Applying Statistical Analysis and Machine Learning to Improve the Ice Sensing Algorithm

by

Lieutenant Junior Grade Lucas A. Herron,

United States Navy

Submitted to the Joint Program in Physical Oceanography
Massachusetts Institute of Technology &
Woods Hole Oceanographic Institution
on August 11, 2023, in partial fulfillment of the requirements for the degree of
Master of Science in Physical Oceanography

Abstract

The detection of sea ice is a major problem faced by Argo floats operating in polar regions. In these areas, the presence of sea ice threatens to damage or destroy floats in the event of an impact at the surface. While methods have been proposed and implemented to combat this danger, the most successful of which is the Ice Sensing Algorithm (ISA), further work is necessary to fully mitigate the risks, particularly in the Arctic. In this analysis, past CTD profiles from the Arctic are compiled and matched with sea ice data to examine the performance of the ISA and recommend potential changes and new methods to further improve its accuracy. This is accomplished by fitting the data to statistical and machine learning models to predict the presence of ice and analyzing the results. Results show that both modifications to current methods and the inclusion of new variables may increase the predictive power of the ISA. Specifically, the analysis shows that the use of point measurements (as opposed to a metric over a pressure range) at the shallowest allowable depth provides the best performance. The additional inclusion of practical salinity and time of year as predictive variables also increases the performance of the algorithm. Results and statistics on the performance of the algorithm are provided and analyzed in various regions.

Thesis Supervisor: Steven R. Jayne
Title: Senior Scientist, Woods Hole Oceanographic Institution
Acknowledgments

I would like to express my gratitude to all those who have supported and guided me as I complete my graduate thesis. Their encouragement, insight, and support have been incredibly helpful and I cannot thank them enough. While graduate school has seemed like a long journey, I am thankful for all those who helped get me there in the first place. Those who sparked my love for science, those who opened new doors, those who pushed me to do my best, and those who supported me along the way. While I cannot name everyone that deserves to be in these categories, I would like to recognize some of the most important.

Firstly, Steve Jayne. While I didn’t know you would become my advisor when we first met during TROPIC before my senior year at USNA, I am so thankful that you took me in. You allowed me take classes I found engaging and explore research that suited my interests. Your hands-off approach allowed me to explore my own interests and learn so much during my time in graduate school.

To my parents. I would like to thank my father for sparking my interest in science. From a very young age, you have always encouraged me to learn about how things work. Whether that was taking things apart or building them for myself, you fostered my interest in asking questions and finding answers and it has led me to where I am today. To my mother, I thank you for always encouraging me to do my best. You’ve opened so many doors for me and pushed me to take chances. Without you, I wouldn’t be where I am today.

To my wife Alex, I thank you for your support through everything. Not only during my time at graduate school but during my time at USNA as well. You’ve supported me and encouraged me to do my best and take on new opportunities even when it requires sacrifice on your part. You are my backbone and I couldn’t do it without you.

I would especially like to thank CAPT Elizabeth Sanabia (ret.). Despite not being an oceanography major at USNA, when I was looking to do research you took me under you wing. Research helped me find my niche at USNA, developed my budding love for programming, and opened my mind to the field of oceanography. You introduced me to the TROPIC program where I was able to apply so many of the things I learned. And during all of it you showed me what it means to means to be a selfless leader. You have opened so many doors for me and helped me to find my calling and for that I am beyond grateful.

Lastly, I would like to thank the Office of Naval Research under grant numbers N00014-18-1-2813, N00014-20-1-2765, and N00014-18-1-2780, and the Civilians Institution Office at the Naval Postgraduate School, who funded and supported my enrollment in the MIT-WHOI Joint Program.
## Contents

1 Introduction ................................................................. 15  
  1.1 Motivation ................................................................. 15  
  1.2 Goal and Purpose ....................................................... 15  

2 Background ................................................................. 17  
  2.1 The Argo Program ....................................................... 17  
  2.2 Previous Work ........................................................... 19  
  2.3 Features and Challenges of the Arctic Ocean .................. 22  

3 Data ........................................................................... 25  
  3.1 Data Overview ........................................................... 25  
  3.2 MAISE-NH Data .......................................................... 25  
  3.3 Argo Data ................................................................. 26  
  3.4 Ice Tethered Profiler Data ............................................. 27  
  3.5 Data Assimilation ....................................................... 28  
  3.6 Representative Sampling ............................................. 30  

4 Statistical and Machine Learning Methods ......................... 35  
  4.1 Overview of Methods Used ............................................ 35  
  4.2 Regression and Neural Network Techniques ....................... 36  
  4.3 Binary Classification Statistics ..................................... 36  

5 Modified Temperature Threshold Method .......................... 39  
  5.1 Overview ................................................................. 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>Best Depth Determination</td>
<td>40</td>
</tr>
<tr>
<td>5.3</td>
<td>Optimal Measurement Strategy</td>
<td>41</td>
</tr>
<tr>
<td>5.4</td>
<td>Results</td>
<td>45</td>
</tr>
<tr>
<td>5.5</td>
<td>Conversion to Temperature Threshold</td>
<td>47</td>
</tr>
<tr>
<td>5.6</td>
<td>Regional Performance and Local Tuning</td>
<td>48</td>
</tr>
<tr>
<td>5.7</td>
<td>Summary</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>Multiple Logistic Regression Method</td>
<td>59</td>
</tr>
<tr>
<td>6.1</td>
<td>Overview</td>
<td>59</td>
</tr>
<tr>
<td>6.2</td>
<td>Best Predictive Variable Combination</td>
<td>59</td>
</tr>
<tr>
<td>6.3</td>
<td>Results</td>
<td>61</td>
</tr>
<tr>
<td>6.4</td>
<td>Regional Performance and Local Tuning</td>
<td>63</td>
</tr>
<tr>
<td>6.5</td>
<td>Summary</td>
<td>65</td>
</tr>
<tr>
<td>7</td>
<td>Neural Network Method</td>
<td>71</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview</td>
<td>71</td>
</tr>
<tr>
<td>7.2</td>
<td>Initial Testing</td>
<td>72</td>
</tr>
<tr>
<td>7.3</td>
<td>LSTM Network</td>
<td>72</td>
</tr>
<tr>
<td>7.4</td>
<td>Results</td>
<td>73</td>
</tr>
<tr>
<td>8</td>
<td>Summary, Conclusions, and Future Work</td>
<td>77</td>
</tr>
<tr>
<td>8.1</td>
<td>Summary and Conclusions</td>
<td>77</td>
</tr>
<tr>
<td>8.2</td>
<td>Future Work</td>
<td>79</td>
</tr>
</tbody>
</table>
List of Figures

2-1 Graphic detailing a typical Argo cycle that repeats in a 10 day pattern. Measurements are recorded during surfacing on each cycle. 18

2-2 Graphic detailing the current Ice Sensing Algorithm used to predict and avoid ice. Current methods use a temperature threshold compared against a mean or median temperature measured over a pressure range. 20

3-1 Location and ice presence of Argo profiles used in this analysis. 26

3-2 Location and ice presence of Ice Tethered Profiler (ITP) profiles used in this analysis. 28

3-3 The regional bounds used for this analysis. Regional bounds are used to roughly approximate expected deployment and operation areas for Argo floats. These regions are used to analyze and locally tune the ISA in subsequent chapters. 29

3-4 Spatial histogram of profiles prior to spatial normalization. The high concentration of profiles in Arctic and Beaufort are likely to introduce a regional bias when fitting models and must be accounted for. 31

3-5 Spatial histogram of profiles after spatial normalization. The more uniform distribution will help prevent a strong regional bias from emerging. 31

5-1 Misclassification rate, false negative rate, and false positive rate for models trained using temperature from different depths. The models show the best performance using measurements from the shallowest depth. 41
A sample temperature profile from the Beaufort region. The plot highlights the potential problem of lagged statistics caused by measuring a mean or median over a pressure range as they are likely to report values that are not representative of the surface conditions.

Misclassification rate of point measurements in comparison to the mean, median, and minimum taken over a 25dbar range.

Misclassification rate of point measurements in comparison to the mean, median, and minimum taken over a 10dbar range.

Performance of the model when trained and tested on the full dataset. Note the misclassification rate minimum does not occur at probability 0.5. The ‘optimal’ value is reported as the threshold that minimizes the misclassification rate, rather than at probability 0.5.

Performance of the model when trained and tested on the full dataset. The x-axis is converted from probability to temperature according to equation 5.5 so the performance of the model can be viewed in terms of current ISA methods that use a temperature threshold.

Performance statistics for the Beaufort Sea region over various probability and temperature thresholds after local tuning.

Performance statistics for the Arctic region over various probability and temperature thresholds after local tuning.

Performance statistics for the Norwegian Sea region over various probability and temperature thresholds after local tuning.

Performance statistics for the Greenland Sea region over various probability and temperature thresholds after local tuning.

Performance statistics for the Baffin Bay region over various probability and temperature thresholds after local tuning.

Contribution of the combined $y_{day\sin}/y_{day\cos}$ predictive variables as a function of the day of the year ($y_{day}$) calculated as: $-0.331(y_{day\sin}) + 0.207(y_{day\cos})$.
Performance of the multiple logistic regression model when trained and tested on the full dataset. ......................... 63
Performance statistics for the Beaufort Sea region over the full range of probability thresholds after local tuning. ..................... 67
Performance statistics for the Arctic region over the full range of probability thresholds after local tuning. ..................... 67
Performance statistics for the Norwegian Sea region over the full range of probability thresholds after local tuning. ..................... 68
Performance statistics for the Greenland Sea region over the full range of probability thresholds after local tuning. ..................... 68
Performance statistics for the Baffin Bay region over the full range of probability thresholds after local tuning. ..................... 69

Predicted probabilities by the RNN model with increasing number of measurements. The model is able to correctly predict the presence of ice once a full profile is input, but further work is required to physically decipher model output from partial profiles. ..................... 74
List of Tables

3.1 Comparison of the proportion of under-ice profiles in the dataset to the year-round average ice presence in each region prior to oversampling. Oversampling corrects this by duplicating samples of the underrepresented class to match the year-round average ice presence in each region. 32

3.2 Change in the number of Argo and ITP profiles during the different steps of data preparation. 33

3.3 Change in the number of ice-free and under-ice profiles during the different steps of data preparation. 33

3.4 Change in the number of Argo and ITP profiles by region during the different steps of data preparation. 33

3.5 Change in the number of Ice-free and Under-ice profiles by region during the different steps of data preparation. 33

4.1 Table of various binary classification metrics used in this analysis including their calculation and interpretation as conditional probabilities. TP represents true positives, TN represents true negatives, FP represents false postives, and FN represents false negatives. 37

5.1 Performance of the logistic regression model at the misclassification rate minimum for each region. 52

5.2 Equations to generate a probability of ice presence as determined by the logistic regression model for each region. 52
5.3 Tabulation of the required probability threshold to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region. 53

5.4 Tabulation of the required temperature threshold (in °C) to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region. 53

6.1 Magnitude of the coefficients when trained using predictive variables that are normalized in the range 0 to 1. The magnitudes can be used to roughly compare the predictive power of variables against each other. 64

6.2 Performance of the multiple logistic regression model at the misclassification rate minimum for each region. 66

6.3 Equations to generate a probability of ice presence as determined by the multiple logistic regression model for each region. 66

6.4 Tabulation of the required probability threshold to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region. 66

7.1 Performance of the recurrent neural network model at the misclassification rate minimum for each region. 75
Chapter 1

Introduction

1.1 Motivation

The detection of ice is a major problem faced by profiling floats operating in polar regions. In these areas, the presence of ice disrupts the regular surfacing operation of these floats by preventing them from reaching the surface. Not only does it prevent the transmission of data and localization by GPS, it also threatens to damage or destroy sensors in the event of an impact. Methods have been proposed and implemented to combat this issue, however further work is still necessary to fully mitigate the risks.

This issue poses a significant problem to the Argo program as it maintains around 4000 autonomous profiling floats across the global oceans [1]. Intentional deployments of Argo floats into areas of sea ice have also been made, and the program seeks to further expand its coverage in areas of seasonal sea ice. This expansion of the program necessitates the development of methods to increase the historically low survivability of profiling floats operating in polar environments.

1.2 Goal and Purpose

The goal and purpose of this analysis is to investigate the performance of the Ice Sensing Algorithm (ISA) based on historical CTD profiles in the Arctic and recommend potential changes and new methods to improve the ISA’s performance. This
includes both recommendations that modify existing strategies and new methods of ice prediction. The goal in introducing new methods of ice prediction is to fit the historical data to statistical and machine learning models to find the optimal policies when predicting the presence of ice above a surfacing float. In this way, the new methods introduced will have a statistical basis and the use of historical data will allow an analysis of each method’s performance.

In this paper, new algorithms and modifications to current methods are proposed that can be implemented on profiling floats that allow for the prediction and avoidance of sea ice using the CTD data available on all Argo floats. Three methods of increasing complexity are presented, and their performance analyzed using historical data in different ocean environments with a specific focus of the Arctic and the spatial and temporal challenges the region poses.
Chapter 2

Background

2.1 The Argo Program

Argo is an international program that aims to collect data from the upper 2000 meters of the world’s ocean [2]. The program utilizes profiling floats that drift with the ocean currents and surface at regular intervals measuring water parameters along vertical profiles. Although there are a number of different models used by the Argo program, each contains a conductivity, temperature and depth sensor (CTD) that measures these parameters during each surfacing cycle. Typical cycles repeat in a 10-day pattern beginning with the float drifting for 10 days at 1000m, then descending to 2000m before ascending to the surface and taking measurements during surfacing. Once surfaced, floats transmit the data recorded during the ascent phase. Floats typically remain on the surface between 20 and 30 minutes during this time [3]. Once data transmission is complete, the float descends to 1000m repeating the pattern for the entirety of its life, typically lasting between four and five years and limited by the capacity of the on-board battery [4].

Although the Argo program is primarily designed to operate in the open oceans, deployments have occurred in areas of seasonal sea ice on numerous occasions. In these environments, profiling floats have a distinct advantage over manned missions as they are able to autonomously and continuously collect data from beneath the surface without the danger or cost of operating in polar environments. In fact, profiling floats
are able to gather CTD profiles year-round while manned missions have historically taken far fewer profiles in the winter compared to the summer [5]. This is a distinct advantage of Argo and calls have been made to further expand the program’s coverage into areas of seasonal sea ice [1].

The presence of ice poses a number of problems however. Namely, it prevents the regular surfacing operation of floats. Without any preventative measures, floats are likely to damage or destroy their antenna or sensors in the event of an impact with ice at the surface. Even if they are able to surface into an ice-free area, they could be crushed between ice floes or become stuck in newly forming ice [6]. Early deployments of Argo floats into areas of seasonal sea ice demonstrate these dangers. Of 35 floats that encountered sea ice in the Southern Ocean between 1999 and 2006 only 13 survived their first season, a survival rate of approximately 37% [7]. This issue necessitated the development of methods to predict and avoid ice to increase the low survivability of Argo floats deployed in polar regions.
2.2 Previous Work

A few solutions have been proposed to combat this problem. These include antenna hardening [8], upward looking sonar [7, 6], and the introduction of an ice sensing algorithm [7]. The first two of these proposals have problems that make their exclusive adoption difficult.

Antenna hardening, although relatively easy to implement, does not fix the issue of being crushed between ice floes or frozen into the ice if a float successfully reaches the surface. This method was first used by researchers at Woods Hole Oceanographic Institution where floats with a hardened antenna would repeatedly attempt to surface if they were initially unsuccessful, essentially probing until an open area can be found [8]. While it may be necessary in the high-Arctic where surfacing is difficult year-round, the continual probing results in high power usage by a float’s pump and is not an ideal solution on its own. Instead, these hardened antennas are typically used as a supplementary safety feature in addition to other methods of ice detection.

Upward looking sonars have also been trialed with nominal success. First employed by the Novel Argo Observing System (NAOS) program, a new Argo float model was developed for use in polar regions. These floats contained an upward looking sonar that was used to detect and predict the thickness of ice above a surfacing Argo float. Results showed high variability in predicted ice thickness between subsequent measurements making its use as an ice avoidance method difficult due to the high risk of false detection. The report concluded that the upward looking sonar may be useful to measure ice thickness when combined with a more accurate depth gauge, but would not be appropriate as an ice avoidance method alone [6].

Finally, the introduction of an ice sensing algorithm (ISA) has proven to be the most viable option for ice detection and avoidance. This method, originally introduced by Klatt et al. [7], uses a float’s temperature sensor to determine the median temperature between two pressure levels during surfacing. It compares this value to a pre-programmed threshold temperature. If the median is below the threshold, it is assumed that ice exists at the surface and the ascent is aborted. For their study,
they used the median value measured between 50m and 20m with a threshold value of -1.79°C which was specifically tuned for their targeted deployment area in the Weddell Sea. The results were positive with a survival rate of around 80%, a marked improvement compared to the 37% survival rate of floats without the algorithm in the same area [7]. Today, the ISA is the predominate ice detection and avoidance method used by Argo floats. The algorithm requires no hardware modification and is simple enough that it can be implemented on a float with negligible power consumption. Current ISA strategies operate on the same principle proposed by Klatt et al. [7] with additional failsafe features including hardened antennas, a pressure timeout and a communication timeout. The latter two exist to abort an ascent if the ISA fails and the float stops ascending because it hits ice and cannot reach the surface.

Figure 2-2: Graphic detailing the current Ice Sensing Algorithm used to predict and avoid ice. Current methods use a temperature threshold compared against a mean or median temperature measured over a pressure range.

While the algorithm has been shown to work well in many deployments, it is not deterministic, and its parameters must be adjusted based on the expected operating
area. As described above, the parameters used by Klatt et al. were selectively tuned for the Weddell Sea. These parameters work well for this area because the Weddell Sea is well mixed during the winter months with mixed layer depths in excess of 100 m [9]. This provides the algorithm with a clear indication of the surface conditions within the 50m-20m range in which the ISA makes its determination. This pressure range and temperature threshold are spatially dependent however, as there are significant regional differences that affect the formation of sea ice including local freezing temperature, stratification profiles, and local mixing processes. For this reason, the ISA must be locally tuned for each deployment based on region. A recent report by the Euro-Argo Research Infrastructure Sustainability and Enhancement Project (EA RISE) collected locally tuned parameters for other areas recommended by various papers. These include the Weddell Gyre, Wilkes Land, Baffin Bay, the Barents Sea, and Nansen Basin. The values of these reports recommend measuring ranges between 10m and 50m with thresholds between -1.79°C and -1.0°C [10]. Still, no threshold values have been recommended in the areas of the Beaufort Sea and East Greenland Current. Both areas are regions of seasonal sea ice that have been targeted by past Argo deployments and would benefit from a recommendation of pressure range and threshold for the ISA.

The ISA would also benefit from some measure of confidence in its prediction of ice. Currently, the use of a temperature threshold value is arbitrary and provides no indication of the likelihood of ice above a surfacing float. If coupled with historical data of known under-ice and ice-free profiles, a probability function could feasibly be introduced to predict the likelihood of ice presence. This would provide Argo floats with a probability of ice presence for the conditions of a given profile and would allow researchers to select a probability threshold based on the risk tolerance of their specific mission requirements.
2.3 Features and Challenges of the Arctic Ocean

The Arctic Ocean poses a number of challenges for the development of the ISA. Unlike the well mixed Southern Ocean that the algorithm was originally developed for, the Arctic is generally well stratified with shallow mixed layer depths and contains separate basins that interact with the different waters of the Atlantic and Pacific Oceans. This makes for strong spatial and temporal differences in water parameters that require local tuning of the ISA.

One of the most significant challenges posed by the Arctic is the presence of shallow mixed layer depths. The Arctic Ocean is generally classified as a $\beta$ ocean, that is, it is generally strongly stratified in salinity but not always in temperature [11]. This results in mixed layer depths that are highly variable and strongly seasonal due to the salinity effects from the freezing and melting of sea ice. This can result in very shallow mixed layer depths during the summer months that can make the prediction of ice presence based on temperature challenging as surfacing floats may not be able to reach this depth to get an indication of surface temperatures before an ascent must be aborted. This is even true during the winter months in the Arctic where mixed layer depths can be as shallow as 25 meters [12]. Because of this, the pressure range in which the ISA makes its determination is particularly important in the Arctic as it must ensure that it is in the mixed layer to accurately predict the presence of ice.

Another significant factor that presents itself in the data is the presence of a warm subsurface layer that exists in Beaufort and Arctic Seas. This warm subsurface layer is the result of intrusion of warm salty water from the Pacific subducting below the less dense arctic waters as it passes through the Bering Strait [13]. This results in a strong temperature inversion in the Beaufort and Arctic Seas that is present in a large number of the profiles used in this paper. The presence of a temperature inversion is likely to affect the ISA because it signals a warm region that may not be indicative of the true surface conditions and therefore the presence of ice. It may also affect the use of a threshold value that is determined by a median taken over a pressure range (as is currently used on Argo floats) because these statistics will represent lagged values.
that are not stationary if a float makes its ISA determination as it passes through such an inversion.
Chapter 3

Data

3.1 Data Overview

To collect historical data and analyze its performance, CTD profiles must be matched with satellite data to determine the presence of ice. In this analysis, CTD data comes from two sources: Argo floats and Ice Tethered Profilers (ITP). Both CTD sources provide a latitude and longitude with each profile. This is cross-referenced with sea-ice extent data provided by the Multisensor Analyzed Sea Ice Extent – Northern Hemisphere (MAISE-NH) dataset that provides information on whether ice is present at the location of each profile.

3.2 MAISE-NH Data

Sea Ice data comes from the Multisensor Analyzed Sea Ice Extent – Northern Hemisphere (MAISE-NH) product produced by the National Snow and Ice Data Center (NSIDC) [14]. The data is derived from another product produced by the National Ice Center (NIC) called the Interactive Multisensor Snow and Ice Mapping System (IMS) which uses visible imagery, passive microwave data, and in situ data to determine the boundaries of sea ice extent in the northern hemisphere [15]. Data from MAISE-NH is provided in 4km and 1km resolutions on a daily basis beginning in 2006 and 2014, respectively. For the purposes of this paper, the 4km resolution is
used because it provides a consistent dataset that covers the majority of time that
relevant Argo and Ice Tethered Profiler data is available to be compared with.

3.3 Argo Data

Argo data comes from the Argo program as previously described. The data used
for this analysis includes all profiles north of 65°N (approximately 31,000 profiles).
The majority of these profiles are in the Greenland and Norwegian Sea consisting
of mostly ice-free profiles with smaller groups in the Beaufort Sea, the Barents Sea,
Baffin Bay, and the Arctic Ocean. It should be noted that relatively few of the total
number of Argo profiles represent under-ice data (see figure 3-1).

Figure 3-1: Location and ice presence of Argo profiles used in this analysis.
3.4 Ice Tethered Profiler Data

To augment the number of under-ice profiles and get greater representation in the Arctic and Beaufort Seas, data from the Ice Tethered Profiler (ITP) program was used [16]. This dataset comes from a Woods Hole Oceanographic Institution (WHOI) program that has installed profiling CTDs on ice floes in the Arctic Ocean since 2004. These instruments are installed on the surface of ice floes with a cord that descends into the water. A CTD (similar to those used by the Argo program) runs along this cord using a traction motor between the surface and the desired depth (typically between 500m and 800m) measuring water parameters during ascent and descent. Both GPS localization and data transmission take place from the surface installation allowing for near year-round availability of the data. It should be noted that the sampling rate of the Ice Tethered Profilers is much higher than that of Argo floats. This results in a much larger number of available profiles that are concentrated in a smaller geographical area. All available profiles from the ITP dataset were collected for use in this analysis (approximately 104,000 profiles) and are concentrated primarily in the Beaufort Sea and Arctic Ocean (see figure 3-2).
3.5 Data Assimilation

In total, CTD profiles comes from two sources: Argo floats and Ice Tethered Profilers. Two sources of CTD measurements are required to gather a sufficient amount of data for analysis. Specifically, ITP data is required to augment the number of profiles available in the Beaufort and Arctic since very few Argo profiles are represented in these regions. Both sources provide similar vertical conductivity, temperature and pressure profiles of the water column.

Profiles are then filtered and assimilated into a single labeled and standardized dataset. To do this, the date of each profile is matched with the corresponding daily sea-ice extent from the MASIE-NH dataset. The profile’s location is matched to the nearest pixel (4km resolution) of the sea ice extent for that day. This is used to determine whether each represents an under-ice or ice-free profile and a corresponding flag is assigned to the data. If no MASIE-NH data is available for the date of a profile,
that profile is excluded from the final dataset.

Once matched with sea ice extent data, the labeled profiles are then further filtered and standardized to ensure each profile represents acceptable and complete data. For this analysis, profiles are linearly interpolated between 10dbar and 200dbar at a 1dbar interval. Any profile with data gaps greater than 15dbar is excluded from the final dataset.

For a regional analysis, each profile is also regionally labeled based on latitude and longitude. The regions analyzed include: the Beaufort Sea, Arctic Ocean, Norwegian Sea, Greenland Sea and Baffin Bay. Each of these regions are used to roughly approximate the different areas Argo floats may be deployed and bin them accordingly so the performance of the algorithm can be analyzed and adjusted on a regional level as necessary (see figure 3-3).

Figure 3-3: The regional bounds used for this analysis. Regional bounds are used to roughly approximate expected deployment and operation areas for Argo floats. These regions are used to analyze and locally tune the ISA in subsequent chapters.
3.6 Representative Sampling

Once assimilated, it was realized that the spatial distribution of profiles was not uniform. Instead, a large number of the profiles in the dataset were concentrated in the Arctic and Beaufort regions. This is a result of the region being primarily sampled by ITP stations which take profiles at a much higher rate than Argo floats. This can be seen in figure 3-4 where the spatial histogram shows significantly more profiles in the two regions. This is problematic for optimization as it signifies unrepresentative data that may introduce a regional bias into the sample. If fit using the unrepresentative data, a model will likely overfit for the oversampled regions and ignore the less sampled regions.

To correct for this and create a more representative sample from the data, a random selection of profiles was chosen from each bin of the spatial histogram. The random selection was capped at a maximum of 110 profiles per bin (corresponding to roughly 1 profile per 40km$^2$) limiting the number of profiles in regions that are over-sampled and producing a more spatially uniform sample. The change in the number of profiles used in the dataset and resulting spatial distribution can be seen in tables 3.2 – 3.5 and figure 3-5 respectively.

It was also sought to match the ratio of under-ice to ice-free profiles to the true year-round distribution of ice that can be expected in each region. It was found in each region that the sampled ratio (the ratio of under-ice to ice-free profiles used in the dataset) did not match the true ratio of ice presence (calculated as the year-round average ice coverage within the bounds of each region given by the MASIE-NH data and shown in table 3.1). The cause of this imbalance tracks with the primary sampling strategies for each region. For example, in the Arctic and Beaufort regions which are primarily sampled by Ice Tethered Profilers, under-ice profiles are overrepresented in comparison to the true year-round average ice coverage in the regions. Likewise, in the Barents, Greenland, and Norwegian Seas which are primarily represented by Argo profiles, ice-free profiles are overrepresented. This is attributed to the sampling strategies of each of the programs. Ice Tethered Profilers are expected to sample
Figure 3-4: Spatial histogram of profiles prior to spatial normalization. The high concentration of profiles in Arctic and Beaufort are likely to introduce a regional bias when fitting models and must be accounted for.

Figure 3-5: Spatial histogram of profiles after spatial normalization. The more uniform distribution will help prevent a strong regional bias from emerging.

under-ice profiles at a much higher rate than the true distribution because they are purposely deployed on multi-year ice. Likewise, Argo floats are typically deployed in ice-free oceans and although they drift freely with the deep ocean currents, will still
<table>
<thead>
<tr>
<th>Region</th>
<th>Proportion of under-ice Profiles Prior to Oversampling</th>
<th>Year-round Average Ice Presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beaufort</td>
<td>97.4%</td>
<td>80.3%</td>
</tr>
<tr>
<td>Arctic</td>
<td>98.9%</td>
<td>96.1%</td>
</tr>
<tr>
<td>Norwegian</td>
<td>1.3%</td>
<td>7.2%</td>
</tr>
<tr>
<td>Greenland</td>
<td>5.2%</td>
<td>27.4%</td>
</tr>
<tr>
<td>Baffin</td>
<td>14.6%</td>
<td>59.4%</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of the proportion of under-ice profiles in the dataset to the year-round average ice presence in each region prior to oversampling. Oversampling corrects this by duplicating samples of the underrepresented class to match the year-round average ice presence in each region.

typically operate in ice-free environments generating a bias towards ice-free profiles in these regions. Together, this imbalance represents an unrepresentative sample that will create misleading results in statistical models if not properly accounted for.

To correct for this regional imbalance, an oversampling strategy was used to match the ratio of under-ice to ice-free profiles with the true year-round average ice coverage in each region. This was done by duplicating a random sample of profiles from the under-represented class until the ratio of profiles matches the true year-round average ice coverage for each region. In this way, the number of under-ice and ice-free profiles is representative of the actual distribution of ice in each region and will avoid bias when fitting statistical models on the dataset. Doing so also allows us to effectively analyze the performance of the algorithm using the dataset as the distributions will match what could reasonably be expected in these regions. Tables 3.2 – 3.5 show the change in the number of profiles in the dataset through the full data assimilation and representative sampling processes.
<table>
<thead>
<tr>
<th></th>
<th>Argo</th>
<th>ITP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>31615</td>
<td>104412</td>
</tr>
<tr>
<td>Matched and</td>
<td>24944</td>
<td>60515</td>
</tr>
<tr>
<td>Standardized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatially</td>
<td>23118</td>
<td>31282</td>
</tr>
<tr>
<td>Normalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oversampled</td>
<td>30805</td>
<td>32559</td>
</tr>
</tbody>
</table>

Table 3.2: Change in the number of Argo and ITP profiles during the different steps of data preparation.

<table>
<thead>
<tr>
<th></th>
<th>Ice-Free</th>
<th>Under-Ice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched and</td>
<td>23637</td>
<td>61822</td>
</tr>
<tr>
<td>Standardized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatially</td>
<td>21866</td>
<td>32534</td>
</tr>
<tr>
<td>Normalized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oversampled</td>
<td>25363</td>
<td>38001</td>
</tr>
</tbody>
</table>

Table 3.3: Change in the number of ice-free and under-ice profiles during the different steps of data preparation.

<table>
<thead>
<tr>
<th></th>
<th>Beaufort</th>
<th>Arctic</th>
<th>Norwegian</th>
<th>Greenland</th>
<th>Baffin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Argo</td>
<td>ITP</td>
<td>Argo</td>
<td>ITP</td>
<td>Argo</td>
</tr>
<tr>
<td>Matched and</td>
<td>336</td>
<td>33704</td>
<td>736</td>
<td>26729</td>
<td>10591</td>
</tr>
<tr>
<td>Standardized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatially</td>
<td>313</td>
<td>13640</td>
<td>705</td>
<td>17560</td>
<td>10223</td>
</tr>
<tr>
<td>Normalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oversampled</td>
<td>2482</td>
<td>14445</td>
<td>1219</td>
<td>17569</td>
<td>10734</td>
</tr>
</tbody>
</table>

Table 3.4: Change in the number of Argo and ITP profiles by region during the different steps of data preparation.

<table>
<thead>
<tr>
<th></th>
<th>Beaufort</th>
<th>Arctic</th>
<th>Norwegian</th>
<th>Greenland</th>
<th>Baffin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ice-free</td>
<td>Under-ice</td>
<td>Ice-free</td>
<td>Under-ice</td>
<td>Ice-free</td>
</tr>
<tr>
<td>Matched and</td>
<td>434</td>
<td>33606</td>
<td>211</td>
<td>27254</td>
<td>10481</td>
</tr>
<tr>
<td>Standardized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatially</td>
<td>365</td>
<td>13588</td>
<td>210</td>
<td>18055</td>
<td>10113</td>
</tr>
<tr>
<td>Normalized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oversampled</td>
<td>3339</td>
<td>13588</td>
<td>733</td>
<td>18055</td>
<td>10113</td>
</tr>
</tbody>
</table>

Table 3.5: Change in the number of Ice-free and Under-ice profiles by region during the different steps of data preparation.
Chapter 4

Statistical and Machine Learning Methods

4.1 Overview of Methods Used

Three methods are proposed to determine the presence of ice above a surfacing Argo float. Each method provides a predicted probability of ice presence and is fit based on the historical CTD profiles gathered for this study. The first method develops an approach similar to Klatt et al. [7] by utilizing temperature to predict the probability of ice presence above a surfacing float. Probabilities are generated by fitting the historical data to a logistic regression model. The second method explores new predictive variables to increase the accuracy of the ISA, and uses temperature, salinity, and time of year to predict the presence of ice above a surfacing float. Again, probabilities are generated by fitting the historical data to a multiple logistic regression model. The final method implements a recurrent neural network to predict the presence of ice given the full CTD profile of a surfacing float. Predicted probabilities are generated by fitting the neural network model using stochastic gradient descent. Results for each method are reported and compared in their respective sections below.
4.2 Regression and Neural Network Techniques

The first two methods utilize logistic regression models to predict the presence of ice above a surfacing float. Logistic regression models are useful to predict binary variables and are capable of generating outputs that can be interpreted as probabilities. These models operate on the equation:

\[
\frac{1}{1 + e^{c_0 + \sum_n c_n x_n}}
\]  

(4.1)

where \(c_0\) and \(c_n\) represent the coefficients determined by the regression and \(x_n\) represents the values of the predictive variables. The output ranges between 0 and 1 and can be interpreted as the predicted probability of the binary variable. Logistic regression models are fit by minimizing the log-loss error between the predicted probabilities and the true binary values.

The last method implements a recurrent neural network. While these are far more complicated than a logistic regression, their output can similarly be interpreted as the predicted probability of the binary variable. Just like logistic regression, these models also attempt to minimize the log-loss error between the predicted probabilities and the true binary values, but do it through stochastic gradient descent.

4.3 Binary Classification Statistics

Much of the analysis in this paper is based on binary classification. This involves classifying a binary variable, in this case under-ice or ice-free profiles, based on predictive variables. While much of this analysis explains and provides examples of the principles and methods used, it is useful to have a working knowledge of such an analysis. Cited are a few papers that provide an introduction to the metrics and methods of binary classification, many of which are used in this paper and may be useful in understanding the subject [17, 18].
<table>
<thead>
<tr>
<th>Metric</th>
<th>Calculation</th>
<th>Interpretation as Conditional Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>( \frac{TP + TN}{TP + TN + FP + FN} )</td>
<td>( P(\text{correctly classified}) )</td>
</tr>
<tr>
<td>Misclassification Rate</td>
<td>( \frac{TP + TN + FP + FN}{FP + FN} )</td>
<td>( P(\text{incorrectly classified}) )</td>
</tr>
<tr>
<td>False Positive Rate (FPR)</td>
<td>( \frac{TN + FP}{TN + FP} )</td>
<td>( P(\text{predicted under-ice}\mid \text{ice-free}) )</td>
</tr>
<tr>
<td>False Negative Rate (FNR)</td>
<td>( \frac{FN + TP}{FN + TP} )</td>
<td>( P(\text{predicted ice-free}\mid \text{under-ice}) )</td>
</tr>
<tr>
<td>False Discovery Rate (FDR)</td>
<td>( \frac{TP + FN}{TP + FN} )</td>
<td>( P(\text{ice-free}\mid \text{predicted under-ice}) )</td>
</tr>
<tr>
<td>False Omission Rate (FOR)</td>
<td>( \frac{TN + FN}{TN + FN} )</td>
<td>( P(\text{under-ice}\mid \text{predicted ice-free}) )</td>
</tr>
</tbody>
</table>

Table 4.1: Table of various binary classification metrics used in this analysis including their calculation and interpretation as conditional probabilities. TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

Because each of these methods predict a binary variable (under-ice or ice-free), classification errors can be thought of as false positives (predicting under-ice when the surface is ice-free) and false negatives (predicting ice-free when the surface is under-ice). This allows for the use of classification metrics such as false positive rate, false negative rate, and misclassification rate that describe the performance of a test that reports a binary variable. These metrics can be used to analyze the performance of the ISA when tested on the dataset and provide insight to how the algorithm will perform. A table of the relevant metrics used in this analysis is given below that describes the calculation of each metric and its interpretation as a conditional probability (see table 4.1).

Given the primary purpose of the ISA to predict and avoid the dangers of ice, it is important to have a low false negative rate. In other words, it is important for a test to successfully predict that the buoy is under-ice given that there is ice present at the surface. While it is easy to adjust a model to minimize this statistic, it comes with the tradeoff of a higher false positive rate (representing more missed surfacing opportunities). This is the primary issue posed by the ISA as it balances the need to protect the float (by minimizing the number of false negatives) with the
need to surface to in order to transmit data (by preventing false positives). This requires a balance between the two metrics, but that balance is subject to the needs and risk tolerance of individual Argo missions. For example, an Argo mission that uses an expensive float model might wish to prioritize survivability, and may accept a larger number of missed surfacing opportunities (false positives) to decrease the risk of surfacing into ice (false negatives). This means that there is no single optimal classification threshold, but it instead dependent upon a mission’s requirements.

To account for this, results are reported in two ways. First, an ‘optimal’ classification threshold is provided that reports the values necessary to minimize the misclassification rate, including both false positives and false negatives. This is considered the ‘optimal’ solution because it minimizes the total number of misclassified profiles. Second, results and plots are provided showing the performance of the algorithms over a wide range of classification thresholds so users can observe the performance of ‘off-optimal’ results and select a classification threshold that fits their needs and risk tolerance. Because the false negative rate (representing surfacing into ice and potentially damaging a float) is the most significant of these metrics, the required thresholds to achieve given values of the FNR are also tabulated.

Lastly, while the false negative rate and false positive rate which describe the performance of the ISA with regard to the true presence of ice are primarily considered in this analysis, it is also important to consider the false discovery rate (FDR) and false omission rate (FOR) which describe the performance of the ISA with regard to the predicted presence of ice (as opposed to the true presence of ice). These are important metrics for the regional models as they are dependent upon the distribution of under-ice and ice-free profiles. If the distributions are highly asymmetric (i.e. the number of under-ice profiles is significantly different from the number of ice-free profiles), the metrics may differ significantly from the FNR and FPR. Described as a conditional probability, the FDR represents the probability that a profile is ice-free given that it is predicted to be under ice and the FOR represents the probability that a profile is under-ice given that it is predicted to be ice-free.
Chapter 5

Modified Temperature Threshold Method

5.1 Overview

Previous work on an ice sensing algorithm has focused on the use of a temperature threshold value to predict the presence of ice above a surfacing Argo float. Current ISA strategies use temperature as the only variable when deciding if ice is above a surfacing float. While this has been shown to work, current methods provide no indication of the confidence or probability of ice presence. In this section, a new method is presented that is capable of generating a predicted probability of ice presence using temperature as the only predictive variable. By using temperature as the only predictive variable, the performance of this new method can be analyzed and compared with current ISA strategies that make use of a temperature threshold.

To generate a probability of ice presence, a logistic regression model is presented that is fit using the dataset of historical profiles matched with ice presence. As discussed in section 4.2, logistic regression models are capable of generating probabilities for a binary classification, in this case the presence of ice. Using temperature as the only predictive variable, the general form of the equation given in section 4.2 becomes:
\[ P(\text{ice}) = \frac{1}{1 + e^{c_0+c_1\text{(temp)}}} \]  
\[ \text{(5.1)} \]

where the coefficients are determined by the regression.

Using the estimated probability of ice output by the regression model, a probability threshold can be used to abort an ascent if the estimated probability of ice is deemed too high. While the probability threshold of 0.5 is often used (probabilities less than 0.5 are predicted to be ice-free and probabilities greater than 0.5 are predicted to be under-ice), the purpose of this is to allow a user to adjust the classification threshold (also referred to as a probability threshold) according to their mission requirements and risk tolerance. Using the historical dataset, the performance of the algorithm can be shown over the full range of probability thresholds. The goal of this is to allow users to observe the performance of the ISA over a range of thresholds and select a threshold according to their needs.

### 5.2 Best Depth Determination

To first find the best depth at which to determine the threshold, the performance of the ISA is compared by performing the regression using temperatures from a range of depths. This is done by conducting the regression 92 times at a 1dbar pressure interval between 10dbar and 100dbar inputting the temperature of all profiles from the respective depth. The relevant metrics vs depth are then compared to determine the pressure at which the ISA performs the best. Results are displayed in figure 5-1 and show that the ISA performs best near the surface where the lowest misclassification rate, false positive rate and false negative rate are all present. This shows that a threshold value is best determined when measurements are taken closest to the surface. This conclusion makes intuitive sense as temperature near the surface is expected to be highly correlated with temperature at the surface and thereby the presence of ice. As discussed in section 2.3, this conclusion also matches with the
physical oceanography of the Arctic where relatively shallow mixed layer depths are likely to affect the performance of the ISA all the way to the near-surface.

![Graph showing misclassification rate, false negative rate, and false positive rate for models trained with temperature from different depths. The models show the best performance using measurements from the shallowest depth.](image)

Figure 5-1: Misclassification rate, false negative rate, and false positive rate for models trained using temperature from different depths. The models show the best performance using measurements from the shallowest depth.

While the exact depth required for an Argo float to abort an ascent varies with ascent rate and conditions [3], results are reported under the assumption that 15dbar is the shallowest allowable depth to make an ISA determination and abort an ascent accordingly. While a number of papers have recommended 10dbar as a minimum depth for the ISA [6, 19], it has been found that certain float models require a greater depth to abort an ascent (14dbar for the case of Apex9 floats) [10]. For the remainder of this analysis 15dbar will be used and referred to as the ‘best-depth’ to make an ISA determination to be inclusive of all float models.

### 5.3 Optimal Measurement Strategy

While testing, it was qualitatively observed that the threshold can be determined more accurately with a single point measurement as opposed to using a mean, median,
or minimum temperature over a pressure range. Upon inspection, it was determined that this warranted investigation as the mean, median, and minimum over a pressure range are likely to report lagged values. This is because these statistics are calculated using past measurements and may not be representative of the most recent data as the float ascends. A sample temperature profile highlights this (see figure 5-2). In the figure provided, if an Argo float uses a median value measured between 50dbar and 15dbar (a pressure range of 35dbar), the median value will be outside the mixed layer depth and will not be representative of the surface conditions when the float reaches 15dbar and makes its ISA determination. This is true of many profiles in the Arctic where shallow mixed layer depths and inversions exist all the way to the near-surface as discussed in section 2.3. In contrast, point measurements report the most recent data which is closest to the surface and a better predictor of ice as shown.

![Sample Temperature Profile](image)

Figure 5-2: A sample temperature profile from the Beaufort region. The plot highlights the potential problem of lagged statistics caused by measuring a mean or median over a pressure range as they are likely to report values that are not representative of the surface conditions.

To confirm this observation, the regression was conducted at a 1dbar interval between 10dbar and 100dbar to compare the performance of point observations with the mean, median, and minimum taken over a 25dbar pressure range with measurements taken every 1dbar. In general, results show a lower misclassification rate by point
measurements over any of the three metrics taken over a pressure range (see figure 5-1). While the difference is generally small, it is most significant in the range where the thermocline can be expected (approximately 25-50dbar) with point measurements being misclassified at a rate 1-2% lower than those taken over a pressure range. This is attributed to the rapidly changing temperature at or near the thermocline that cause the metrics taken over a pressure range to report significantly lagged values thereby decreasing performance. At the best-depth of 15dbar, point measurements outperform the mean and median by 0.9% and 1.3% respectively. While this may seem small, the misclassification rate is around 5%, so a 1% increase represents a 20% jump in the number of misclassified profiles which is significant.

While this is true using a 25dbar pressure range, it was also tested using a smaller 10dbar range to see if the effect is less pronounced. Figure 5-4 displays the results using a 10dbar range and shows that point measurements still slightly outperform the metrics taken over a pressure range by a margin of around 0.5-1.0%. Although the difference is small, this is true at the best-depth of 15dbar where the mean and median have a higher misclassification rate by 0.3% and 0.4% respectively. This still suggests that even with a pressure range of 10dbar, point measurements still slightly outperform any of the statistics taken over a pressure range.

Because point measurements are the best metric at the best-depth of 15dbar, the use of point measurements is recommended over a mean, median, or minimum taken over a pressure range. The use of point measurements also significantly simplifies the ISA and this analysis in general. For the remainder of this paper, point measurements are utilized and results are reported accordingly. It should be recognized however that the use of a pressure range provides some degree of redundancy in the event of erroneous temperature readings. For that reason, it is recommended that point measurements are utilized, but their precision verified by past observations. Alternatively, in the case of current ISA strategies that require the use of a pressure range, it is recommended that a small pressure range ($\leq 10$dbar) be selected so as not to significantly affect the performance of the ISA.
Figure 5-3: Misclassification rate of point measurements in comparison to the mean, median, and minimum taken over a 25dbar range.

Figure 5-4: Misclassification rate of point measurements in comparison to the mean, median, and minimum taken over a 10dbar range.
5.4 Results

Knowing the best-depth of 15dbar and that the use of point measurements is the preferred measuring strategy, the performance of the model is analyzed when trained using only a single measurement taken at 15dbar. Coefficients of the regression are reported in table 5.2 and results at the misclassification rate minimum are displayed in table 5.1. The performance of the regression over the full range of probability thresholds can also be found in figure 5-5.

Overall, results are positive with the model reporting a misclassification rate of 4.2% of profiles in the dataset. Figure 5-5 provides a view of the performance of the model over the full range of probability thresholds. In the plot, the misclassification rate represents the fraction of profiles that are classified incorrectly, the false positive rate represents the fraction of ice-free profiles that are incorrectly predicted to be under ice (representing missed surfacing opportunities), and the false negative rate represent the fraction of under-ice profiles that are incorrectly classified as ice-free (representing surfacing into ice and potentially damaging the float). The purpose of this is to allow a user to select a probability threshold based on their risk tolerance and mission requirements and analyze how it would perform on the full dataset of profiles in the Arctic. It should be noted that this model is trained on the full dataset and so is not indicative of the local or regional performance of the algorithm (see section 5.6 for regional performance and local tuning). Instead, it can be interpreted as the performance of the algorithm over the Arctic in general since it is represented by the full dataset.

Looking again at figure 5-5, the relationship between the false positive rate and false negative rate can be seen if the probability threshold is adjusted. A lower threshold classifies profiles more conservatively favoring a lower false negative rate at the expense of a higher false positive rate. Conversely, a higher threshold classifies profiles less conservatively, favoring a lower false positive rate at the expense of a higher false negative rate. This highlights the balance in adjusting the threshold as the tradeoff between missed surfacing opportunities (false positives) and risk of
damage (false negatives) is affected by the probability threshold.

While 0.5 is often used as the default classification threshold, it can be observed that the misclassification rate does not reach its minimum at this value, instead occurring around probability 0.8. This is an artifact of the loss function and distribution of temperatures in the dataset. Whereas an under-ice profile will almost always be around freezing temperature, ice-free profiles can be any range of higher temperatures. Because of this, a single misclassified under-ice profile is likely to incur a much higher loss than a misclassified ice-free profile. This will cause the model to favor type I (false-positive) errors to reduce the total loss, and will shift the misclassification rate minimum higher than 0.5. To account for this imbalance in the dataset, ‘optimal’ results are reported at the misclassification rate minimum rather than at probability 0.5. We refer to the misclassification rate minimum as the optimal threshold because it results in the smallest number of total misclassifications. This optimal threshold is reported in table 5.1 where the corresponding misclassification rate, false negative rate, and false positive rate are also displayed.

![Graph showing performance of the model when trained and tested on the full dataset.](image)

**Figure 5-5:** Performance of the model when trained and tested on the full dataset. Note the misclassification rate minimum does not occur at probability 0.5. The ‘optimal’ value is reported as the threshold that minimizes the misclassification rate, rather than at probability 0.5.
5.5 Conversion to Temperature Threshold

The model trained on the regionally aggregated data reports optimal coefficients according to the equation provided in table 5.2. Because the model is trained using temperature as the only predictive variable, the equation can be reversed to convert probability thresholds into temperature thresholds. Reversing the general form of the equation to solve for temperature, results in:

\[ temp = \frac{\ln \left( \frac{1}{P(\text{ice})} - 1 \right) - c_0}{c_1} \]  \hspace{1cm} (5.2)

Using the probability value at the misclassification rate minimum (0.78), the optimal temperature threshold over the full dataset equates to -0.782°C. Again, it should be noted that this represents the optimal temperature threshold over the full dataset and may not be suitable for certain regions (see section 5.6 for regional performance and local tuning). The same can be done for the full range of probability thresholds. Figure 5-6 converts the range of probability thresholds given in figure 5-5 to temperature thresholds. In this way, the performance of the model can be viewed in terms of current ISA methods that use a temperature threshold.

Figure 5-6 displays the performance of the model in terms of a temperature threshold, and can be used to determine the performance of the ISA over the full dataset. In the figure, there is a clear decrease in performance with threshold values lower than approximately -1.2°C. Below this point, the overall misclassification rate and false negative rate increase significantly as under-ice profiles are increasingly predicted to be ice-free. Likewise, above approximately -0.5°C, the misclassification rate and false positive rate increase, although more gradually, because ice-free profiles are increasingly classified as under-ice. The optimal threshold value is reported as -0.782°C and occurs at the minimum of the misclassification rate. This can be described as the optimal temperature threshold required to minimize the total number of misclassifi-
cations over the full dataset.

Figure 5-6: Performance of the model when trained and tested on the full dataset. The x-axis is converted from probability to temperature according to equation 5.5 so the performance of the model can be viewed in terms of current ISA methods that use a temperature threshold.

5.6 Regional Performance and Local Tuning

While the model performs well over the full dataset, regional tuning has been shown to improve the performance of the ISA [10]. As discussed in section 2.3, the differences between the various water masses of the Arctic are significant and regional tuning can account for many of the large-scale differences that may affect the performance of the algorithm. Since the dataset has been oversampled to match the regional distributions of ice presence, models trained on the individual regions will also reflect the expected performance within those regions. This is used to conduct regional tuning of the threshold and analyze the expected performance within the regions.

To accomplish regional tuning of ISA, a model is trained for each of the individual regions. This is done using only the profiles from their respective regions, again
inputting only a single temperature measurement at 15dbar. Much like what was done for the full dataset, the optimal probability and temperature thresholds that minimize the misclassification rate within each region are reported (see table 5.1). We also report the coefficients determined by each model (see table 5.2) and provide a plot of the performance over a range of probability and temperature thresholds (see figures 5-7 – 5-11).

Looking at the optimal misclassification rates for each region provided in table 5.1, it can be observed that some regions have much higher misclassification rates than others. The Greenland region has the highest of these misclassification rates at 8.6% meaning that it is the least predictable of the regions. At the same time, it has a false-negative rate of 24.1% showing that even using the threshold that minimizes the total number of misclassifications, almost a quarter of under-ice profiles will incorrectly surface into ice. This can still be decreased however by raising the temperature threshold, but comes at the expense of increasing the overall misclassification rate and false positive rate.

While the thresholds reported in table 5.1 minimize the misclassification rate including both false-positives and false-negatives, it is important to consider the primary purpose of the ISA to predict and avoid the dangers of ice. For this reason, the false negative rate (representing the rate at which under-ice profiles incorrectly surface into ice and risk damage) is an important statistic to consider and minimize. For that reason, tables 5.3 and 5.4 are provided that present threshold values that are required to achieve a desired false-negative rate as provided by the dataset. Returning to the example in the Greenland region, if the FNR of 24.1% at the misclassification rate minimum is considered unacceptable, the temperature threshold can be adjusted to instead target an FNR of 0.1 (10%) for example using this table. This would require a temperature threshold of 0.283°C as provided by the dataset. While the same information can be derived from the performance plots provided for the individual regions, the table provides an easy to understand format that can be quickly scanned for potential thresholds. The plots should still be considered however as they provide additional information regarding the other classification statistics that may need to
be considered. Returning again to the example, if a 10% FNR is targeted using a temperature threshold of 0.284°C, it will come at the expense of a very high false positive rate and false discovery rate as shown in figure 5-10. While the ‘optimal’ result is reported as the threshold that minimizes the total misclassification rate, these plots allow users to analyze the expected performance of the ISA under different conditions and select a different threshold if desired.

Lastly, all of the regional models are combined to analyze the performance of the ISA over the full dataset after regional tuning is accomplished. This is done by applying each profile to the respective model for its region and combining the results. The results can then be compared with the performance of the model trained on the full dataset to determine how much local tuning increases the performance of the ISA. The result is displayed in table 5.1 and shows that the regionally tuned models only improve the performance of the ISA by 0.1% when optimizing for the misclassification rate minimum. This suggests that regional tuning does not significantly improve the overall performance of the ISA. It does not however negate the purpose of regional tuning as the regional plots still allow users to view the expected performance of the algorithm in their respective regions and select thresholds accordingly.

Using the models, tables, and plots provided in this section, users are able to analyze the performance of the ISA over the dataset to determine an appropriate threshold for their needs. While the ‘optimal’ threshold is reported as the value that minimizes the total misclassification rate, users should take into account risk tolerance, regional distributions of ice, and expected performance of the algorithm including the misclassification rate, FNR, FPR, FDR, and FOR to determine a threshold that fits their requirements.

5.7 Summary

To summarize, in this section, it has been shown that the ISA shows the best performance when a threshold is determined closest to the surface. According to past papers on the topic, this is assumed to be 15dbar. It is also shown that point
measurements are preferred over any metrics taken over a pressure range; alternatively a small pressure range should be used if point measurements are not available.

Regarding the models, a logistic regression model is introduced that is capable of generating an expected probability of ice presence based on historical data and use this to determine optimal thresholds (both as probabilities and temperatures). While the optimal thresholds are reported as the thresholds that minimize the misclassification rate for each model, tables and plots are provided that allow users to analyze the expected performance of each model to determine the threshold appropriate for their own needs and risk tolerance.
<table>
<thead>
<tr>
<th></th>
<th>Optimal Probability Threshold</th>
<th>Optimal Temperature Threshold</th>
<th>Misclassification Rate</th>
<th>False Negative Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>0.774</td>
<td>-0.782</td>
<td>0.042</td>
<td>0.041</td>
<td>0.045</td>
</tr>
<tr>
<td>Beaufort</td>
<td>0.466</td>
<td>-0.716</td>
<td>0.056</td>
<td>0.016</td>
<td>0.218</td>
</tr>
<tr>
<td>Arctic</td>
<td>0.869</td>
<td>-1.153</td>
<td>0.005</td>
<td>0.002</td>
<td>0.082</td>
</tr>
<tr>
<td>Norwegian</td>
<td>0.631</td>
<td>-0.797</td>
<td>0.015</td>
<td>0.169</td>
<td>0.003</td>
</tr>
<tr>
<td>Greenland</td>
<td>0.624</td>
<td>-0.647</td>
<td>0.086</td>
<td>0.241</td>
<td>0.027</td>
</tr>
<tr>
<td>Baffin</td>
<td>0.786</td>
<td>-0.949</td>
<td>0.066</td>
<td>0.084</td>
<td>0.040</td>
</tr>
<tr>
<td>After Regional Tuning</td>
<td>N/A</td>
<td>N/A</td>
<td>0.041</td>
<td>0.038</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Table 5.1: Performance of the logistic regression model at the misclassification rate minimum for each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{0.782+2.575(\text{temp})}}$</td>
</tr>
<tr>
<td>Beaufort</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{3.723+5.010(\text{temp})}}$</td>
</tr>
<tr>
<td>Arctic</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{6.084+6.918(\text{temp})}}$</td>
</tr>
<tr>
<td>Norwegian</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{0.703+1.554(\text{temp})}}$</td>
</tr>
<tr>
<td>Greenland</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{0.468+1.507(\text{temp})}}$</td>
</tr>
<tr>
<td>Baffin</td>
<td>$P(\text{ice}) = \frac{1}{1 + e^{0.979+2.402(\text{temp})}}$</td>
</tr>
</tbody>
</table>

Table 5.2: Equations to generate a probability of ice presence as determined by the logistic regression model for each region.
**Table 5.3:** Tabulation of the required probability threshold to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>FNR=0.2</th>
<th>FNR=0.1</th>
<th>FNR=0.05</th>
<th>FNR=0.025</th>
<th>FNR=0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>0.952</td>
<td>0.925</td>
<td>0.838</td>
<td>0.523</td>
<td>0.067</td>
</tr>
<tr>
<td>Beaufort</td>
<td>0.958</td>
<td>0.909</td>
<td>0.803</td>
<td>0.631</td>
<td>0.355</td>
</tr>
<tr>
<td>Arctic</td>
<td>0.994</td>
<td>0.991</td>
<td>0.988</td>
<td>0.986</td>
<td>0.979</td>
</tr>
<tr>
<td>Norwegian</td>
<td>0.729</td>
<td>0.363</td>
<td>0.058</td>
<td>0.023</td>
<td>0.001</td>
</tr>
<tr>
<td>Greenland</td>
<td>0.513</td>
<td>0.283</td>
<td>0.094</td>
<td>0.037</td>
<td>0.004</td>
</tr>
<tr>
<td>Baffin</td>
<td>0.942</td>
<td>0.805</td>
<td>0.539</td>
<td>0.366</td>
<td>0.004</td>
</tr>
</tbody>
</table>

**Table 5.4:** Tabulation of the required temperature threshold (in °C) to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>FNR=0.2</th>
<th>FNR=0.1</th>
<th>FNR=0.05</th>
<th>FNR=0.025</th>
<th>FNR=0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>-1.464</td>
<td>-1.279</td>
<td>-0.942</td>
<td>-0.339</td>
<td>0.719</td>
</tr>
<tr>
<td>Beaufort</td>
<td>-1.367</td>
<td>-1.202</td>
<td>-1.024</td>
<td>-0.850</td>
<td>-0.624</td>
</tr>
<tr>
<td>Arctic</td>
<td>-1.618</td>
<td>-1.559</td>
<td>-1.517</td>
<td>-1.495</td>
<td>-1.435</td>
</tr>
<tr>
<td>Norwegian</td>
<td>-1.089</td>
<td>-0.090</td>
<td>1.341</td>
<td>1.960</td>
<td>3.991</td>
</tr>
<tr>
<td>Greenland</td>
<td>-0.345</td>
<td>0.306</td>
<td>1.193</td>
<td>1.852</td>
<td>3.351</td>
</tr>
<tr>
<td>Baffin</td>
<td>-1.568</td>
<td>-0.998</td>
<td>-0.473</td>
<td>-0.179</td>
<td>1.889</td>
</tr>
</tbody>
</table>
Figure 5-7: Performance statistics for the Beaufort Sea region over various probability and temperature thresholds after local tuning.
Figure 5-8: Performance statistics for the Arctic region over various probability and temperature thresholds after local tuning.
Figure 5-9: Performance statistics for the Norwegian Sea region over various probability and temperature thresholds after local tuning.
Figure 5-10: Performance statistics for the Greenland Sea region over various probability and temperature thresholds after local tuning.
Figure 5-11: Performance statistics for the Baffin Bay region over various probability and temperature thresholds after local tuning.
Chapter 6

Multiple Logistic Regression Method

6.1 Overview

As an alternative to using temperature as the only predictive variable, new variables are proposed to better predict the presence of ice above a surfacing Argo float. Specifically, the addition of practical salinity and time of year proves a better predictor of ice presence than temperature alone. In this section, this method is proposed, tested, and analyzed on the dataset of historical profiles.

6.2 Best Predictive Variable Combination

Previous work on the ISA has used temperature as the sole predictor of ice above a surfacing Argo float. While this is the most significant variable in determining the presence of ice, it was sought to determine if any other available measurements could be used to increase the overall accuracy of the ISA. The number of available measurements from an Argo float are limited including temperature, practical salinity, pressure, and date. Using these, multiple derivative variables were determined that may be helpful in making an ISA determination including density, rate of change of temperature, and rate of change of salinity.

To provide a useful metric regarding time of year, the date is modeled as two functions: a sine and a cosine function according to the equations:
\[ y_{day, \sin} = \sin \left( \frac{2\pi y_{day}}{365} \right) \] (6.1)

\[ y_{day, \cos} = \cos \left( \frac{2\pi y_{day}}{365} \right) \] (6.2)

with \( y_{day} \) representing the day of the year as a number. The purpose of this is to represent the time of year as a cyclical number that repeats in a regular cycle to capture the yearly freezing/thawing cycle in the Arctic. This requires splitting the date into two features represented by both a sine and cosine function with a period of one year (365 days) to differentiate the full cycle. This method is commonly used in machine learning and is an effective encoding method for logistic regression models [20, 21]. The inclusion of time as a predictive variable will give the model an indication of what stage of the yearly freezing/thawing cycle the profile is in, potentially improving the performance of the model.

Additionally, temperature rate of change and salinity rate of change are both calculated as the rate of change over 5 dbar between the best-depth of 15 dbar and 20 dbar. This is done by simply subtracting the measurement at 15 dbar from the measurement at 20 dbar and dividing by 5 dbar to determine the rate of change in temperature or salinity per dbar. These measurements are thought to be potentially useful because if the rate of change is non-zero, it may give an indication to changing conditions and thereby value at the surface.

To determine what, if any, of these predictive variables might be used to increase the overall accuracy of the ISA, all possible combinations of the variables described (temperature, salinity, time of year, density, temperature rate of change, and salinity rate of change) were fit to multiple logistic regression models and their performance compared (both on the full dataset and regionally, using the same methods as in chapter 5). After evaluating each of these models, it was determined that using temperature, salinity, and time of year provided that best results.
6.3 Results

Using point observations from the best depth of 15dbar for the best combination of variables described above (temperature, salinity, and the year-day sin/cos functions), the logistic regression model reports optimal coefficients according to the equation given in table 6.3 when trained on the full dataset. Results are displayed in table 6.2 and 6.4 and figure 6-2 and show a minimum misclassification rate of 4.1%. This is an improvement of 0.1% in comparison to the threshold method of section 5.4. While this improvement may seem insignificant, when regionally tuned, its performance improves dramatically as will be discussed in section 6.4.

Using the coefficients provided by the model, the effect of the different variables on the predicted probability of ice presence is examined. The sign and magnitude of the coefficients show how the predicted probability of ice presence changes for individual variables with all else being equal. For an increase in any variable, a positive coefficient results in a lower probability of ice presence and a negative coefficient results in a higher probability. As expected, the positive coefficient in front of temperature shows that an increase in temperature results in a lower probability of ice.

The same is true of salinity where an increase in salinity results in a decrease in the predicted probability of ice presence with all else being equal due to its positive coefficient. This is also to be expected as higher salinity causes a decrease in the freezing temperature of sea water thereby decreasing the likelihood of ice for a given temperature.

The contribution of the combined year-day sine and cosine function can be plotted as a function of the day of the year to see that the combined term decreases during the winter and spring and increases during the summer and fall (see figure 6-1). Because a decreasing term causes an increased probability, this results in increasing probabilities of ice during the winter and spring peaking around day 123 of the year (beginning of May) and decreasing during the summer and fall reaching a minimum around day 306 (beginning of November). While this roughly corresponds with the freezing/thawing cycle in the Arctic, the cycle generated by the regression is shifted
nearly two months later in the year than the true average where the sea-ice maximum and minimum occur in mid-March and mid-September respectively. The exact cause of this is unknown, but could be the result of the fact that Argo floats are deployed in ice-free areas and often do not experience ice until later in the season resulting in a shift in the fit of the model.

Figure 6-1: Contribution of the combined $y_{day_{sin}}/y_{day_{cos}}$ predictive variables as a function of the day of the year ($y_{day}$) calculated as: $-0.331(y_{day_{sin}}) + 0.207(y_{day_{cos}})$

To analyze the relative importance of the different variables, the regression is conducted again with normalized predictive variables (each predictive variable is normalized in the range 0 to 1) so the magnitude of the coefficients can be compared relative to each other (see table 6.1). Doing so, it can be seen that temperature has the highest relative coefficient magnitude followed by salinity and the year-day sin/cos encoding (relative weight for year-day sin/cos calculated as sum of the magnitude of the two coefficients). This shows that temperature is the most significant variable in predicting the presence of ice followed by salinity and time of year due to its high weight. While the same is generally true for the regional models, the relative importance of salinity and time of year can be much higher in some cases as will be discussed in section 6.4.
6.4 Regional Performance and Local Tuning

Just as in section 5.6, models are locally tuned by fitting only the data from their respective region. Results, including performance statistics at the misclassification rate minimum and coefficients to the logistic regression for each region can be found in tables 6.2 and 6.3 respectively. Table 6.4 also displays the necessary probability threshold to achieve a range of desired false-negative rates. Figures 6-3 – 6-7 display the performance of each of the models over the full range of probability thresholds.

Again, the goal of these tables and figures is to provide a user with the relevant information to determine an acceptable ISA threshold if following this method. While ‘optimal’ results are reported as the thresholds that minimize the total misclassification rate, different thresholds can be chosen according to the plots to achieve the desired performance. Because the dataset was adjusted to match the true distribution of under-ice and ice-free profiles, the performance of these plots should be representative of what can be expected in each region.

While we will not rediscuss how to interpret all of the statistics and metrics offered
in the tables and plots as was done in chapter 5, we will highlight some of the notable inclusions. Firstly, from table 6.1 showing the relative importance of the variables, it can be seen that salinity and the year-day sin/cos function play more significant roles regionally than they do when trained on the full dataset. For example, in the Baffin region, the combined coefficient for the time of year is nearly as large as the coefficient for temperature signifying that the time of year is almost as good of a predictive variable as temperature in the region. This is because the Baffin region varies between being fully ice covered and ice-free on a yearly cycle, unlike the other regions which maintain a mix of under-ice and ice-free profiles year-round. Similarly, salinity plays a greater role in some of the regional models. For example, in the Arctic region, the magnitude of the coefficient for salinity is larger signifying that the salinity is a relatively strong predictor of ice presence in the region.

Lastly, each of the regional models are combined to analyze the performance of the ISA over the full dataset after regional tuning is accomplished. This is done by applying each profile to the model for its respective region and combining the results at the misclassification rate minimum. Overall regional tuning results in a 0.8% drop in the misclassification rate. This is significant as the original misclassification rate was just 4.1% representing an almost 20% improvement. The false-negative rate also decreases significantly, falling from 3.9% to just 2.6% when regionally tuned representing a 33% decrease in the false negative rate. This is important because the
false-negative rate represents the fraction of profiles that will unintentionally surface into ice, potentially resulting in damage to the float. The substantial improvement of the false-negative rate without increasing the false-positive rate shows that the regional tuning of these models is successful and is likely to improve the performance of the ISA if implemented.

6.5 Summary

To summarize, logistic classification using temperature, salinity, and time of year outperforms the use of temperature alone. While the inclusion of the new predictive variables results in a more complex algorithm, the significant performance gains after regional tuning may warrant their use. After regional tuning is applied, it is shown that the multiple logistic regression model results in a 20% decrease in the misclassification rate and a 32% decrease in the false negative rate in comparison to the modified temperature threshold method of chapter 5.
Table 6.2: Performance of the multiple logistic regression model at the misclassification rate minimum for each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Optimal Probability Threshold</th>
<th>Misclassification Rate</th>
<th>False Negative Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>0.760</td>
<td>0.041</td>
<td>0.039</td>
<td>0.044</td>
</tr>
<tr>
<td>Beaufort</td>
<td>0.377</td>
<td>0.051</td>
<td>0.010</td>
<td>0.221</td>
</tr>
<tr>
<td>Arctic</td>
<td>0.894</td>
<td>0.004</td>
<td>0.002</td>
<td>0.048</td>
</tr>
<tr>
<td>Norwegian</td>
<td>0.620</td>
<td>0.011</td>
<td>0.111</td>
<td>0.004</td>
</tr>
<tr>
<td>Greenland</td>
<td>0.566</td>
<td>0.074</td>
<td>0.185</td>
<td>0.032</td>
</tr>
<tr>
<td>Baffin</td>
<td>0.611</td>
<td>0.039</td>
<td>0.035</td>
<td>0.044</td>
</tr>
<tr>
<td>After Regional Tuning</td>
<td>N/A</td>
<td>0.033</td>
<td>0.026</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Table 6.3: Equations to generate a probability of ice presence as determined by the multiple logistic regression model for each region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>( P(ice) = \frac{1}{1 + e^{-0.464+2.548(temp)+0.038(psal)-0.331(yday_sin)+0.207(yday_cos)}} )</td>
</tr>
<tr>
<td>Beaufort</td>
<td>( P(ice) = \frac{1}{1 + e^{-10.086+3.846(temp)+0.366(psal)-3.665(yday_sin)-6.605(yday_cos)}} )</td>
</tr>
<tr>
<td>Arctic</td>
<td>( P(ice) = \frac{1}{1 + e^{-27.525+5.516(temp)+0.936(psal)-2.103(yday_sin)+0.208(yday_cos)}} )</td>
</tr>
<tr>
<td>Norwegian</td>
<td>( P(ice) = \frac{1}{1 + e^{-44.514+1.684(temp)+1.307(psal)+0.309(yday_sin)+1.463(yday_cos)}} )</td>
</tr>
<tr>
<td>Greenland</td>
<td>( P(ice) = \frac{1}{1 + e^{-12.228+1.646(temp)+0.357(psal)+1.011(yday_sin)-0.685(yday_cos)}} )</td>
</tr>
<tr>
<td>Baffin</td>
<td>( P(ice) = \frac{1}{1 + e^{-67.392+1.876(temp)+1.999(psal)-7.919(yday_sin)-0.937(yday_cos)}} )</td>
</tr>
</tbody>
</table>

Table 6.4: Tabulation of the required probability threshold to achieve various values of the false negative rate (representing the rate at which profiles will surface into ice, risking damage) for each region.

<table>
<thead>
<tr>
<th>Required Probability Threshold to Achieve:</th>
<th>FNR=0.2</th>
<th>FNR=0.1</th>
<th>FNR=0.05</th>
<th>FNR=0.025</th>
<th>FNR=0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>0.948</td>
<td>0.932</td>
<td>0.842</td>
<td>0.533</td>
<td>0.068</td>
</tr>
<tr>
<td>Beaufort</td>
<td>0.922</td>
<td>0.880</td>
<td>0.786</td>
<td>0.603</td>
<td>0.386</td>
</tr>
<tr>
<td>Arctic</td>
<td>0.994</td>
<td>0.988</td>
<td>0.980</td>
<td>0.975</td>
<td>0.972</td>
</tr>
<tr>
<td>Norwegian</td>
<td>0.753</td>
<td>0.532</td>
<td>0.263</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td>Greenland</td>
<td>0.583</td>
<td>0.265</td>
<td>0.108</td>
<td>0.048</td>
<td>0.007</td>
</tr>
<tr>
<td>Baffin</td>
<td>0.940</td>
<td>0.790</td>
<td>0.669</td>
<td>0.470</td>
<td>0.042</td>
</tr>
</tbody>
</table>
Figure 6-3: Performance statistics for the Beaufort Sea region over the full range of probability thresholds after local tuning.

Figure 6-4: Performance statistics for the Arctic region over the full range of probability thresholds after local tuning.
Figure 6-5: Performance statistics for the Norwegian Sea region over the full range of probability thresholds after local tuning.

Figure 6-6: Performance statistics for the Greenland Sea region over the full range of probability thresholds after local tuning.
Figure 6-7: Performance statistics for the Baffin Bay region over the full range of probability thresholds after local tuning.
Chapter 7

Neural Network Method

7.1 Overview

Finally, we seek to determine if a pre-trained neural network can be used to determine the presence of ice above a surfacing Argo float. While significantly more complicated than both the threshold and logistic regression methods previously described, neural networks have shown incredible predictive abilities in other fields. Given that a neural network is essentially a large set of matrix operations, a simple pre-trained neural network could reasonably be developed for use on an Argo float to determine the probability of ice during surfacing. In this section, the feasibility of this method is explored and a simplified network that could reasonably be implemented on future Argo floats is developed.

Using the dataset of historical profiles matched with ice presence, neural network models are fit to predict the presence of ice above a surfacing Argo float using the full temperature, salinity, and pressure profiles to 15dbar. Optimization is done using stochastic gradient descent to minimize the log-loss error (also called binary cross-entropy) of the classification and probabilities are generated as the output of the trained network.
7.2 Initial Testing

To begin, a simple densely connected neural network was tested to predict the presence of ice. It takes as input the full profile including temperature, salinity and pressure at a 1dbar interval between 20 and 200dbar as well as the year-day sin/cos function used in section 6.2 and outputs a predicted probability of ice presence. Once trained, the neural network was successfully able to predict the presence of ice with a 3.7% misclassification rate. In fact, results slightly outperformed both the threshold and logistic regression methods. This initial success prompted further investigation into the feasibility of a neural network for use in ice detection. Upon further examination, it was determined that a neural network would have to be both small (to limit computational complexity and power usage) and flexible (to allow for irregular measurements) for it to be feasibly implemented on an Argo float.

7.3 LSTM Network

To accomplish this, a second neural network was developed. This network is structured as a recurrent neural network (RNN) that makes use of long-short term memory (LSTM) nodes to predict the presence of ice. RNNs are a specialized type of neural network that are commonly used to process and classify sequential data. They are frequently used in time series forecasting as they contain memory that is able to recall encodings of past observations. While not explicitly a time series, measurements taken by a surfacing Argo float represent sequential data where current observations are related to previous measurements. In this way, an RNN may be a viable option to predict the presence of ice above a surfacing float. An added benefit of the RNN’s structure is that it does not require a rigid input shape and is instead flexible in the number of observations that are input. In this way, an RNN could be used to input measurements as they are available instead of mandating an exact pressure in which they are taken. This characteristic also allows a model to predict the presence of ice at any point, changing its confidence as it gains subsequent measurements. Together,
these benefits make the RNN network a robust and adaptable algorithm that could be usefully employed on an Argo float.

The RNN developed utilizes four hidden layers, each containing ten nodes. The first layer is an LSTM layer followed by three densely connected layers. The network was designed to be as small as possible, so as to minimize computational complexity and power usage if deployed a float, while still allowing for accurate prediction. The resultant network is extremely small containing only 981 parameters which would represent negligible computational load and power usage if implemented on an Argo float. The network takes as input temperature, practical salinity, pressure, and the yday sin/cos functions and outputs a predicted probability of ice being above the surfacing float. The input shape of the network is \((1, n_{\text{measurements}}, 5)\) allowing the network to adapt to the number of measurements taken by an Argo float. To prevent overfitting and ensure that the RNN would be able to generalize to irregular measurements, dropout was applied to the data during training and testing.

### 7.4 Results

The network was trained and evaluated using an 80/20 train/test split of randomly selected profiles to prevent overfitting. Results from the RNN show slightly lower performance in comparison to the first neural network with a 4.0% misclassification rate, but slightly outperform the metrics of the modified temperature threshold and multiple logistic regression methods when trained on the full dataset. Results are displayed in table 7.1.

In order to visualize how the predicted probabilities change with increasing number of measurements, two plots are presented where each shows the predicted probabilities of ice as the float ascends (see figures 7-1). The first shows the predicted probabilities without any dropped measurements (measurements taken every 1dbar between 200dbar and 20dbar) and the second shows the same probabilities with some measurements dropped (representing irregularly spaced data). Here, it can be observed that even with dropout of the data, the algorithm is able to generalize to the
profile and accurately predict the presence of ice when it reaches 15dbar.

It should be noted however that the neural network does not continuously increase the confidence of its predictions as it receives more information. Instead it takes seemingly random swings that are difficult to physically decipher. While it is still able to successfully determine the presence of ice once the full profile is input (even with dropped data), earlier predictions seem arbitrary. This goes against the original assumption that a recurrent neural network would be able to continuously update its prediction of ice, gaining confidence as it approaches the surface. Further work is required to determine if the output of an incomplete profile can be usefully interpreted.

Figure 7-1: Predicted probabilities by the RNN model with increasing number of measurements. The model is able to correctly predict the presence of ice once a full profile is input, but further work is required to physically decipher model output from partial profiles.

Just as in sections 5.6 and 6.4, the models are locally tuned by fitting a model
for each region using the respective data. Results are shown in table 7.1. After local tuning, the models are able to predict the presence of ice with a 3.5% misclassification rate over the full dataset. While this is better than predicting on temperature alone as in section 5.6, it is not as good as the multiple logistic regression method presented in section 6.4.

Because of its added complexity, questionable interpretation given a partial profile, and the fact that it does not significantly outperform either of the previous methods, a neural network is not found to be an appropriate replacement for the ISA. Although it was originally thought that a neural network might significantly outperform the other methods warranting the additional complexity, its mediocre increase in performance has shown otherwise. For this reason, the performance plots and a sample implementation are neglected and the neural network method is not explored further.

<table>
<thead>
<tr>
<th>Region</th>
<th>Optimal Probability Threshold</th>
<th>Misclassification Rate</th>
<th>False Negative Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Dataset</td>
<td>0.59</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>Beaufort</td>
<td>0.24</td>
<td>0.045</td>
<td>0.011</td>
<td>0.176</td>
</tr>
<tr>
<td>Arctic</td>
<td>0.033</td>
<td>0.003</td>
<td>0.001</td>
<td>0.056</td>
</tr>
<tr>
<td>Norwegian</td>
<td>0.61</td>
<td>0.015</td>
<td>0.181</td>
<td>0.003</td>
</tr>
<tr>
<td>Greenland</td>
<td>0.51</td>
<td>0.072</td>
<td>0.180</td>
<td>0.032</td>
</tr>
<tr>
<td>Baffin</td>
<td>0.38</td>
<td>0.039</td>
<td>0.031</td>
<td>0.052</td>
</tr>
<tr>
<td>After Regional Tuning</td>
<td>N/A</td>
<td>0.035</td>
<td>0.029</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Table 7.1: Performance of the recurrent neural network model at the misclassification rate minimum for each region.
Chapter 8

Summary, Conclusions, and Future Work

8.1 Summary and Conclusions

In this paper, multiple changes to the ice avoidance algorithm are recommended. This includes recommendations that modify current strategies and the introduction of new methods to increase the performance of the ISA.

In terms of the current ISA, it is shown that the use of a pressure range results in lower performance of the algorithm because the reported values (mean, median, or minimum) represent lagged statistics and should be replaced in favor of point measurements whose precision is verified by past measurements. It is also shown that ice is most predictable nearest the surface and an ISA determination should always be made at the shallowest allowable depth (assumed to be 15dbar).

In terms of new ISA strategies, three potential methods are introduced. Each method generates a predicted probability of ice presence and is fit using historical data. The introduction of a probability allows users to select a probability threshold based on their risk tolerance and specific mission requirements rather than adjusting an arbitrary temperature threshold as current methods dictate. Fitting the models using historical data of ice presence allows for a quantitative approach to each method.

The first method proposed uses temperature as the only predictive variable (sim-
imilar to current ISA methods), and fits the data on the historical profiles. Results are positive with a 4.2% misclassification rate over the full dataset. Models are locally tuned for various regions by fitting only the data from their respective region. Local tuning results in only a 0.1% decrease in the misclassification rate.

The second method introduces new predictive variables and finds the addition of practical salinity and time of year to be better predictors of ice than temperature alone. Results improve upon the first method with a 3.3% misclassification rate after local tuning. This a significant improvement over the performance of the temperature threshold method as it represents a nearly 20% decrease in the misclassification rate. Results also show a 31% decrease in the false negative rate reaching 2.6%. These results suggest the inclusion of the new predictive variables might improve the performance of the ISA.

The final method attempts to implement a recurrent neural network to predict ice presence given the full temperature, salinity, and pressure profiles in addition to the time of year. It is concluded that the added complexity, problems with the interpretations of partial profiles, and insignificant improvement in performance do not warrant its use.

Careful consideration is made to provide a historical dataset that is representative of the true distribution of ice coverage in the Arctic. This requires both spatial normalization to prevent regional bias, and an oversampling strategy to match the ratio of under-ice to ice-free profiles in various regions. Doing so allows for an analysis of the dataset so performance of the ISA can be locally tuned and the expected performance of the algorithms can be observed on a regional level.

With the first two methods, optimal threshold values are reported as the values that minimize the overall misclassification rate. While this is reported as the ‘optimal’ solution, plots and tables are provided that allow users to select their own threshold according to their needs and analyze the expected performance in various regions.
8.2 Future Work

Future work would benefit from refining the data and methods used in this analysis. Namely, the dataset suffers from significant bias caused by the sources of data used. Although accounted for as best as possible, new data or methods could be introduced to create a better dataset from which to predict.

The methods developed in this paper can also be refined. While they introduce new ways to predict ice presence using probabilities and are generally successful at classifying ice presence, they are still fairly naïve. If a better, more representative dataset is presented, new predictive variables may still be introduced to predict ice presence that have not been discussed in this paper.

The use of a neural network may also still prove viable through the work of a more expert researcher in the field. Initial testing showed promise and although a simple RNN could not be developed to significantly improve the performance of the algorithm, further work may yield results.

Finally, the actual implementation of the ISA methods recommended in this paper will show their true effectiveness. While the methods discussed show promising performance on the historical profiles used, the effectiveness of their actual implementation will still need to be evaluated in real world conditions. Only the application of the recommendations made in this analysis will prove their usefulness.
Bibliography


[15] IMS Daily Northern Hemisphere Snow and Ice Analysis at 1km, 4km, and 24km Resolutions. U.S. National Ice Center and National Snow and Ice Data Center.


