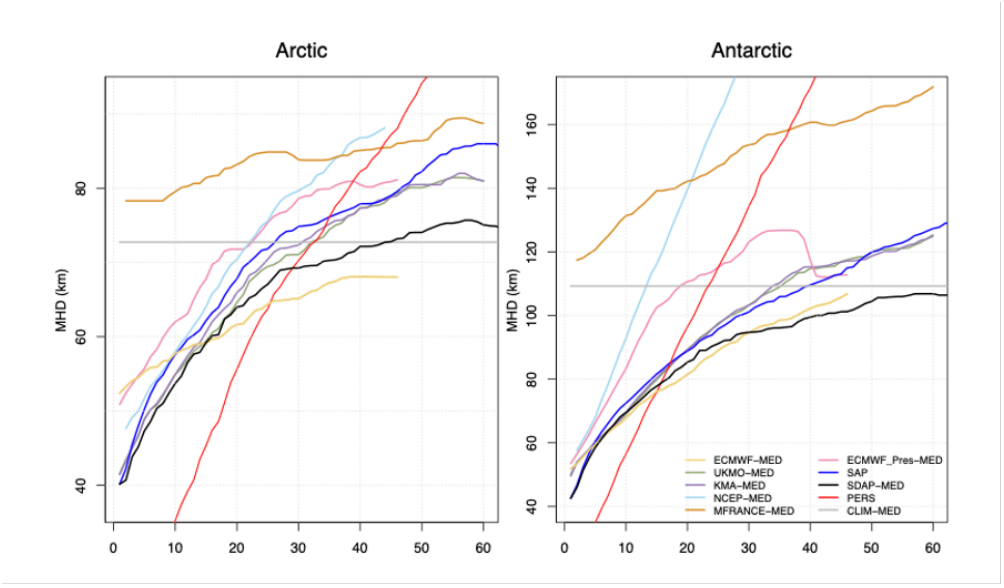

FIGURE A.1: Same as figure 1.8, but for Modified Hausdorff Distance (MHD), using the 50% probability threshold to create the binary forecasts. For the probabilistic forecasts, the median probability of ice presence is used to determine the predicted ice edge. Due to the resolution of the S2S grid and the disconnected segments of the ice-edge, the MHD measurements do not precisely mirror the results of the SPS measurements. Nevertheless, the ranking of the models is similar to that found using SPS and we can confirm that the SDAP forecasts perform better than all models in the S2S database except ECMWF.

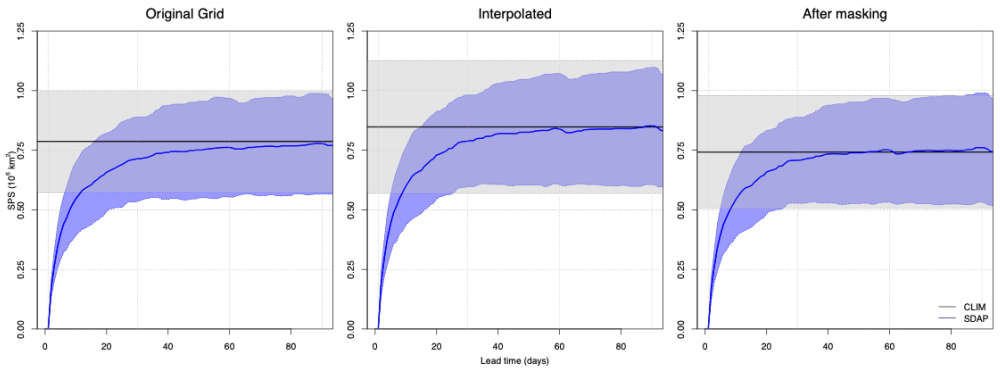

FIGURE A.2: A comparison between the SPS of climatological forecast (CLIM) and the Spatial Damped Anomaly Persistence forecast (SDAP) in the original grid, after interpolation to the common S2S grid and after applying a common mask. The mask is derived by finding all grid-cells, where each of the forecast systems being compared has either ice or ocean. Results show the average of all forecasts for years 1999 to 2010, while the shaded areas show +/- 1 standard deviation.

B

Supplementary Information for Chapter 2

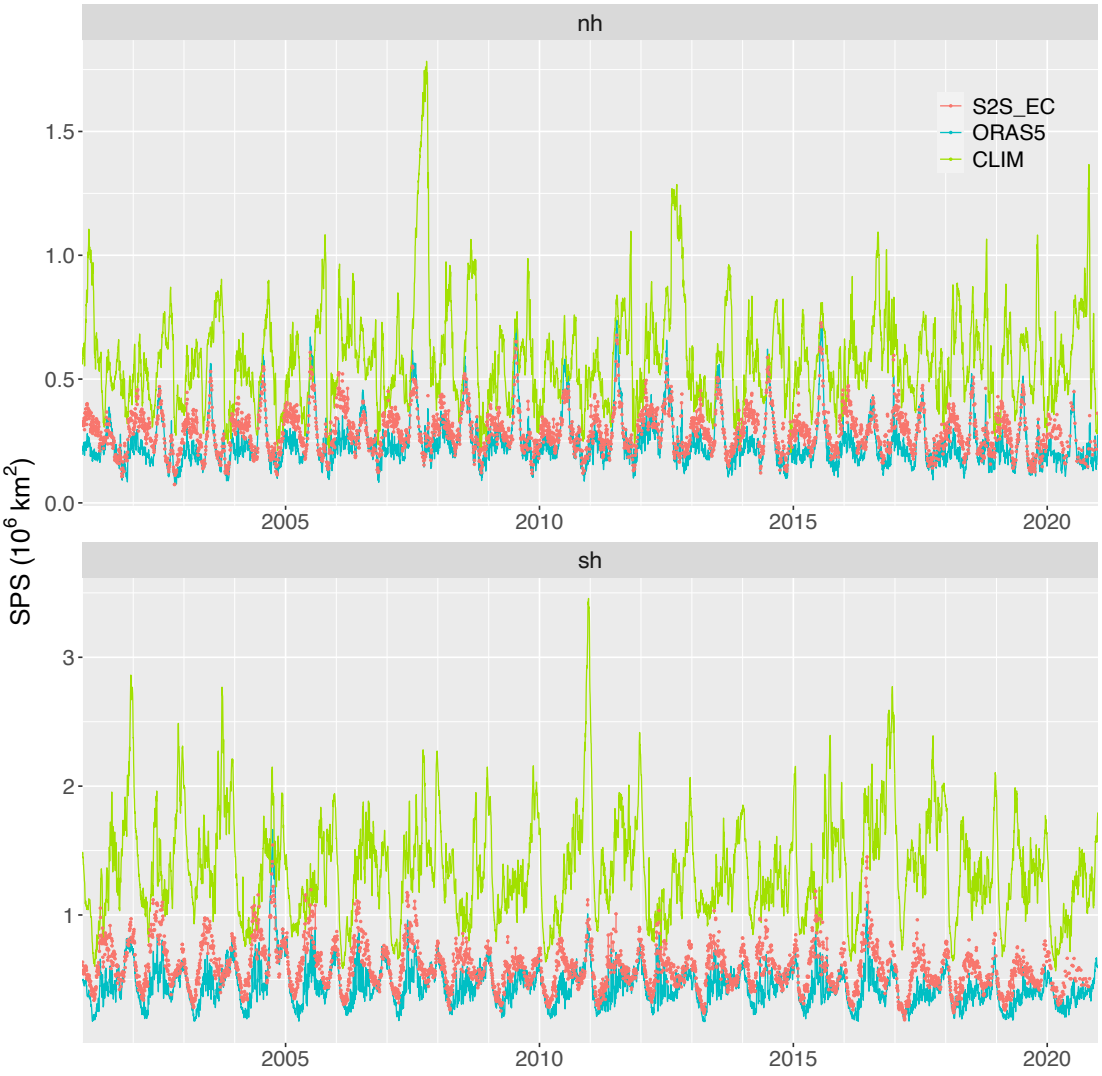

FIGURE B.1: Daily timeseries of Spatial Probability Score (in each hemisphere) of S2S_EC, ORAS5 and CLIM measured against the observed ice edge from OSISAF.

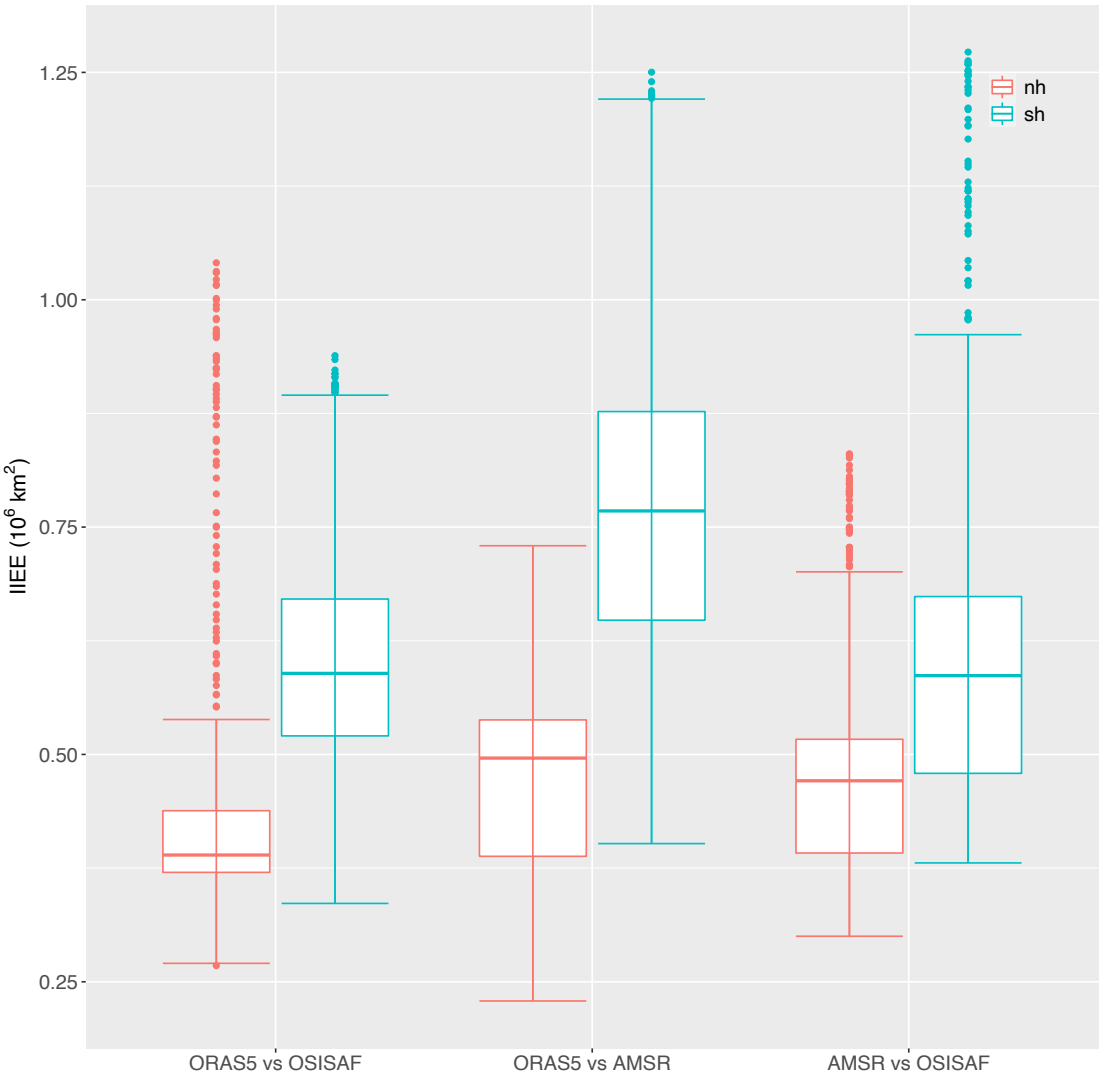

FIGURE B.2: Boxplot of Integrated Ice Edge Error measured pairwise between OSISAF, AMSR and ORAS5 (ensemble median) in each hemisphere.

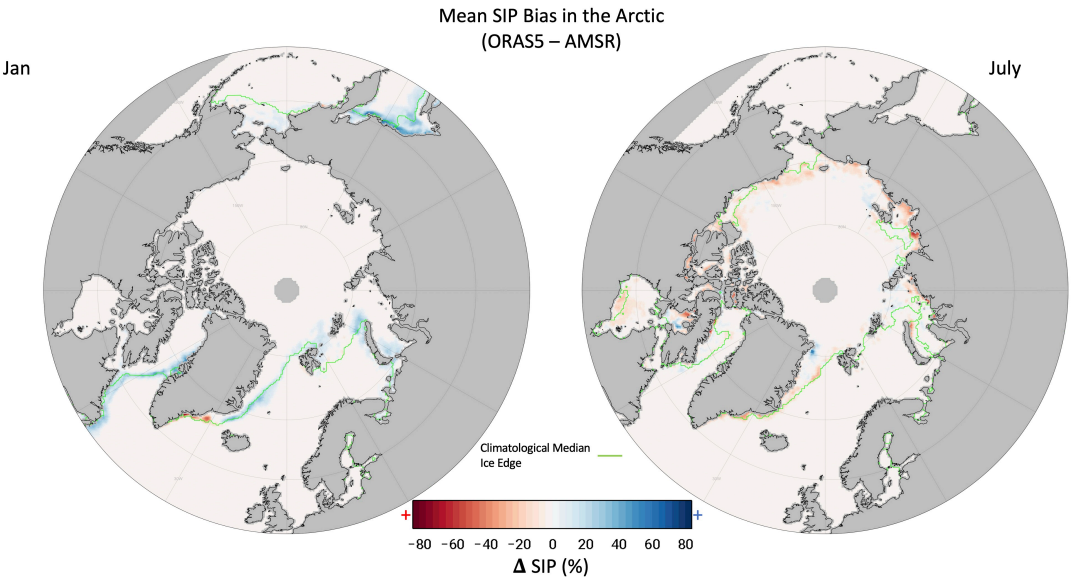

FIGURE B.3: Same as Figure 2.4, but for the bias against observations from AMSR. The mask has been changed to match the coverage of the AMSR dataset.

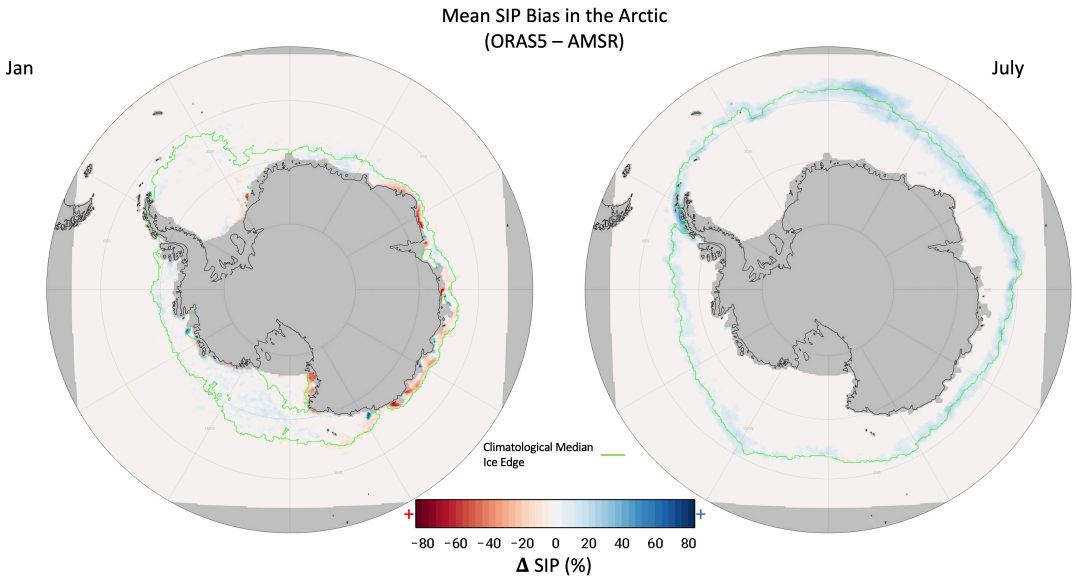

FIGURE B.4: Same as Figure 2.5, but for the bias against observations from AMSR. The mask has been changed to match the coverage of the AMSR dataset.

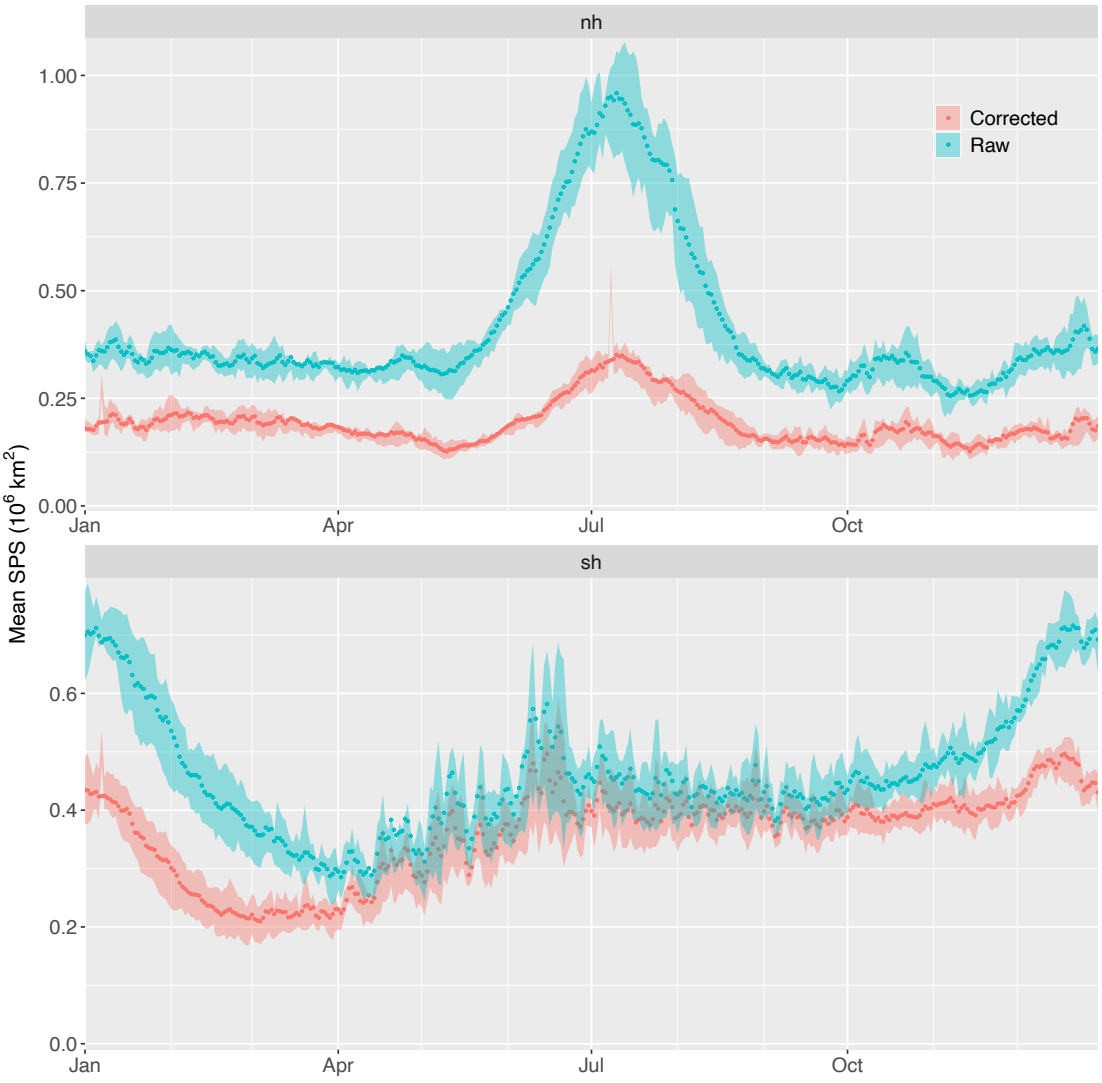

FIGURE B.5: Reduction in mean SPS of ORAS5 analysis (against observations from OSISAF) due to simple bias correction. The upper and lower quartile of the mean are shown by the shaded region.

C

Supplementary Information for Chapter 3

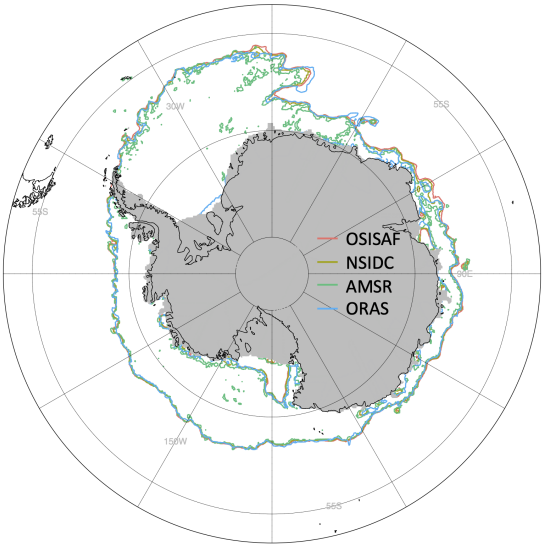

FIGURE C.1: Ice Edge contours in the Antarctic from the different datasets used in this study for 15th of December, 2015.

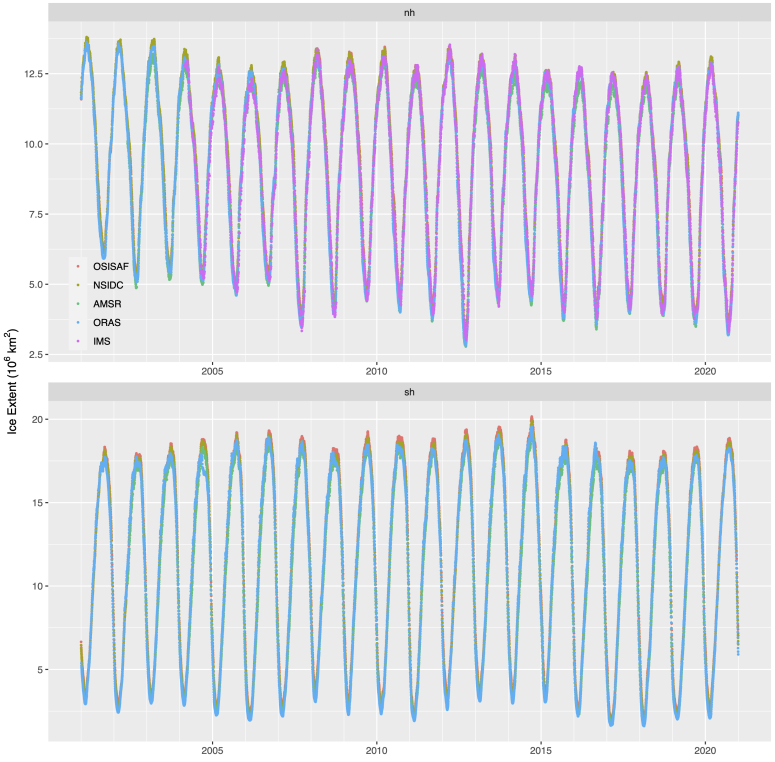

FIGURE C.2: Daily ice extent estimates for each hemisphere from the different datasets used in this study, all (except IMS) computed using the 15% SIC threshold.

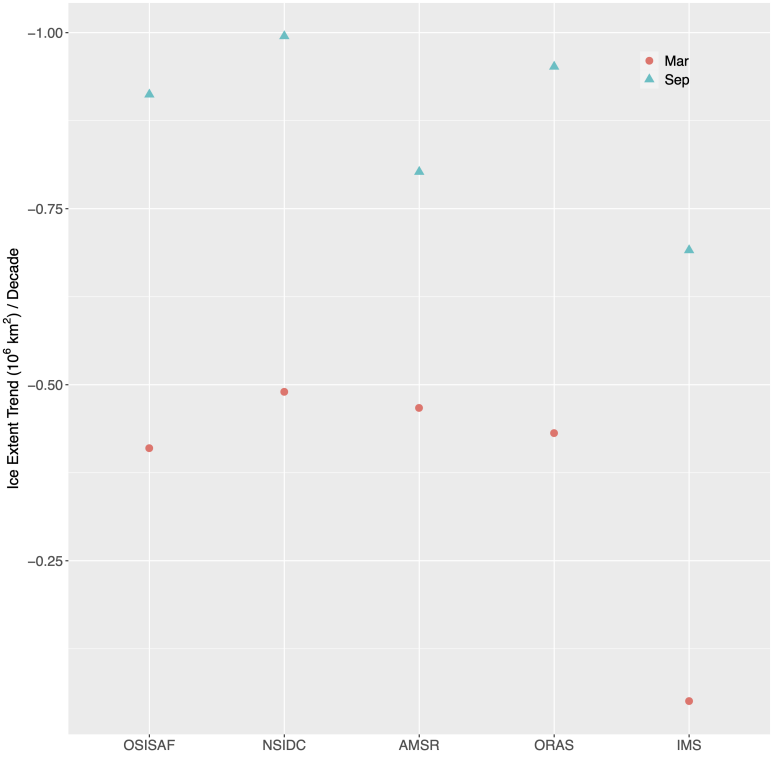

FIGURE C.3: Decadal trend of Arctic ice extent decrease (in million square kms per decade) for the months of September and March, according to each dataset. Antarctic trends were not found to be statistically significant.

FIGURE C.4: Seasonality of IIEE against IMS using 15% and 40% concentration threshold to determine ice presence in the 3 SIC datasets.

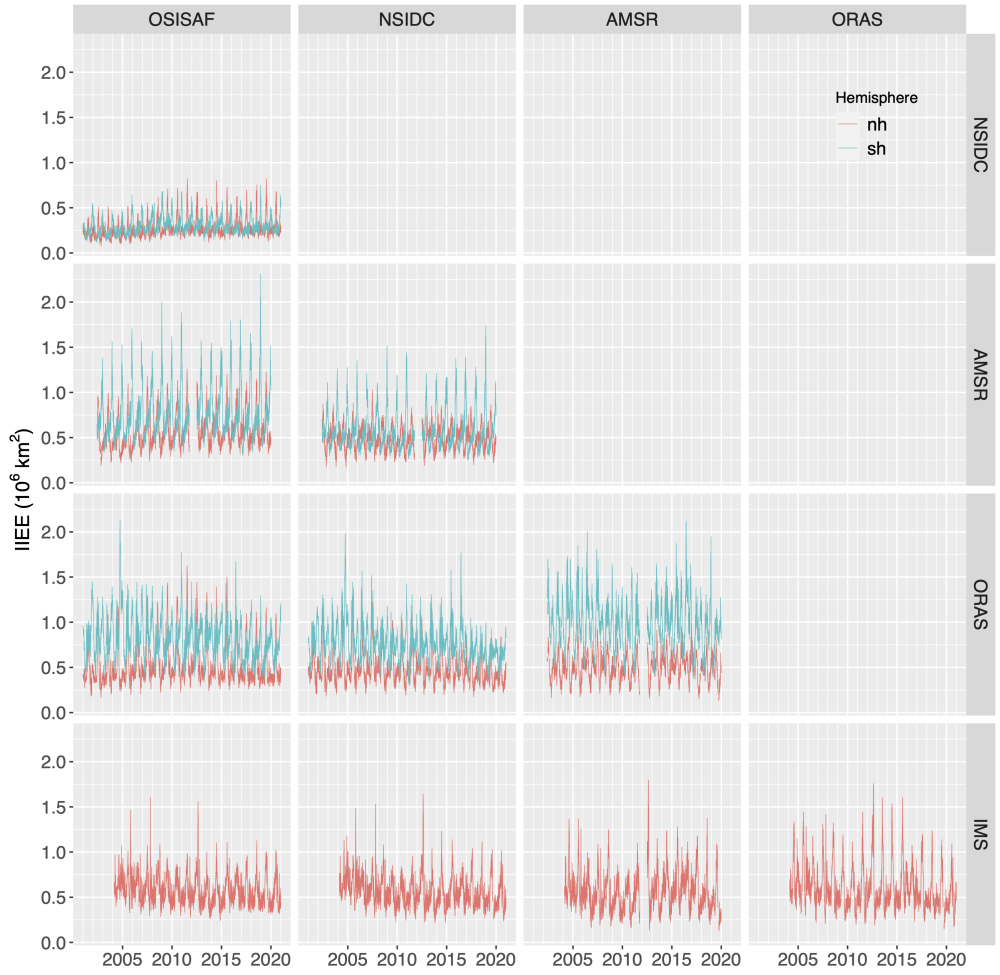

FIGURE C.5: Daily timeseries of Integrated Ice Edge Error computed pairwise between all datasets in both hemispheres. The pairs are marked by the row and column heads.

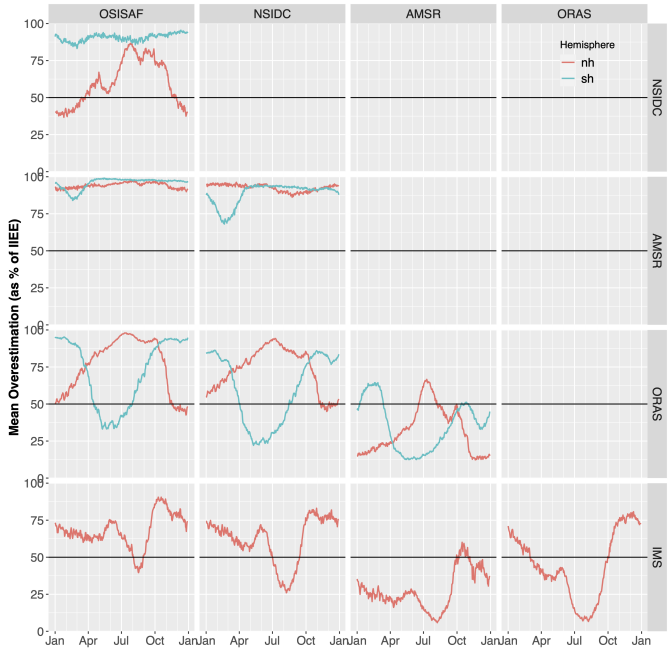

FIGURE C.6: Mean seasonality of Overestimation (by dataset on column head, over dataset on row hea), as % of IIEE. Overestimation from one can be considered as Underestimation from the other one. The 50 % line shows a well balanced relation between the two datasets, such that they are both equally over (or under) estimating compared to the other.

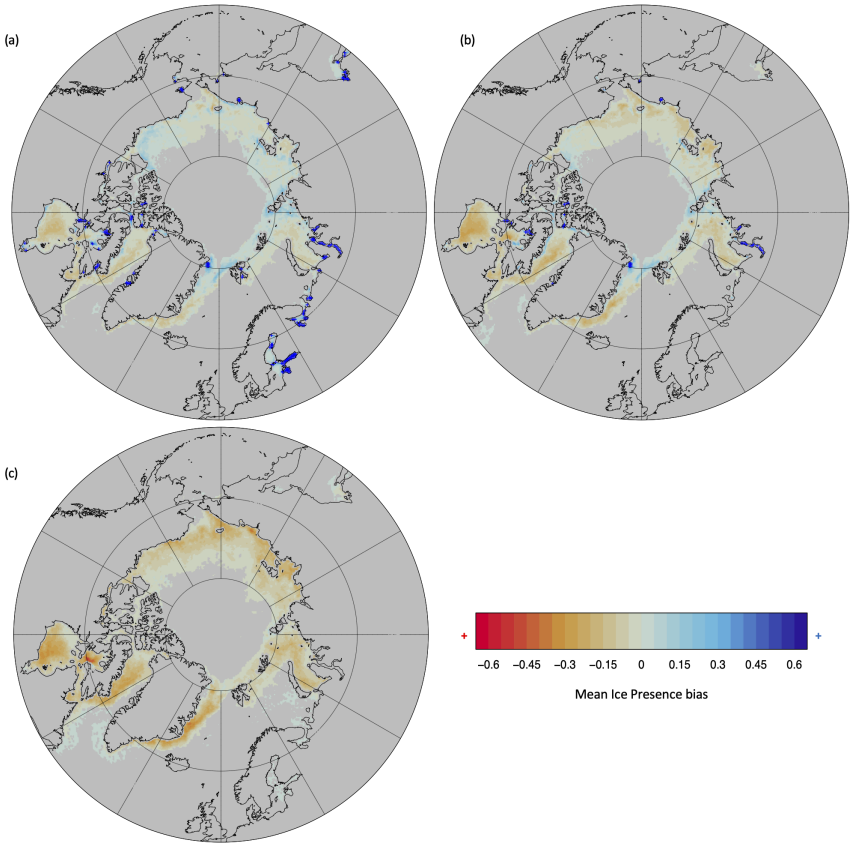

FIGURE C.7: Same as Fig.3.4 (showing mean bias in the Arctic during July), but for (a) OSISAF minus IMS, (b) NSIDC minus IMS and (c) AMSR minus IMS.

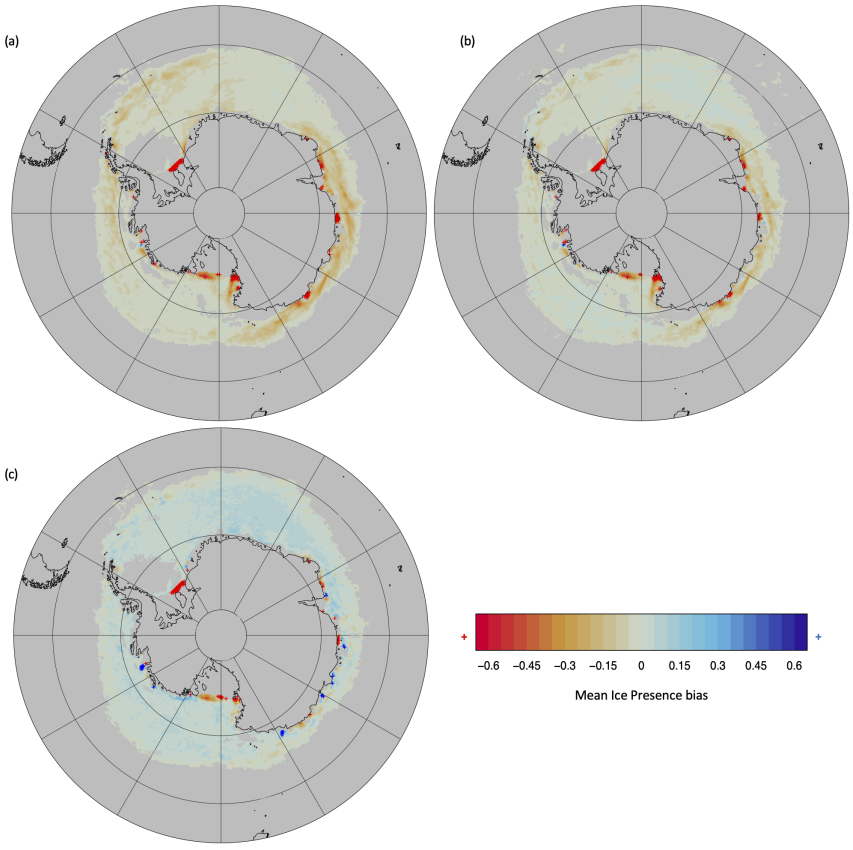

FIGURE C.8: Same as Fig.3.5 (showing mean bias in the Antarctic during December), but for (a) ORAS minus OSISAF, (b) ORAS minus NSIDC and (c) ORAS minus AMSR.