The Application of Statistical Learning Techniques to Studying Arctic Sea Ice Survivability

by

Matthew Nathaniel Tooth

B.S., University of Nevada, Reno, 2013

M.S., University of Colorado Boulder, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Aerospace Engineering Sciences

2018

This thesis entitled:
The Application of Statistical Learning Techniques to Studying Arctic Sea Ice Survivability
written by Matthew Nathaniel Tooth
has been approved for the Department of Aerospace Engineering Sciences

_____

Prof. Jeffrey Thayer

_____

Dr. Mark A. Tschudi

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Tooth, Matthew Nathaniel (Ph.D., Aerospace Engineering)

The Application of Statistical Learning Techniques to Studying Arctic Sea Ice Survivability

Thesis directed by Prof. Jeffrey Thayer

Arctic sea ice extent has continued to decline in recent years, and the fractional coverage of multi-year sea ice has decreased significantly during this period. These changes impact the hardiness of the ice pack against future climate forcing, and will affect the future evolution of the Arctic ice cover. These changes will also have substantial effects on Arctic inhabitants, military operations, commercial exploration, and global climate. There exist many sources of remotely sensed Arctic data that can be used to study these changes and determine which predictors account for much of the change in the Arctic ice cover. This thesis assesses the impact of several remotely sensed sea ice parameters on the survival of sea ice in the summer melt season.

A Lagrangian track-based sea ice data product that combines sea ice parcel locations with coincident satellite-derived data is described herein. This database is used in conjunction with several statistical learning classifiers to determine the optimal technique for predicting sea ice extent at the end of the melt season. These statistical learning classifiers are then used to assess which remotely sensed sea ice parameters have the greatest impact on sea ice survival for the pan-Arctic domain. These methods are further combined with airborne data from NASA's Operation IceBridge to investigate sea ice survival in the Beaufort Sea from 2009-2016.

It is shown that sea ice parcel latitude and thickness prior to the onset of melt are the most important variables in estimating parcel survival. As the melt season progresses, broadband albedo becomes the greatest predictor of summer survival. Additionally, downwelling longwave radiation is observed to contribute to melt onset and the triggering of the sea ice albedo feedback in the Beaufort Sea. Coincident airborne ice thickness and snow depth offer less conclusive results, with some years exhibiting higher mean thicknesses and depths in the melted population. The statistical learning techniques described herein are relatively underutilized methods that will prove valuable

in future studies of changing predictor importance in the Arctic.

Dedication

To Dr. Monica Rice

We will always remember your loving heart and brilliant mind

There was a man who sat each day looking out through a narrow vertical opening where a single board had been removed from a wooden fence. Each day a wild ass of the desert passed outside the fence and across the narrow opening - first the nose, then the head, the forelegs, the long brown back, the hindlegs, and lastly the tail.

One day the man leaped to his feet with a light of discovery in his eyes and he shouted for all who could hear him: "It is obvious! The nose causes the tail!"

Frank Herbert, *Heretics of Dune*

Acknowledgements

Projects like these aren't possible without support from family and friends. This list isn't exhaustive, but the contributions from those listed here are unquantifiable:

Paul Tooth, Catherine Dubord, Heather Tooth, Jonathan Tooth, Kimberley Mueller, Kimberly Schoep, the Schoep family, Muir & Rachel Morrison, Warren Smith, Sabrina Cochrane, and Darren Zuro.

I would also like to thank my committee and professional colleagues:

Mark Tschudi, Jeff Thayer, Jim Maslanik, Tomoko Matsuo, Bob Leben, Jennifer Kay, Cora Randall, Mike McGrath, Rory Barton-Grimley, Annie Brookover, and Sharon Anderson.

I have also received direct and indirect support from the following people and organizations:

Lt Bill Coombs, Jim Yungel, John Woods, Sarah Melssen, Steve Hart, the Colorado Center for Astrodynamics Research, NASA, the Office of Naval Research, Coast Guard Station Kodiak, Coast Guard Station Sacramento, the National Snow and Ice Data Center, NOAA Staff at MacDill AFB, the Department of Education Graduate Assistance in Areas of National Need Fellowship, the American Geophysical Union, Melodi & Teddy Rodrigue, Radu Presura, Victor Kantsyrev, Alla Safronova, Tamara Valentine, Marsha Urban, Raytheon, and the University of Colorado Department of Aerospace Engineering Sciences.

Contents

Chapter

Executive Summary **1**

1 Introduction **4**

1.1 Background: Sea Ice in the Arctic 5

1.2 Background: Statistical Learning Techniques 11

1.3 Project Overview 15

2 Lagrangian Tracking of Sea Ice Parcels **18**

2.1 Motivation and Previous Work 18

2.2 Parcel Tracking Overview 19

2.3 Parcel Database Description 20

2.4 Methods 23

2.4.1 Initialization of Field 23

2.4.2 Weekly Data Acquisition 24

2.4.3 Storing Parcel Data 24

2.4.4 Melting and Formation of Parcels 24

2.4.5 Production of Final Database 25

2.4.6 Example Outputs From Database 25

2.5 Publication of Database 29

2.6 Potential Future Expansion of Database 29

2.7 Discussion of Error . . . 30
2.8 Summary . . . 30

3 Applying Statistical Learning to Predicting Arctic Sea Ice Survival 32

3.1 Introduction . . . 32
3.2 Support Vector Machine Approach . . . 35
3.2.1 Methods: Support Vector Machine . . . 37
3.2.2 Results: Support Vector Machine . . . 38
3.3 Binary Decision Tree Classifier . . . 40
3.3.1 Methods: Binary Decision Tree . . . 43
3.3.2 Results: Binary Decision Tree . . . 44
3.4 Random Forest Approach . . . 44
3.4.1 Methods: Random Forest . . . 48
3.4.2 Results: Random Forest . . . 48
3.5 Neural Networks . . . 50
3.5.1 Methods: Neural Network . . . 51
3.5.2 Results: Neural Network . . . 53
3.6 SIPN Example Case . . . 56
3.7 Discussion . . . 58
3.8 Summary . . . 59

4 Relative Importance of Predictors in the Statistical Learning Analysis 61

4.1 Introduction . . . 61
4.2 Predictor Importance in the Random Forest . . . 62
4.2.1 Methods: Predictor Importance in Random Forest . . . 62
4.2.2 Results: Predictor Importance in Random Forest . . . 63
4.3 Predictor Importance in the Neural Network . . . 68
4.3.1 Methods: Predictor Importance in the Neural Network . . . 68

4.3.2 Results: Predictor Importance in the Neural Network 68
4.4 Discussion 74
4.5 Summary 77

5 Study of Beaufort Sea Ice Parcels from 2009-2016 78

5.1 Introduction 79
5.2 Study Area and Data Sources 80
5.2.1 Lagrangian Tracking Product 80
5.2.2 Operation IceBridge Data 81
5.3 Methodology 83
5.4 Results of Analysis of Satellite Data 87
5.4.1 Parcel Ages 87
5.4.2 Ice Surface Temperatures 88
5.4.3 Ice Thickness 88
5.4.4 Mean Latitudes 88
5.4.5 Downwelling Radiative Fluxes 91
5.4.6 Surface Albedos 91
5.4.7 Pearson Correlations between Major Variables and Concentration Change . . 92
5.5 Results for Parcels With Coincident Operation IceBridge Data 94
5.6 Predictor Importance in the Beaufort Sea 97
5.7 Discussion 99
5.8 Summary 101

6 Conclusions 103

6.1 Major Findings 104
6.2 Potential Future Advancement 107

Bibliography **109**

Appendix

A Author's Publications and Presentations **123**

A.1 Publications 123

A.2 Presentations 123

B Acronyms **125**

C Ancillary Data Product Descriptions **128**

C.1 EASE-Grid Sea Ice Motion 128

C.1.1 Product Description 128

C.1.2 Processing Steps 129

C.2 EASE-Grid Sea Ice Age 130

C.2.1 Product Description 130

C.2.2 Processing Steps 130

C.3 SSMI/SSMI-S Sea Ice Concentration Product 131

C.3.1 Product Description 131

C.3.2 Processing Steps 131

C.4 PIOMAS Sea Ice Thicknesses 132

C.4.1 Product Description 132

C.4.2 Processing Steps 133

C.5 MODIS Ice Surface Temperatures 133

C.5.1 Product Description 133

C.5.2 Processing Steps 134

C.6 APP-x Atmospheric Products 135

C.6.1 Product Description 135

C.6.2 Processing Steps . . . 135
C.7 IceBridge Airborne Data . . . 136
C.7.1 Product Description . . . 136
C.7.2 Processing Steps . . . 136
C.8 Sea Ice Convergence Fields from EASE-Grid Sea Ice Motion Data . . . 137

D Map Coordinate Systems **140**
D.1 EASE-Grid 2.0 . . . 140
D.2 Polar Stereographic Grid . . . 140
D.3 Generalized Orthogonal Curvilinear Coordinate Grid System . . . 142
D.4 Converting Between Grid Coordinates . . . 142
D.4.1 Converting Grids With Coordinate Locations . . . 142
D.4.2 Filling Empty Grid Cells After Converting Coordinates . . . 145

E Additional Information on Tracking Product **148**
E.1 History of Product Development . . . 148

F Additional Aiborne Data Sources in the Beaufort Sea **151**
F.1 BESST . . . 151
F.2 CULPIS-X . . . 152
F.2.1 Instrument Package Overview . . . 152

Tables

Table

2.1 Data contained in the Lagrangian tracking database 22

3.1 Number of parcels utilized by classifiers in years 2002-2016 34

3.2 Calendar dates of weeks in a non-leap year. 35

3.3 Binary tree-based classifier week 12 prediction accuracies for validation years 2012-2016 and various training years 44

3.4 Mean random forest classification accuracy for various numbers of trees in the forest 50

3.5 Random forest week 30 prediction accuracies for validation years 2012-2016 and various training years 51

3.6 Neural network peak classification accuracy for various hidden layer sizes 53

3.7 Neural network peak classification accuracy for various numbers of hidden layers of size 10 55

3.8 Neural network week 30 prediction accuracies for validation years 2012-2016 and various training years 56

3.9 Spatial match percentages between random forest predictions and truth for 2008-2017. 58

4.1 2001-2016 pan-Arctic week 20 Gini index rank for predictors in the random forest classifier 66

4.2 2001-2016 pan-Arctic week 30 Gini index rank for predictors in the random forest classifier 66

4.3 Table of variables included in random forest test cases and their mean 2002-2016 accuracies. 1 indicates that the variable was included in a test, while 0 indicates that it was excluded. 67

4.4 OFAT change in accuracy when removing specified variable during neural network prediction in week 20 72

4.5 OFAT change in accuracy when removing specified variable during neural network prediction in week 30 72

4.6 Table of variables included in neural network test cases and their mean 2002-2016 accuracies. 1 indicates that the variable was included in a test, while 0 indicates that it was excluded. 73

5.1 Number of melted and surviving parcels analyzed for each year of study. 84

5.2 Weeks of IceBridge data for years 2009–2016. 85

5.3 Pearson correlations between studied variables and total change in concentration for 2009–2016. Values with $p \in (0.05, 0.3]$ are shown in blue. Values with $p \in (0.3, 1]$ are shown in orange 92

5.4 Mean Operation IceBridge and satellite derived spring surface conditions of melted (**red**) and surviving (**black**) parcels 96

5.5 Week 20 predictor importance in the Beaufort Sea 97

D.1 Basemap instance parameters for grid conversions 145

Figures

Figure

1.1 Arctic sea ice extent for 2012-2017 with 1981-2010 interquartile & interdecile ranges [1] 9

1.2 (a) Sea ice age coverage by year 1984-2017. Sea ice age coverage maps for March 1984 (b) and 2017 (c) [129] 10

1.3 Map of some major statistical learning techniques, and the decisions that lead to choosing a technique from the *SciKit Learn* documentation. 13

2.1 Tracked sea ice characteristics from [163] (Figures 5 and 6) 19

2.2 Example of sea ice parcel tracks derived from the Lagrangian tracking database. . . 21

2.3 Procedural flow for generating a year of parcel tracks with ancillary data. 23

2.4 Selected example parcel drift tracks from 2014 with IST and concentration for the parcel indicated by an asterisk in the left panel. 26

2.5 Beaufort Sea example region with associated parcel temperatures for 2014. Red scatter points indicate parcels that melted, while blue points indicate parcels that survived to the end of the year. 27

2.6 Tracks for four parcels from 2001-2004 with associated ancillary data. 28

3.1 Broad categories of statistical learning with some examples from subcategories shown. 36

3.2 Linear support vector separation (left), Radial SVM separation (center), and Polynomial of degree 4 separation (right) [54] 37

3.3 2010 SVM accuracy with varying training portions and a radial basis function kernel 39

3.4 2010 SVM accuracy with varying kernel functions 39
3.5 Classification tree grown for week 12 of 2016 with terminal nodes labeled as 0 for melted and 1 for survived. 41
3.6 Example classification splits for week 12 of 2016. X1 corresponds to latitude, while X3 represents PIOMAS thickness. 42
3.7 Mean binary tree-based classifier accuracy with varying training portions 45
3.8 Two example trees with associated decision regions (left, center), and a forest ensemble with example member decision regions (right) [52] 47
3.9 2002-2016 mean random forest classification accuracy with varying training portions. Dashed lines indicate one standard deviation away from the main curve. 49
3.10 Two-layer pattern recognition neural network representation created using a Matlab network training tool 52
3.11 Mean pattern (1) and fitting (2) neural network accuracies with varying training portions 54
3.12 Predictions of September sea ice extent in 2017 made during (A) June, (B) July, (C) August. The true extent is shown in (D). 57

4.1 Weekly mean predictor importance in the random forest for years 2002-2016 65
4.2 Weekly mean predictor importance in the neural network for years 2002-2016 71
4.3 Mean PIOMAS thickness for parcels of different age classes 76

5.1 Beaufort Sea study region shaded in black. 82
5.2 Example plots of four studied variables for melted (**red**) and surviving (**black**) parcels during 2009. Dotted lines indicate one standard deviation away from the main curve. 85
5.3 Operation IceBridge flight track sections utilized for 2009–2016. 86
5.4 Percent of population in age categories 1–5+ years for surviving (**black**) and melted (**red**) ice parcels in the Beaufort Sea during 2009–2016. 89

5.5 Ice surface temperature trends for melted (**red**) and surviving (**black**) parcels during 2009–2016. 89

5.6 PIOMAS model ice thickness of melted (**red**) and surviving (**black**) sea ice parcels during 2009–2016. 90

5.7 Mean latitudes of sea ice parcel populations during 2009–2016. Melted parcels are shown in **red**, surviving parcels are shown in **black**. 90

5.8 Mean downwelling shortwave radiation in W/m^2 for surviving (**black**) and melted (**red**) parcels during 2009–2016. 91

5.9 Mean downwelling longwave radiation in W/m^2 for surviving (**black**) and melted (**red**) parcels during 2009-2016 93

5.10 Mean albedo for melted (**red**) and surviving (**black**) parcels during 2009–2016. . . . 93

5.11 Weekly mean predictor importance in the Beaufort Sea 98

C.1 Arctic regions used for IceBridge data 138

C.2 IceBridge flight track from April 5^{th}, 2009 with corresponding snow thickness data. . 138

C.3 EASE-Grid cell with vector components used to compute convergence values shown 139

D.1 EASE-Grid 2.0 map of the Arctic, generated using the Python Basemap package . . 141

D.2 Polar Stereographic map of the Arctic, generated using Basemap. 143

D.3 Representation of the PIOMAS GOCC grid 143

D.4 PIOMAS model coordinates plotted on EASE-Grid 144

D.5 Polar Stereographic coordinates plotted on EASE-Grid 144

D.6 Un-filled sea ice concentration field for August 30^{th}, 2014 146

D.7 Filled sea ice concentration field for August 30^{th}, 2014 147

F.1 Assembled BESST instrument [35] 153

F.2 CULPIS-X Package installed in a flare tube prior to a June 2016 Arctic Flight . . . 153

F.3 CULPIS-X flight track with coincident ice concentration data from June 15^{th}, 2016 . 154

F.4 Visible image of ocean surface from CULPIS-X (left), and k-means clustered image with three clusters (right) 155

Executive Summary

This thesis focuses on using statistical learning techniques to assess which factors most influence the survival of sea ice during the summer melt season. Arctic sea ice has been declining in extent over the past several decades, and the areal share of older ice has also declined during this period. This decline will impact plans for future resource extraction, military operations, native communities, and the global climate as a whole. While there have been many field expeditions to obtain in-situ data in the Arctic, the large size of the Arctic domain and coverage of field experiments have prevented these observations from capturing larger scale variation. Remotely sensed measurements from satellites offer a means through which researchers, policy makers, and analysts can view Arctic change over time.

Remote sensing of sea ice has been growing since the earliest satellite images were obtained of sea ice in 1964. Observations from satellite-based sensors have provided time histories of numerous surface characteristics of the Arctic ice cover. These data suggest that the Arctic is undergoing rapid and drastic changes that will have lasting impacts on the future of ice in the Arctic. Determining which parameters most strongly influence sea ice survival will help scientists focus future modeling and remote sensing efforts.

This thesis investigates which remotely sensed sea ice parameters most influence sea ice survival during the summer melt season. This is accomplished through the development of a new Lagrangian track-based sea ice data product and the use of several statistical learning techniques. These statistical learning techniques are used to estimate which sea ice "parcels" survive a given summer melt season. The importance of predictors in determining which parcels melt is also de-

rived from these methods. Lastly, the major factors influencing the survival of sea ice parcels in the Beaufort Sea are investigated using these statistical learning techniques and coincident airborne data from NASA's Operation IceBridge.

The data product developed in this thesis combines sea ice motion vectors with ancillary satellite data products to form a database that contains weekly sea ice parcel positions and their co-located data. The product is available from 2001-2016 on a 12.5km EASE-Grid through the Pangaea Data Publisher. The co-located data tracked for each parcel include ice thickness, ice surface temperature, ice concentration, convergence values, downwelling longwave/shortwave radiation, albedo, and age. Each parcel is also issued a unique ID that allows it to be tracked between years. The process through which this database is generated is described in this thesis, along with background information on each ancillary product. Further discussion of how each product is converted to the 12.5km EASE-Grid is also included in an appendix. This database serves as the foundation of the rest of the thesis, which takes the approach of discovery from data.

Four statistical learning techniques are explored in this thesis for short term prediction of summer ice survival: support vector machines, binary tree-based classification, random forests, and neural networks. These techniques were tested by splitting data from the Lagrangian tracking database into training and testing sets. Splitting the data in this manner allowed for training and testing to be performed using the same overall set of tracks during this study. We observe that the random forest and neural network yield the greatest short term prediction accuracies: 96% and 94% respectively. We also determine that the optimal random forest design includes 100 trees, while the optimal neural network is made up of two ten-node layers. It is also determined that when predicting summer sea ice survival, the inclusion of more data during training serves to increase the prediction accuracy. This implies that short term predictive work would benefit best from using random forests and neural networks with at least five to ten years of training data.

The random forest and neural network are used to determine the relative importance of seven predictors from the Lagrangian tracking database: latitude, ice surface temperature, ice thickness, albedo, downwelling longwave/shortwave radiation, and age. These predictors are ranked using

one-factor-at-a-time removal in the neural network, and the Gini index in the random forest. The results from this study suggest that latitude and ice thickness during the spring are the most important predictors in determining which parcels survive summer melt. During the summer melt season the Gini index-derived importance of albedo grows, and is near those of latitude and thickness. Seasonal changes in relative predictor importance are more apparent in the random forest, which indicate that it is more suited for this type of analysis. As a result, the random forest was chosen to be used in the Beaufort Sea ice survival case study.

Lagrangian track data and coincident airborne data from Operation IceBridge are compared for parcels in the Beaufort Sea from 2009-2016. In addition, a random forest is used to determine weekly predictor importance values for these data as a point of comparison against more traditional statistical techniques. The results of this case study suggest that initial latitude and thickness of parcels prior to melt onset best separate melted and surviving parcels during the summer melt season. This agrees with results from the random forest, which also implicates these variables during melt onset. Change in albedo most negatively correlates with change in sea ice concentration in these data, which underscores the importance of the sea ice albedo feedback during summer melt. Parcels with IceBridge data offer mixed results, with melted parcels often having greater ice thicknesses and snow depths than their surviving counterparts.

The development of a Lagrangian tracking database and statistical learning techniques to study it has opened up new avenues for sea ice research. It is the hope of the author that the database described herein will continue to be applied to future sea ice research that can benefit from studying ice parcel histories. In addition, the statistical learning techniques will potentially improve short term sea ice forecasting while offering readily available predictor importance values that can be used to direct future studies.

Chapter 1

Introduction

The energy balance of Arctic sea ice is governed by an interconnected budget that is driven by several competing feedbacks. The exploration of these feedbacks and the parameters that govern them have been identified as grand challenges in the cryospheric sciences. This thesis focuses on the use of discriminative classifiers to assess which remotely sensed sea ice parameters most impact sea ice survival during the summer melt season. It does this through the application of statistical learning techniques that are not utilized in a widespread manner in cryospheric studies.

During this research project, the use of discriminative classifiers in short term prediction of summer sea ice extent has been investigated. These techniques were applied to assessing which remotely sensed parameters had the greatest impact on sea ice survival during the Arctic melt season. A case study in the Beaufort Sea is also explored, with a focus on the major characteristics of parcels that do and do not survive summer melt. In this case study the results from a random forest are compared against more traditional statistical methods. These techniques demonstrate new ways in which relationships between sea ice characteristics and survival can be investigated.

1.1 Background: Sea Ice in the Arctic

Arctic sea ice has been recognized as an important component of the global climate system, and has been well studied throughout the modern era [5, 59, 77, 174, 175]. Changes in ice cover throughout the Arctic Basin are expected to have a lasting impact on both the Arctic and global climate [152, 180]. Seasonal and annual changes in the ice pack influence shipping efforts, military operations, resource extraction, local weather, and biological processes in the Arctic [33,37,103,104]. Recent changes during the summer melt season have been the focus of many research efforts, as the Arctic ice cover continues to decline in extent, age, and thickness (Figures 1.1 & 1.2) [84, 91, 129]. Significant effort has been put forth to better understand and predict changes in ice extent, ice properties, and characteristics of the summer melt season [144].

The melt of Arctic sea ice during the summer is controlled by a large set of properties and processes that serve to input energy into an individual sea ice floe. The net surface flux in the Arctic can be represented by:

$$F_{sfc} = -(R_{sfc} + Q_H + Q_E) \tag{1.1}$$

where R_{sfc} is the net radiation at the surface while Q_H & Q_E are turbulent sensible (H) and latent (E) heat fluxes that are driven by conduction and turbulent eddies in the atmosphere. In this convention, F_{sfc} is positive upward while R_{sfc}, Q_H, and Q_E are positive downward. The net surface flux can be further partitioned into three major terms:

$$F_{sfc} = -(M + C + B) \tag{1.2}$$

where M represents melted ice, C represents conduction of heat in/out of the underlying column, and B represents bulk absorption in the column.

The R_{sfc} term is defined as the sum of the net shortwave (SW) and longwave (LW) radiation components at the surface:

$$R_{sfc} = SW_{down}(1 - \alpha) + LW_{down} - \epsilon\sigma T_s^4 \tag{1.3}$$

where α represents the surface albedo, ϵ represents the surface emissivity, and T_s represents the surface temperature. The quantity $\epsilon\sigma T_s^4$ is the upwelling longwave term LW_{up}. During the Arctic winter when there is no direct sunlight the SW_{down} term is effectively zero. During the summer the downwelling shortwave term is an appreciable source of energy input at the surface due to seasonal changes in albedo, the presence of clouds, and solar zenith angle [16, 137, 178]. Albedos over dry snow during the spring can be 0.7-0.8, while accumulation of melt ponds and exposure of bare ice can degrade the albedo to 0.4-0.5 [6, 76, 122]. Albedo can be further impacted by surface morphology [16]. The emissivity of sea ice and snow are relatively constant near 0.98 [128]. The net longwave portion of R_{sfc} is typically negative, but it can be positive in the presence of low, warm cloud cover [137].

When R_{sfc} is positive there is energy input into the surface. This energy input serves to melt snow and ice at the surface, and can contribute to the bulk absorption term B in open water areas. This causes the formation of melt ponds during summer melt, which further contribute to melting the ice cover [41, 134]. Areas of open water that form near ice allow for absorption of energy at the ocean surface that is conducted into the ice cover through bottom and lateral melt [161]. This creates more open water areas that contribute to the sea ice albedo feedback [27, 121]. Properties and processes such as surface temperature, albedo, ice thickness, and downwelling longwave/shortwave radiation, along with parcel location, can contribute to the parcel's potential survival or extinction during summer melt. It is valuable to obtain data describing these collective sea ice parameters so that their impact on sea ice survival can be assessed.

Given the large areal coverage of Arctic sea ice, it is desirable to obtain data from satellite-based remote sensing. Such observations can be further augmented by airborne remote sensing and in-situ observations from a variety of platforms and missions. The earliest images of sea ice from satellites were obtained in 1964 via the Nimbus 1 satellite [102]. The next major milestone in remote sensing came with the launch of Nimbus-7 and the scanning multichannel microwave radiometer (SMMR) in 1978. This was followed by the SSM/I and SSMIS sensors, which have formed a 35-year record of daily global sea ice observations [165]. The advantage of these passive

microwave sensors is that they can sense emission in the presence of clouds and at night. These observations provide a long-term record of sea ice extent and concentration from which studies of changes in the Arctic climate system and their impacts on sea ice can be observed.

Additional satellite sensors have been launched throughout the past several decades that have helped gather Arctic data, such as the Advanced Very High Resolution Radiometer (AVHRR), the Moderate Resolution Imaging Spectroradiometer (MODIS), and the Visible Infrared Imaging Radiometer Suite (VIIRS). These instruments gather optical and infrared data that have been used to derive ice concentration [32,36], melt pond fraction [95,173], surface temperature [49,71], and other sea ice properties. Observations of concentration and melt pond fraction have proven valuable, as they have aided in determining concentration from microwave measurements [23, 24]. Data from these and other instruments have served as inputs to models that provide estimates of other Arctic surface properties and processes [75].

During the 1980s and 1990s it was recognized from satellite data that the Arctic was undergoing changes in ice composition, age, and thickness [20, 46, 100, 119]. Arctic ice extent in the summer was found to have declined -0.6%/ each year from 1978-1995, with an overall loss of 6% of area that was previously occupied by ice [100]. Multi-year ice was found to have reduced by 14% in winter areal coverage between 1978 and 1998 [62]. Change in ice age was also found to be correlated with substantial changes in ice thickness [62]. It was suggested that these changes in the ice cover would lead to different ice regimes in the Arctic that would have lasting impacts on heat and momentum exchanges in the region.

Arctic sea ice continued to decline in age and extent throughout the 2000s and 2010s [22, 99, 114, 120, 132, 148], with ten of the lowest record sea ice extent years occurring between 2006 and 2017 (Figure 1.1) [120]. In 2017 the September monthly average trend in sea ice extent for the Arctic had reached -13% per decade relative to the 1981-2010 average, with a -2.7% per decade trend in March [129]. This loss of extent has coincided with a significant reduction in the areal fraction of older sea ice in the Arctic (Figure 1.2) [99,129]. The extent of ice older than four years of age has decreased from 2.54 million km^2 to 0.13 million km^2 from March 1985 to March 2017 [129].

The loss of perennial sea ice further weakens the ice cover against future warming, as older ice is typically thicker and more likely to survive summer melt [170].

The Arctic climate system is driven by a large number of feedbacks that range from strongly positive to strongly negative [56, 68, 96, 115, 117, 126]. While remotely sensed data cannot capture all of the parameters that influence these feedbacks and the melt of Arctic sea ice, they can offer insights into surface changes that can aid in further development and testing of models. The change in the surface energy balance required to explain the loss of ice cover over the past several decades is only $1 W m^{-2}$ [89], so the attribution of this additional heat is meaningful [56]. Future Arctic research and modeling efforts will benefit from additional remotely sensed sources of Arctic data, and from additional tools through which these data can be analyzed. Identifying the major causes of sea ice loss in the Arctic has been identified as a "grand challenge" of climate science [56, 67]. Therefore, further exploration of the parameters that drive the Arctic energy budget and its associated feedbacks is warranted. The investigation of the impact of changing surface processes, atmospheric processes, and sea ice properties on sea ice survivability can be aided through the novel application of tools from the field of statistical (machine) learning. These tools can aid in both short term prediction for efforts like the Sea Ice Prediction Network and further inference of predictor importance from remotely sensed data.

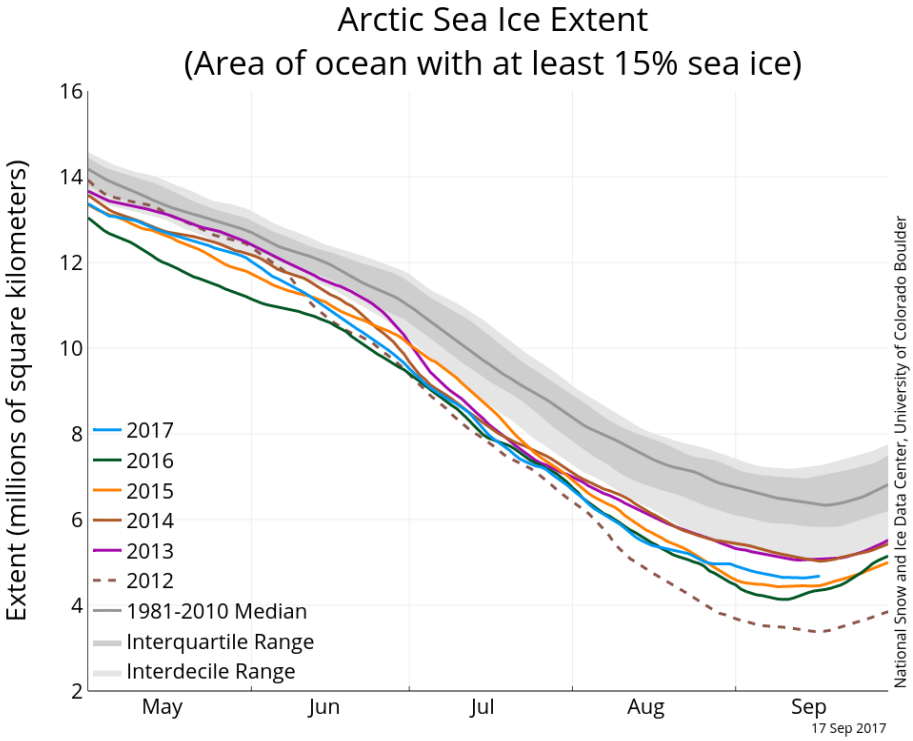

Figure 1.1: Arctic sea ice extent for 2012-2017 with 1981-2010 interquartile & interdecile ranges [1]

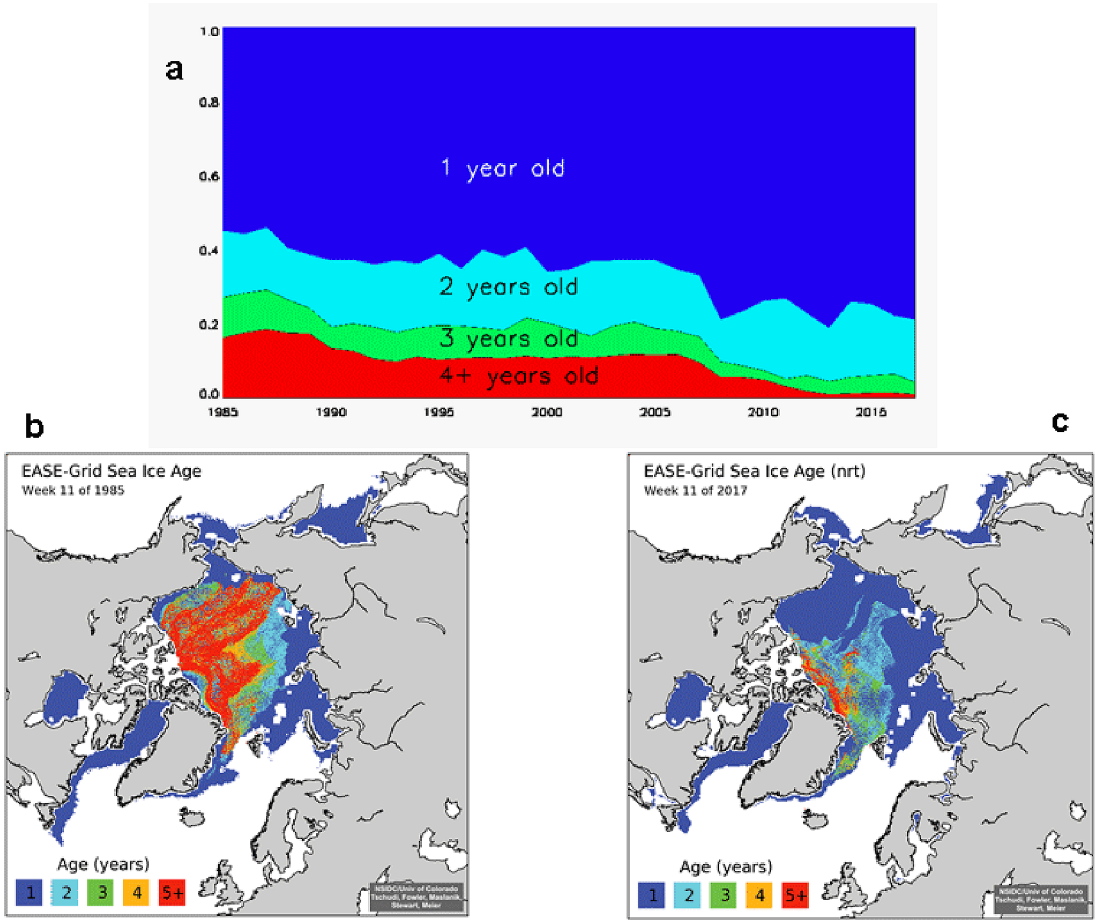

Figure 1.2: (a) Sea ice age coverage by year 1984-2017. Sea ice age coverage maps for March 1984 (b) and 2017 (c) [129]

1.2 Background: Statistical Learning Techniques

In this dissertation the effectiveness of several statistical learning techniques in classifying sea ice parcels into melted and survived categories is assessed. Further inference of the importance of individual predictor variables is also studied. These techniques are leveraged to create test cases in which predictor importance can be compared against more traditional statistical techniques for studying Arctic sea ice parcels.

The field of statistical learning is made up of a broad set of tools that are used in understanding data. These tools generally fit into two major categories: supervised and unsupervised learning. In the supervised case, the data are composed of observations $x_i, i = 1, ..., n$ and associated responses $y_i, i = 1, ..., n$ that can be utilized in training an algorithm to generate responses from input observations. The algorithm seeks to form the relationship:

$$Y = f(X) + \epsilon \tag{1.4}$$

where f represents the systematic information that X provides about Y, and ϵ represents the error terms. Typically this relationship f is used either in prediction or inference. Prediction seeks to generate an f that is used to generate predictions $Y^{'}$ from new inputs $X^{'}$, while inference seeks to determine how Y is affected as the X values change. Examples of supervised learning include support vector machines (SVM), neural networks, and forest/tree methods. In the unsupervised case, the observations $x_i, i = 1, ..., n$ do not have associated responses, and the algorithms instead seek to describe relationships between the variables or between observations. For example, in the case of clustering the algorithm seeks to determine whether X contains relatively distinct groups that can be formed using differences between the data points in X. The results provide a means through which the difference between clusters can be inferred with respect to some property of interest in X. Examples of these types of statistical learning techniques include k-means clustering, k-nearest neighbors, and fuzzy clustering.

The beginning of statistical learning could be considered to be the point at which Legendre and Gauss first published papers on the subject of least squares, which was an early form of lin-

ear regression [140, 142]. This approach was employed primarily in predicting quantitative values, but was not applied to qualitative values. The prediction of qualitative values was addressed by Fisher's Linear Discriminant Analysis (LDA) technique in 1936 [39]. The alternative approach of logistic regression was further developed throughout the 1940's [7, 40, 94]. In the early 1970's the term *generalized linear models* was coined by Nelder and Wedderburn to cover statistical learning methods that include both linear and logistic regression as special cases [113].

Until the 1980's computing technology could not fit non-linear relationships due to computational restraints. During the 1980's technology improved to the extent that classification and regression trees were possible. This improvement in computational power led to the introduction of these methods by Breiman, Friedman, Stone, and Olshen [13]. Their work was among the first to demonstrate the implementation of cross-validation for model selection. Work by Hastie and Tibshirani in 1986 further coined the term "generalized additive models" to encompass non-linear extensions to generalized linear models [53].

Access to statistical learning methods has increased with the advent of more powerful personal and research computers. Major programming languages such as Matlab, Python, and R are provided with toolboxes that can perform many statistical learning tasks that were unfeasible a few decades ago. The modern toolset of statistical learning is made up of a variety of techniques for analyzing data based on the quantity and type of data available (Figure 1.3).

Techniques that classify data into one or more categories are valuable for remote sensing and scientific applications [3, 8, 45, 110]. Much of the current Arctic research focuses on classification of surface types from remote sensing data using statistical learning classifiers [4,29,51,90,111,116,138]. There has been some exploration of using statistical learning techniques for prediction of sea ice extent [21], but there exist many areas in which this research can be expanded and improved upon. These techniques have not been applied to Lagrangian tracking algorithms, which offers a new avenue from which melt season states can be investigated. Additionally, the effectiveness of short term prediction through the use of various classifiers has not been explored. The use of ranking predictors through methods like the Gini index [54, 61] also provides a valuable tool that has been

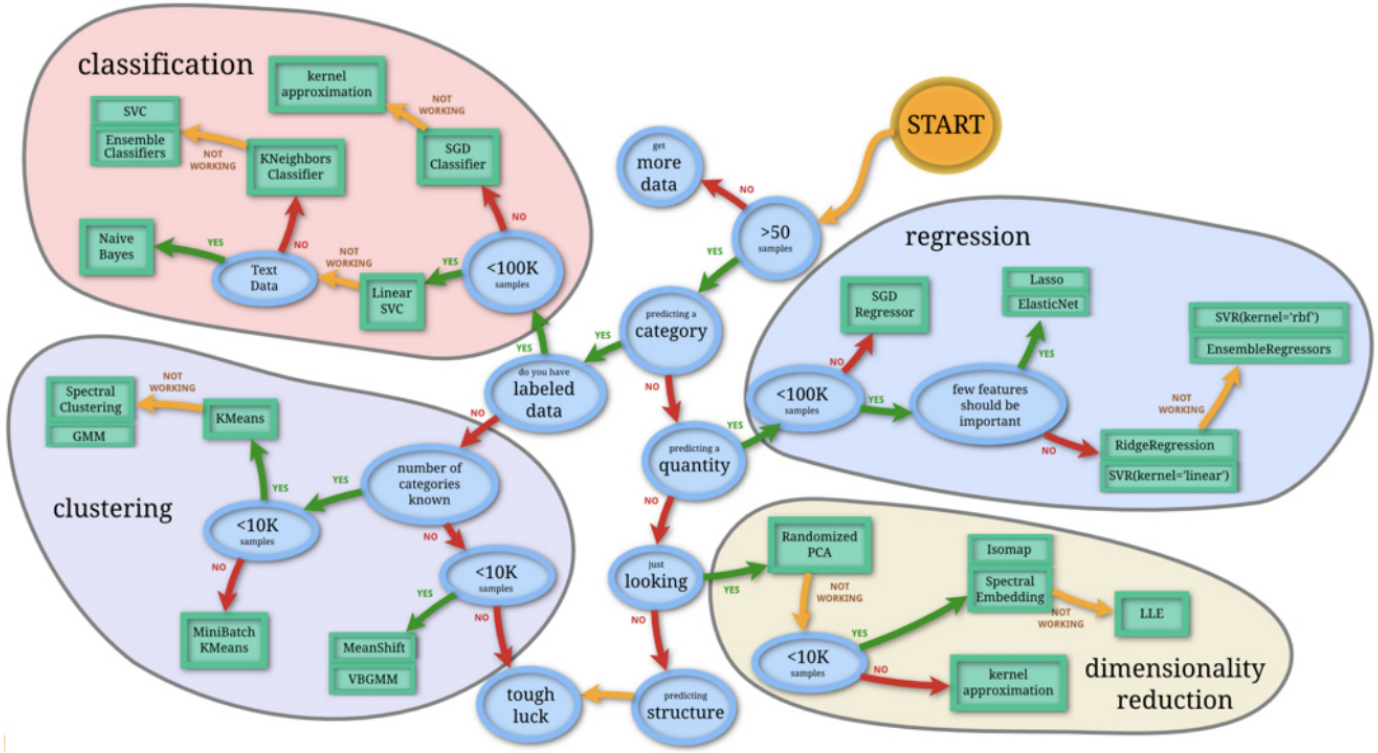

Figure 1.3: Map of some major statistical learning techniques, and the decisions that lead to choosing a technique from the *SciKit Learn* documentation.

utilized successfully in other fields [10].

1.3 Project Overview

Initial investigation into the Arctic studies discussed in Section 1.1 and the statistical learning techniques discussed in Section 1.2 revealed several areas of research that could be explored through the novel application of specialized statistical learning techniques to the investigation of Arctic sea ice. **This investigation assesses the impact of remotely sensed sea ice parameters on the survival of sea ice in the summer melt season**. To address this, the following science questions are pursued:

(1) **Which sets of remotely sensed data are both available and appropriate for studying sea ice survival in the Arctic?**

There are numerous satellite-based data products that provide records of different ice surface properties and sources of energy input available for use in Arctic studies. Lagrangian tracking can serve to combine these measurements with positions of parcels throughout a given year to form a combined data product that includes a representative set of data for use in the study of sea ice survivability. The dataset described herein contains a variety of properties, parameters, and processes that influence sea ice in the Arctic. The data product spans 2001-2016, and was created in order to combine ice surface temperature (IST), modeled ice thickness, ice concentration, ice age, convergence, downwelling shortwave/longwave, and albedo data with Lagrangian tracks of sea ice parcels. The generation of this product and the resulting database is accomplished in Chapter 2, and further information on the data products is included in Appendix C.

(2) **Does a particular statistical learning technique perform best in short-term prediction of sea ice survival during the summer melt season?**

Many prior studies of sea ice that utilized statistical learning techniques focus on surface classification, but the techniques used in those studies can also be utilized for classifying sea ice parcels into melted and surviving categories to form predictions. Each classifier's resulting prediction accuracy can be influenced both by the structure of the data and by

the parameters chosen to optimize the classifier. Statistical learning classifiers are broadly separated into generative and discriminative categories. Generative classifiers seek to use assumptions about underlying distributions in datasets to classify points after training. The results from these classifiers are more robust in the presence of outliers, are more interpretable, and are capable of accounting more for the underlying uncertainties in the data used during training and testing. Discriminative classifiers instead seek to learn decision boundaries directly from the data used in training, and are more likely to yield an accurate decision boundary after training as compared to generative classifiers. Four discriminative classifiers were chosen in order to determine the optimal statistical learning technique to use for short-term prediction of sea ice survival. The prediction accuracy and optimal parameters for support vector machines, binary tree-based classification, random forests, and neural networks are obtained in Chapter 3. The two most accurate techniques were a random forest and neural network. These techniques were further utilized to study sea ice survivability throughout the rest of this thesis.

(3) **How does the importance of predictors change during summer melt, and do different statistical learning techniques report different results?**

The melt of Arctic sea ice is driven by a large number of factors that have associated feedbacks. Statistical learning techniques and predictor importance derived from these factors can provide additional insight into their changing importance in determining which sea ice parcels will survive a given melt season. Measures of importance, derived using the random forest Gini index and neural network OFAT analysis, are compared for both the start and end of the melt season in Chapter 4. The changing importance of predictors between these two points of comparison are also discussed in the context of changes in the ice cover during the melt season.

(4) **Which predictors of sea ice survival in the Beaufort Sea during summer are ranked as most important, and how do these results compare to results from**

other statistical techniques?

The Beaufort Sea has acted as a sink for sea ice over the past decade, as it has accounted for a significant portion of the total areal loss of sea ice each year [43, 57]. The region is also the extinction location for many of the multi-year ice floes that advect from the Canadian Archipelago [44, 83]. The Lagrangian tracking database described in Chapter 2 is utilized to investigate differences between melted and surviving populations of sea ice parcels in the Beaufort Sea from 2009 to 2016 in Chapter 5. Additional airborne data from NASA's Operation IceBridge are also considered, along with correlations between major satellite-derived data and change in sea ice concentration. The statistical learning techniques developed in Chapters 3 & 4 of this thesis are also applied to the data and compared against the initial results of the case study to assess the performance of these classifiers against more traditional techniques.

This research is divided into six chapters. Chapter 1 has provided background on Arctic sea ice and statistical learning, the major motivation of this work, and the research questions selected to address the motivating hypotheses. Chapter 2 describes a Lagrangian tracking sea ice parcel database, its contents, and how it is produced. This chapter serves to introduce the data product that is utilized throughout the rest of the study. Chapter 3 contains a discussion of four major discriminative classifiers used in prediction, the optimization of these classifiers, and their prediction accuracy. Chapter 4 is a discussion of predictor importance for Arctic sea ice parcels from 2002-2016 as derived from the statistical learning techniques discussed in Chapter 3. Chapter 5 presents an analysis of Beaufort sea ice parcels from 2009-2016 using the Lagrangian tracking database described in Chapter 2 and coincident NASA Operation IceBridge observations. This chapter also presents results from applying the statistical learning techniques discussed in Chapter 3 & 4 to the product described in Chapter 2, and compares them with the results of the initial case study. This document is then concluded with a summary of the research results, major findings, and future research possibilities in Chapter 6.

Chapter 2

Lagrangian Tracking of Sea Ice Parcels

This chapter provides an overview of a Lagrangian tracking database of sea ice parcel positions and associated satellite data. The database was developed at the University of Colorado by M. Tooth and M. Tschudi, and is further discussed in [158]. The majority of the chapter details the methods behind the creation of the data product, and provides an overview of its contents. The data products included in the Lagrangian tracking database are further discussed in Appendix C.

This chapter is adapted from *Tooth M. and Tschudi M., 2017* [158].

2.1 Motivation and Previous Work

The near-continual movement of most sea ice parcels in the Arctic exposes them to different regimes and climate conditions throughout a given year. This creates a need to better quantify how the ice responds to the variable conditions it encounters in order to improve forecasts and to better understand changing ice conditions in specific regions. Lagrangian tracking of key variables that influence the survivability and conditions of the sea ice will help quantify the effects of these changing regimes on ice conditions. This is particularly of interest in regions such as the Beaufort Sea, which has become a sink for sea ice in the summer [79, 127, 145], and is of economic and political interest [33].

Early ancillary data product tracking was performed by Chuck Fowler for use in Tschudi et al (Figure 2.1) [163]. There was significant merit in continuing to produce these tracks and ancillary data files, which led to interest in continuing this work through research led by M. Tschudi.

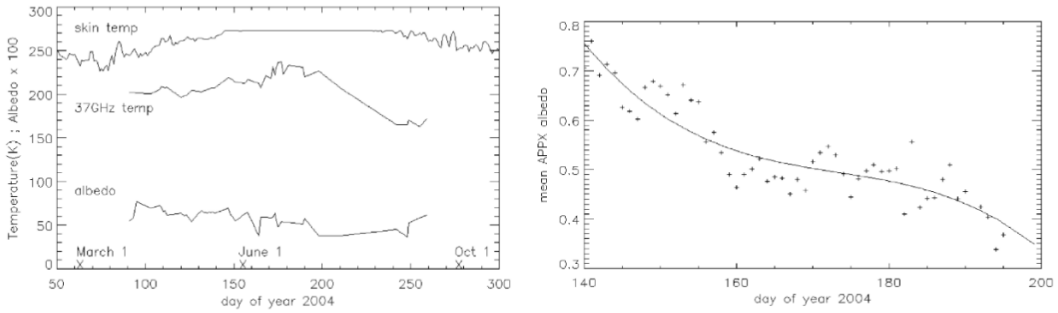

Figure 2.1: Tracked sea ice characteristics from [163] (Figures 5 and 6)

Initial work on this project focused on sea ice "gate" codes that were developed by J. Maslanik at the University of Colorado Boulder. These codes used sea ice motion fields to estimate the flux of sea ice across user-defined boundaries in the Arctic in order to derive estimates of the inflow and outflow of ice. Interest expressed by Jinlun Zhang at the University of Washington led to the development of code for tracking parcel concentrations, positions, and thicknesses that was eventually developed into a fully fledged tracking program under the direction of M. Tschudi. This code has used to generate the database described in this chapter. Further information on the development history of the tracking product code is provided in Appendix E.

2.2 Parcel Tracking Overview

The result of this work was the production of a sea ice parcel tracking database that combines ancillary thermodynamic and dynamic variables with sea ice parcel positions. The database contains individually identified sea ice parcels and their locations on the EASE-Grid with weekly ancillary data values when they are available. These tracks provide a means through which changes in parcel features and ancillary data can be tracked through time for both individual parcels and regions of interest. An example motion field that demonstrates the successful storage of parcel tracks is shown in Figure 2.2. Motion patterns such as the Trans-polar Drift Stream and Beaufort Gyre are present in the motion field.

2.3 Parcel Database Description

The variables included in the parcel tracking database are summarized in Table 1. Ancillary data products include EASE-Grid Sea Ice Motion Vectors [165], MODIS Terra ice surface temperature (IST) [50], SSM/I & SSMIS ice concentration [19], PIOMAS ice thickness [136], EASE-Grid convergence values in the u & v directions, APP-X all-sky broadband albedo, APP-X all-sky shortwave up/downwelling radiation, APP-X all-sky longwave up/downwelling radiation [75], and EASE-Grid Sea Ice Yearly Age values [164]. Further information on each of these data products is provided in Appendix C. Weekly values of each variable are stored for individually tracked parcels on a 12.5km EASE-Grid. Data that are available on daily timescales are averaged for each week, and data that are not provided on EASE-Grid are converted to a 12.5km EASE-Grid as described in Appendix D.

The data are archived through the Pangaea Data Publisher, and are separated by year from the start of 2001 to the end of 2016 [156]. Each year's data have been separated into individual files containing parcel IDs, grid positions, and variable data to keep file sizes below 100MB. Further information on how to read the data files and example Python 2.7 codes are available in the product's Pangaea archive [156]. Fill values of 9999 and 9999.0 have been used for missing data, and a fill value of 9998.0 has been used for the pole-hole present in some products.

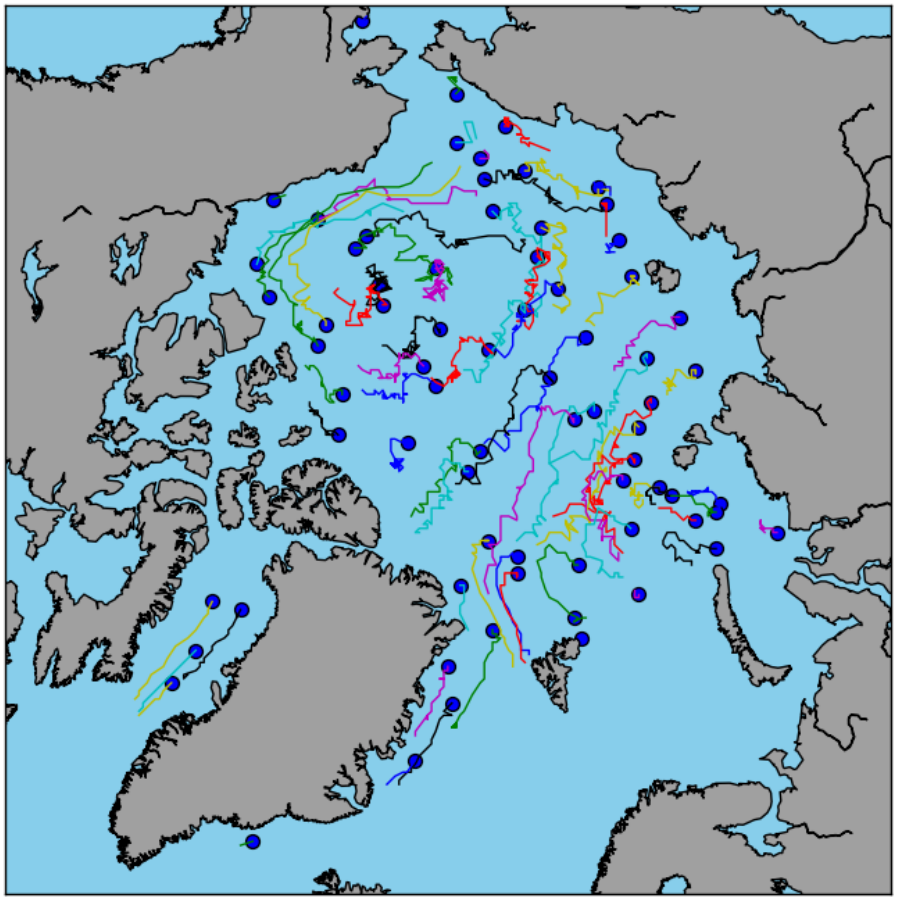

Figure 2.2: Example of sea ice parcel tracks derived from the Lagrangian tracking database.

Table 2.1: Data contained in the Lagrangian tracking database

Variable	**Type**	**Units**
Parcel ID	Int	N/A
Parcel Age	Int	Weeks
EASE-Grid Ice Age Product	Int	Years
u Position	Int	12.5km EASE-Grid Cell
v Position	Int	12.5km EASE-Grid Cell
U Vector	Float	cm/s
V Vector	Float	cm/s
Ice Surface Temperature	Float	Kelvin
Ice Concentration	Float	Percentage
Model Ice Thickness	Float	Meters
Convergence in V Direction	Float	cm/s
Convergence in U Direction	Float	cm/s
Albedo, Morning	Float	N/A
Albedo, Afternoon	Float	N/A
Shortwave Downwelling Radiation, Morning	Float	W/m^2
Shortwave Downwelling Radiation, Afternoon	Float	W/m^2
Shortwave Upwelling Radiation, Morning	Float	W/m^2
Shortwave Upwelling Radiation, Afternoon	Float	W/m^2
Longwave Downwelling Radiation, Morning	Float	W/m^2
Longwave Downwelling Radiation, Afternoon	Float	W/m^2
Longwave Upwelling Radiation, Morning	Float	W/m^2
Longwave Upwelling Radiation, Afternoon	Float	W/m^2

2.4 Methods

The parcel database is generated using weekly mean motion vectors available from the National Snow and Ice Data Center (NSIDC) [165]. An initial field is sampled for the first week of January, and then weekly changes in the position of parcels are stored along with available ancillary data for the parcel's location for the rest of the year through December 31st. Newly frozen parcels are also accounted for, along with parcels that have melted during a weekly time step. The main flow of the program is shown in Figure 2.3.

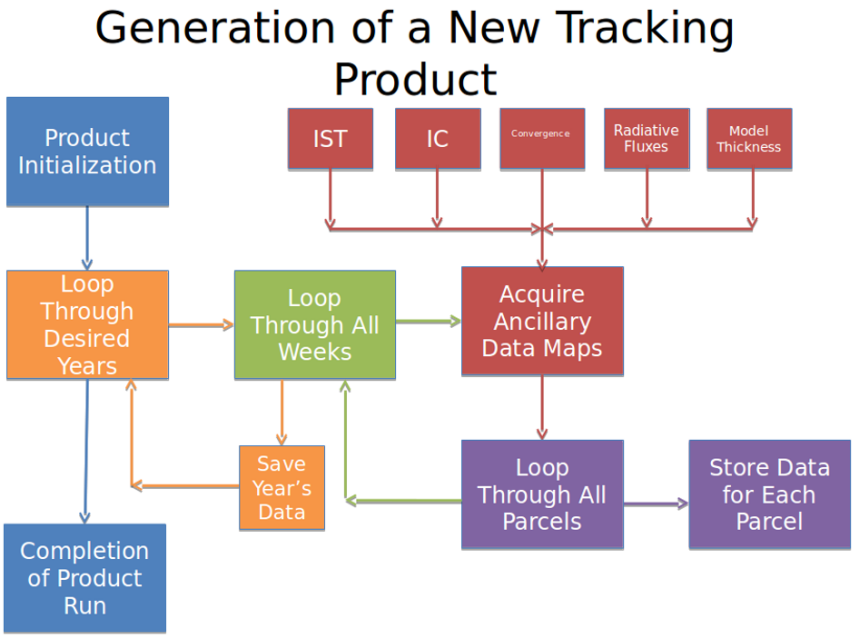

Figure 2.3: Procedural flow for generating a year of parcel tracks with ancillary data.

2.4.1 Initialization of Field

In order to initialize the ice parcel field, the program begins by searching for sea ice parcels that are of greater than 15% sea ice concentration in week one of the year. Parcels that meet these criteria are assigned a parcel ID and are given space in a Python dictionary object. The initial fields for the ancillary variables are also stored in the initialization phase for week one.

2.4.2 Weekly Data Acquisition

After the initialization steps are completed, the program loops through every week of the year to save the weekly ancillary data for each parcel. It begins by verifying that data are available, and loading in maps for each ancillary data product for that week. Once these maps are loaded, and any averaging of daily values to weekly values is performed, the program enters a loop that goes through each parcel to save their ancillary data.

2.4.3 Storing Parcel Data

After the ancillary data are loaded, the program loops through each parcel in the database in order to update each parcel's position and ancillary data values. The previous week's sea ice motion vectors are applied to the parcel's EASE-Grid location, which provides a new location for the current week. Once this new location is saved, the grid coordinates are used to find and store ancillary data for that particular parcel. If data are missing, a fill value is assigned. After the ancillary parcel data are stored, the program also increments the weekly age value of the parcel by one. This incremented age value is separate from the EASE-Grid age product hosted at NSIDC, which is generated through a different process [164].

2.4.4 Melting and Formation of Parcels

During the parcel sub-loop in the program, a check is performed for parcels that have dropped below 15% sea ice concentration. At that point the parcel is considered melted, and is assigned fill values in lieu of ancillary data values. Another data field stores the week in which that parcel melted for verification after the product run. In addition to searching for melted parcels, the program checks parcel positions against a landmask and terminates parcels that make landfall.

In order to account for parcel formation later in the year, the program checks for new parcels in areas that are not occupied during the year's freeze-up period. After the minimum extent week is reached for a particular year, the program begins to search all non-occupied locations for sea ice parcels that contain greater than 15% sea ice concentration. These newly frozen parcels are

assigned a space in the parcel database, and are tracked in subsequent weeks. In order to maintain the proper database size and spacing, weeks prior to the formation of the particular parcel are filled with fill values for the ancillary variables and locations.

2.4.5 Production of Final Database

Once all of the weeks have been sampled, and the database has been composed, the product saves the data in a combined comma-separated value (CSV) file for the year. Further steps break this file up into smaller files for individual ancillary variables in order to maintain the 100MB size limit set by Pangaea.

2.4.6 Example Outputs From Database

An example of tracks and ancillary data from the 2014 data is shown in Figure 2.4. The major motion patterns of the Arctic are visible from the drifts of the example parcels in the left panel of the figure, while the right panel contains the ice surface temperature (IST) and concentration history for an individual parcel that advected from the Canadian Archipelago region toward the Beaufort Sea. As the melt season began, the IST of this parcel rose to the melting point, which led to some of the parcel melting away and a corresponding decrease in concentration. This database enables quick searches for relationships like these for individual tracks and different variables from the beginning of 2001 to the end of 2016.

Another example use of this database to gather bulk regional statistics for 2014 is shown in Figure 2.5. The left panel shows a user-defined Beaufort Sea region highlighted in red, while the right panel is a scatter plot of the ice surface temperatures of parcels in that region for the duration of 2014. Parcels that melted (red) experienced higher temperatures than their surviving counterparts (blue) prior to the start of the melt season, which may have played a role in their survivability during the summer. Searching regions like the one shown in Figure 2.5 enables users to examine relationships between different cohorts of parcels in areas of their own interest.

An example of using this database to study inter-annual changes in parcel properties is shown

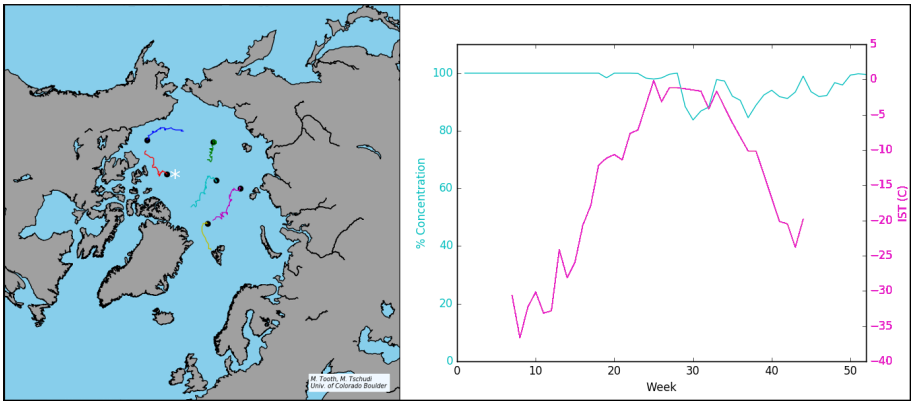

Figure 2.4: Selected example parcel drift tracks from 2014 with IST and concentration for the parcel indicated by an asterisk in the left panel.

in Figure 2.6. The top left panel shows the positions of four parcels as they advect through the Arctic from 2001-2004. The upper right panel shows the changing ice surface temperatures during each year as they rise and fall with each melt season. The lower left panel shows the trends in ice concentration for each year. Lastly, the lower right panel shows changes in thickness over time for each of these parcels as they seasonally gain and lose ice each year.

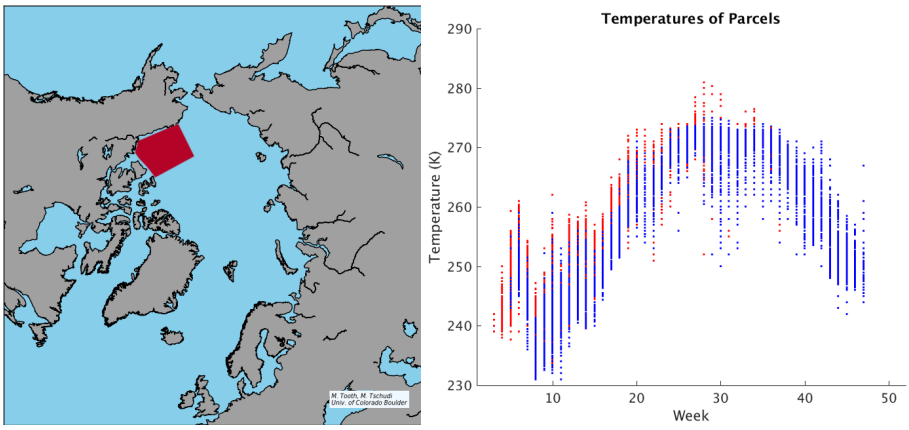

Figure 2.5: Beaufort Sea example region with associated parcel temperatures for 2014. Red scatter points indicate parcels that melted, while blue points indicate parcels that survived to the end of the year.

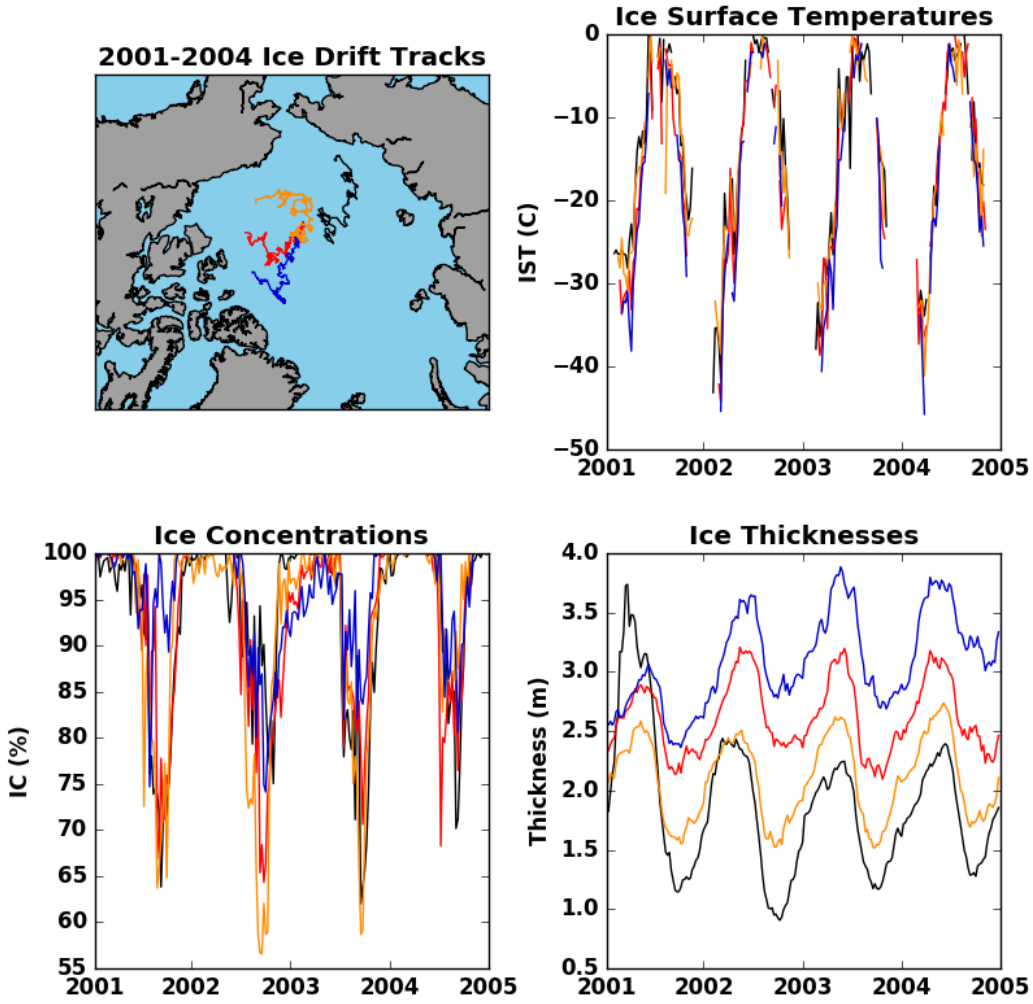

Figure 2.6: Tracks for four parcels from 2001-2004 with associated ancillary data.

2.5 Publication of Database

The Pangaea Data Archive had several requirements that differed from the original file format of the parcel database output, so codes were developed to re-format the files to fit the Pangaea specifications. Files were to be stored in tab delimited format text files that did not exceed 100MB each. In order stay below the file size limits the parcel data were separated into individual variable files for the year. Each file contained the U and V positions of parcels along with a single variable such as IST or concentration. This both solved the size issue and simplified access for users wishing to investigate a subset of the data contained in the parcel database. The original Pangaea release in March 2016 contained NaN values, a small error in PIOMAS data, and spanned 2001-2015. The second release in March 2018 included 2001-2016, fixed the error in the PIOMAS data, and updated fill values to the new standards described above. It is also important to note that the Pangaea releases contain integer positions, while the database natively uses float positions.

In addition to the release of the data, codes were supplied in the Pangaea repository for users to access the data from the files provided. These codes were provided in Python, C++, and Matlab for the convenience of potential users. Additional updates, such as the inclusion of 2016-2017 data, are possible through updating the Pangaea link without a change to the data DOI that was originally assigned by Pangaea. This will ensure that users are able to continually find and utilize the data as it is improved, updated, and re-released.

2.6 Potential Future Expansion of Database

One of the major strengths of this database is that it is modular, and can be expanded to include more data products or alternate sources of data products. Desired data products can be added through the use of additional read/write functions, and the code that generates the parcel database has been constructed in a way that makes it easy to do so. This enables future runs of the database to include these alternatives at the discretion of those running the code. Some potential future products include snow depth, surface winds, and alternative ice thickness sources.

2.7 Discussion of Error

Work by Szanyi et al [150] argued that there are small error contributions in the underlying sea ice motion data that come about as a result of the weighting scheme used to combine the various motion sources used in the generation of the motion product [165]. These "buoy-affected-domains" can contribute to errors in motion tracks that are used in the generation of this database. The use of weekly sea ice motion data mitigates the effect of this error on the motion data however, and enabled the publication of the database with minimal influence from the buoy-affected domains. An update to the underlying motion tracks has resolved this issue as of Fall 2017 [165]. Comparisons of parcel tracks with IABP buoys [130], CRREL buoys [34, 125], and the SHEBA track [143] yielded errors on the order of those found in [163].

Due to the nature of how the database and the weekly data products are published, each week of ancillary data should be considered independent with regard to error. The individual error for each data product field does not depend on the previous week's state, and therefore enables careful use of the data in this product with consideration of the error present in each individual data product. Further information about the error present in these products can be found in Appendix C.

2.8 Summary

The near continual movement of Arctic sea ice can expose individual ice parcels to varying conditions throughout the year. Several satellite-based sources of sea ice data were combined with Lagrangian track derived positions in this chapter to yield a database of ice parcel positions with coincident data. The satellite data chosen capture key properties, processes, and parameters of sea ice in the Arctic that are used to investigate sea ice survivability throughout this manuscript. Chapter 3 will use the product developed here to investigate the optimal statistical learning tools for assessing which of the parameters described here most impact sea ice survival. Chapter 4 further utilizes these data to determine predictor importances using statistical learning techniques.

Chapter 5 compares these data against Operation IceBridge data in the Beaufort Sea.

Chapter 3

Applying Statistical Learning to Predicting Arctic Sea Ice Survival

This chapter provides background on statistical learning-based classification, and the process through which four statistical learning techniques were selected for this study. The optimization and parameter selection for each classification technique is discussed, along with the underlying theory behind each algorithm. Classification accuracies are also presented, with a discussion of the optimal combination of training data to use during analyses. Section 3.1 introduces the classifiers, while Sections 3.2, 3.3, 3.4, and 3.5 provide additional information about each of the four classifiers. A discussion of the results is provided in Section 3.7, in which we determine that the random forest and neural network classifiers are the optimal tools for this work. The techniques described and explored in this section are utilized further in Chapters 4 and 5 to derive the relative importance of predictors from the database described in Chapter 2.

This chapter is partially adapted from *Tooth, M.; Tschudi, M.; and Matsuo, T. 2018* [160].

3.1 Introduction

Over the past several decades statistical learning techniques have become more widely used with the advent of more powerful computers and accessible methods. These efforts have focused primarily on surface type classification, dimensionality reduction, and image feature classification using discriminative classifiers [3,8,45,110]. Arctic work using these techniques often falls into these categories, as these classifiers are valuable tools for remote sensing data analysis [51, 90, 106, 138].

Recent efforts toward utilizing statistical learning to form summer ice extent predictions have

outperformed other methods [21], which indicates that further exploration of these techniques is warranted. Statistical techniques yielded the most accurate estimates for the 2017 Sea Ice Prediction Network (SIPN) outlook, but these methods were made up of regression and correlation based estimates [144]. The application of statistical learning techniques to prediction of summer ice extent will provide additional tools for the sea ice community that may serve to improve estimates such as those made for the SIPN outlook.

This study utilizes the Lagrangian parcel track database described in Chapter 2, along with sea ice age data from NSIDC (Appendix C) [164] to compare the prediction accuracy of four statistical learning classifiers in the pan-Arctic domain. The predictors chosen for use in classification are:

(1) Latitude

(2) IST

(3) Thickness

(4) Albedo

(5) Downwelling Shortwave Radiation

(6) Downwelling Longwave Radiation

(7) Ice Age

where these predictors are supplemented by the addition of a binary value, Y, that represents whether a parcel melted (0) or survived (1) in a given melt season. Data are assembled from the Lagrangian tracking database for years 2002-2016 and exclude parcels that were not present at the start of their respective year. The total number of 12.5km by 12.5km EASE-Grid parcels, along with the number of melted and surviving parcels, are shown in Table 3.1. These data were used to predict summer sea ice survival based on snapshots of values at each study week in order to demonstrate the utility of these techniques.

Table 3.1: Number of parcels utilized by classifiers in years 2002-2016

Year	Melted	Survived	Total	Percent Survived
2002	8854	24288	33142	73
2003	8442	24560	33002	74
2004	8180	23564	31744	74
2005	8445	19297	27742	70
2006	8647	24439	33086	74
2007	10192	16855	27047	62
2008	10045	21449	31494	68
2009	10054	22857	32911	69
2010	9710	18040	27750	65
2011	11419	23019	34438	67
2012	12435	18075	30510	59
2013	8650	17363	26013	67
2014	9501	20401	29902	68
2015	10168	23644	33812	70
2016	14477	17630	32107	55

The goal of this analysis is to utilize the truth data present in the Lagrangian tracking database to train classifiers to predict whether a given sea ice parcel in the database would melt or survive during the melt season. The classifiers were chosen after a preliminary investigation of available techniques in the broad field of statistical learning (Figure 3.1). The presence of truth data eliminated clustering and other unsupervised methods from consideration, which focused this study on supervised classification. Four discriminative classifiers were chosen in order to directly learn decision boundaries from the datasets studied. The chosen classifiers were: support vector machines, binary tree-based classifiers, random forests, and neural networks. These methods were chosen for both their predictive capability, and their associated use in inference about the relationship between predictors (X) and class (Y) after training [61]. Time is expressed in weeks throughout this chapter, as the underlying Lagrangian tracking data have a weekly temporal resolution. A table showing the dates of several weeks in the study period has been included for the reader's reference (Table 3.2).

Each classifier chosen for this study offers a different way of analyzing the data, but none of them rely on an underlying statistical distribution in the data. The support vector machine

method is a useful tool for finding non-linear boundaries in a dataset, and attempts to minimize how well the determined boundary fits the data. The neural network is a very flexible technique that aims to minimize mean squared prediction error with respect to the training data. The binary tree-based classifier is fast and easily interpreted, but it is not considered comparatively accurate or portable. The random forest provides greater tree-based accuracy through growing an ensemble of decorrelated trees that use bootstrap samples of the original dataset and restricting the number of parameters that can be used at each split in a tree. The work presented in this chapter seeks to determine which of the four chosen classifiers works best with the Lagrangian tracking data we utilize herein.

Table 3.2: Calendar dates of weeks in a non-leap year.

Week	**DOY**	**Calendar Date**
10	63	Mar 4 - Mar 10
15	98	Apr 8 - Apr 14
20	133	May 13 - May 19
25	168	Jun 17 - Jun 23
30	203	Jul 22 - Jul 28
35	238	Aug 26 - Sep 1

3.2 Support Vector Machine Approach

The support vector machine algorithm is a classifier that is an extension of the maximum margin classifier and the support vector classifier [25, 61]. It is capable of creating non-linear decision boundaries in a dataset through the use of the "kernel trick" which enables the algorithm to operate in a vector space where linear relations exist between the data. This provides a means by which linear algorithms can be applied to data with non-linear relationships. The linear results obtained in the transformed space are used to form non-linear boundaries in the original space that are difficult to calculate in the original data space.

In the linear case, separation of data that are composed of two classes can be attained through placing a linear boundary between points on a plot. This is not unique however, as there

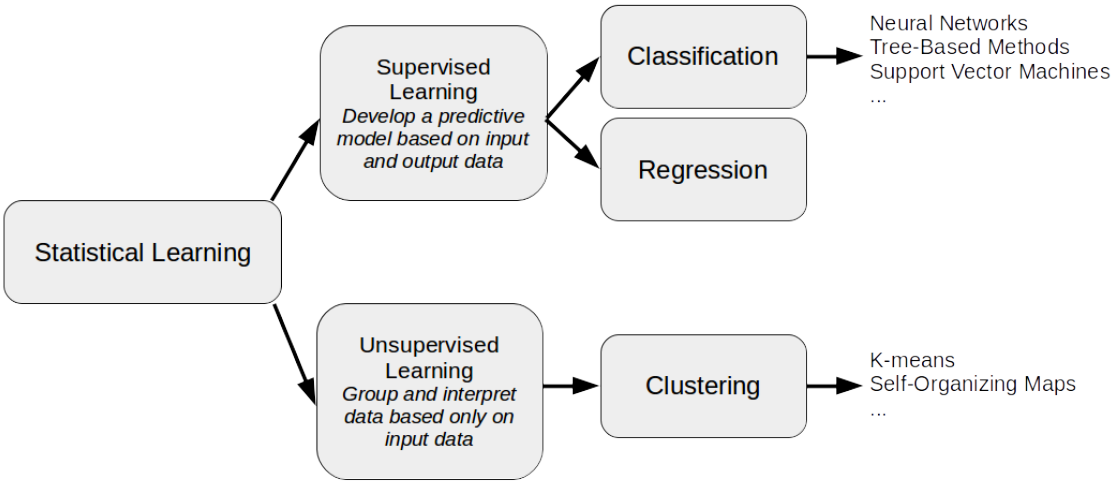

Figure 3.1: Broad categories of statistical learning with some examples from subcategories shown.

exist a large set of suboptimal lines that divide a dataset. The maximum margin classifier seeks to estimate a decision boundary such that the distance of the nearest data points to the linear boundary are maximized for both classes (Figure 3.2) [9,61]. This provides a means through which the subset of data that are closest to the boundary are used to classify new points without a need for considering the entire dataset. The subset of points that define the decision boundary are referred to as "support vectors". When data are not linearly separable, the kernel trick is applied in order to produce a nonlinear decision boundary [61, 107, 149].

3.2.1 Methods: Support Vector Machine

The support vector machine technique is implemented using the Matlab fitcscvm() function. The effect of varying the portion of the original dataset utilized in training is assessed for the cases of a quarter and a half of the available training data. Due to the computational complexity of the algorithm, the tests are limited to the year 2010, which was chosen as the first year for initial testing. In addition, several kernel functions are tested to find an optimal choice for classification of parcels into melted and survived categories. These included a Gaussian, linear, and quadratic kernel. These tests are also limited to the year 2010 due to computation time, which was on the order of six hours for each week of each study year.

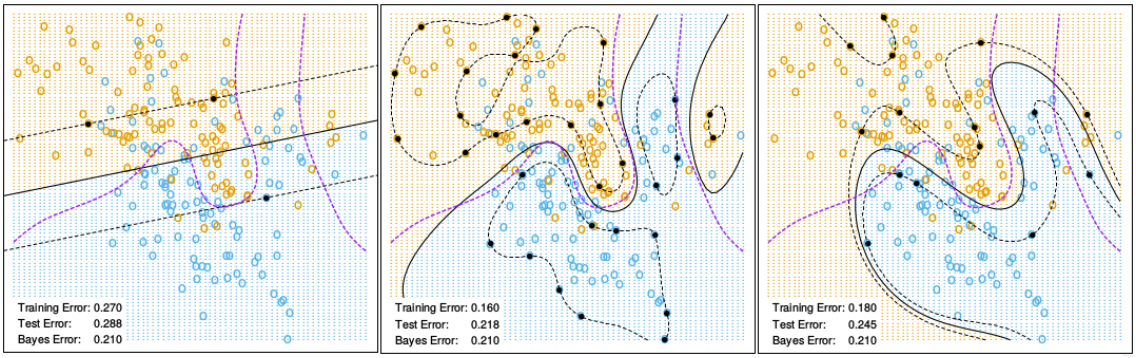

Figure 3.2: Linear support vector separation (left), Radial SVM separation (center), and Polynomial of degree 4 separation (right) [54]

3.2.2 Results: Support Vector Machine

Peak survival prediction accuracy of parcel survival is obtained in week 18, with an accuracy near 78% for both training portion cases (Figure 3.3). This peak is followed by a gradual decline toward accuracies in the 60% range later in the melt season. This is likely a result of the increase in missing data values as the melt season progresses. Varying the portion of data utilized during training has a minimal effect on the classification accuracy for the SVM technique. Increasing from a quarter to a half of the data increases peak accuracy in week 18 by less than 1%, while causing several drops in accuracy during weeks 16 and 24. This suggests that utilizing a quarter of available training data is sufficient to form a decision boundary, and that increasing the amount of data included in training does not offer the algorithm additional information that can be utilized for classification.

Varying the kernel function used to form the boundary during training has a significant impact on the classification accuracy, but does not yield superior results as compared to the Gaussian case (Figure 3.4). The Gaussian test case peaks in accuracy during both weeks 12 and 18, with peaks near 80%. The linear kernel function yields a peak accuracy in week 17, with a peak near 77%. The quadratic kernel significantly reduces overall accuracy, and yields a peak near 65% accuracy in weeks 10 and 33. The Gaussian and linear cases exhibit the general behavior of peaking early in the season, and declining later in the melt season, while the quadratic case resembles a bathtub curve. This suggests that the Gaussian kernel function is the optimal choice, as it yields higher accuracies than its counterparts.

Varying input training years for testing of prediction is not shown for the SVM technique due to low accuracies found during initial tests and computational cost. The classifiers described in the following sections outperform the SVM technique in all cases, and perform better in the presence of missing data than SVM.

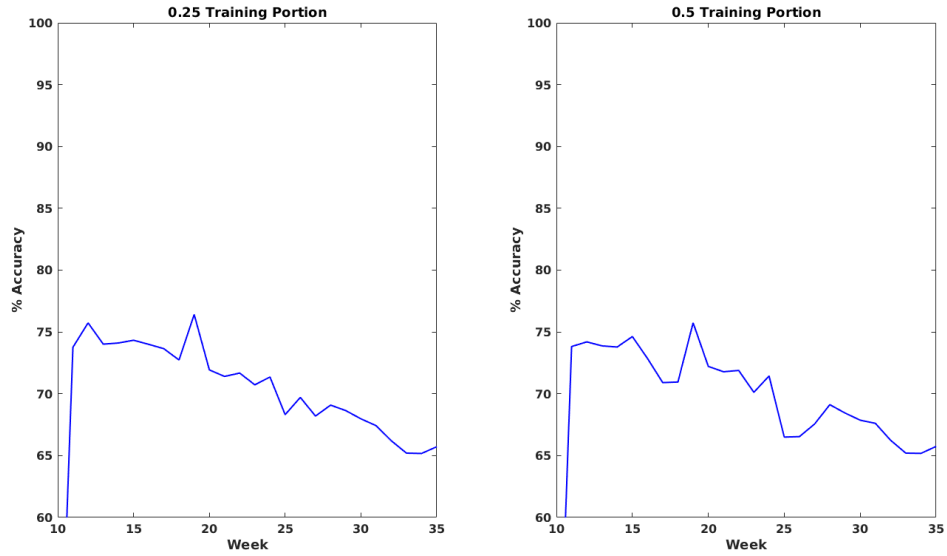

Figure 3.3: 2010 SVM accuracy with varying training portions and a radial basis function kernel

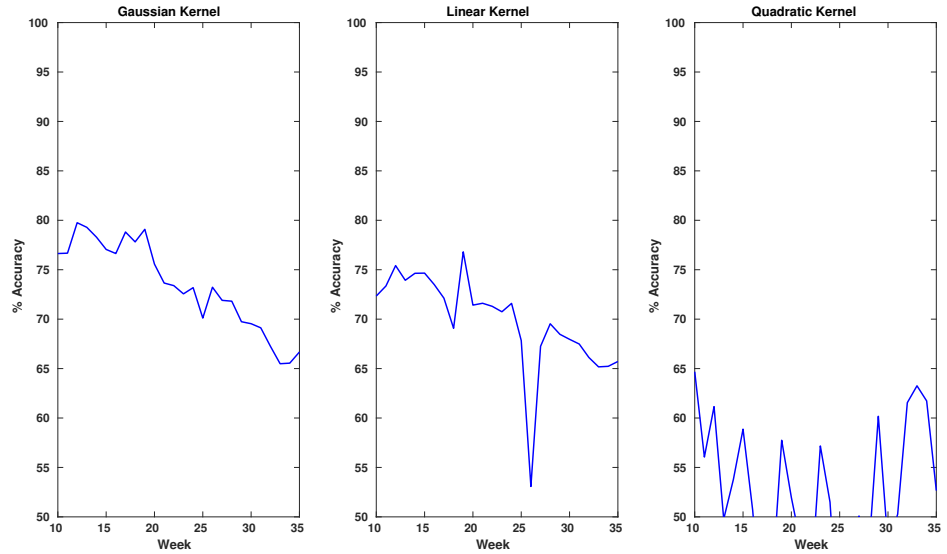

Figure 3.4: 2010 SVM accuracy with varying kernel functions

3.3 Binary Decision Tree Classifier

The binary decision tree classifier is a tree-based method that seeks to classify points through dividing data into M rectangular regions:

$$R_1, R_2, ..., R_m, ..., R_M \tag{3.1}$$

where each region R_m is known as a terminal node in the tree, or as a leaf. This process is performed iteratively, with the dataset being split at internal nodes along the tree analogue [13, 61]. The end result is a set of regions that are labeled as one of K classes: $c_m \in 1, ..., K$.

In an example with training data (x_i, y_i); $i = 1, ..., n$; where $y_i \in (1, ..., K)$ are class labels; suppose a given region R_m contains n_m points. Then the predicted class probability of class $k \in K$ is given by:

$$p_k(R_m) = \frac{1}{n_m} \sum_{x_i \in R_m} 1_(y_i = k) \tag{3.2}$$

The predicted class in R_m is determined by:

$$c_m = max(p_k(R_m)), k \in K \tag{3.3}$$

The tree building process takes a top-down, "greedy", approach. The algorithm begins with the entire dataset and subdivides it through successive splits in the tree, with each split forming two new branches in feature space. The algorithm is considered greedy due to it choosing the optimal split at each step instead of choosing the split that would most optimally split the tree in future steps. This process is repeated until a large tree is grown that can later be pruned at the end of the creation of the tree [61, 93]. An example tree grown using week 12 of 2016 as training data is shown in Figure 3.5. Individual splits in the tree are formed based on underlying data. For example, the top split in Figure 3.5 is formed using a boundary of 76° of latitude, with a successive split to the left that uses 75.7° as the boundary (Figure 3.6). This second split is followed by a node that uses thickness to split along a boundary of 2.16m thickness. The values that form these splits were determined by the classification tree during training.

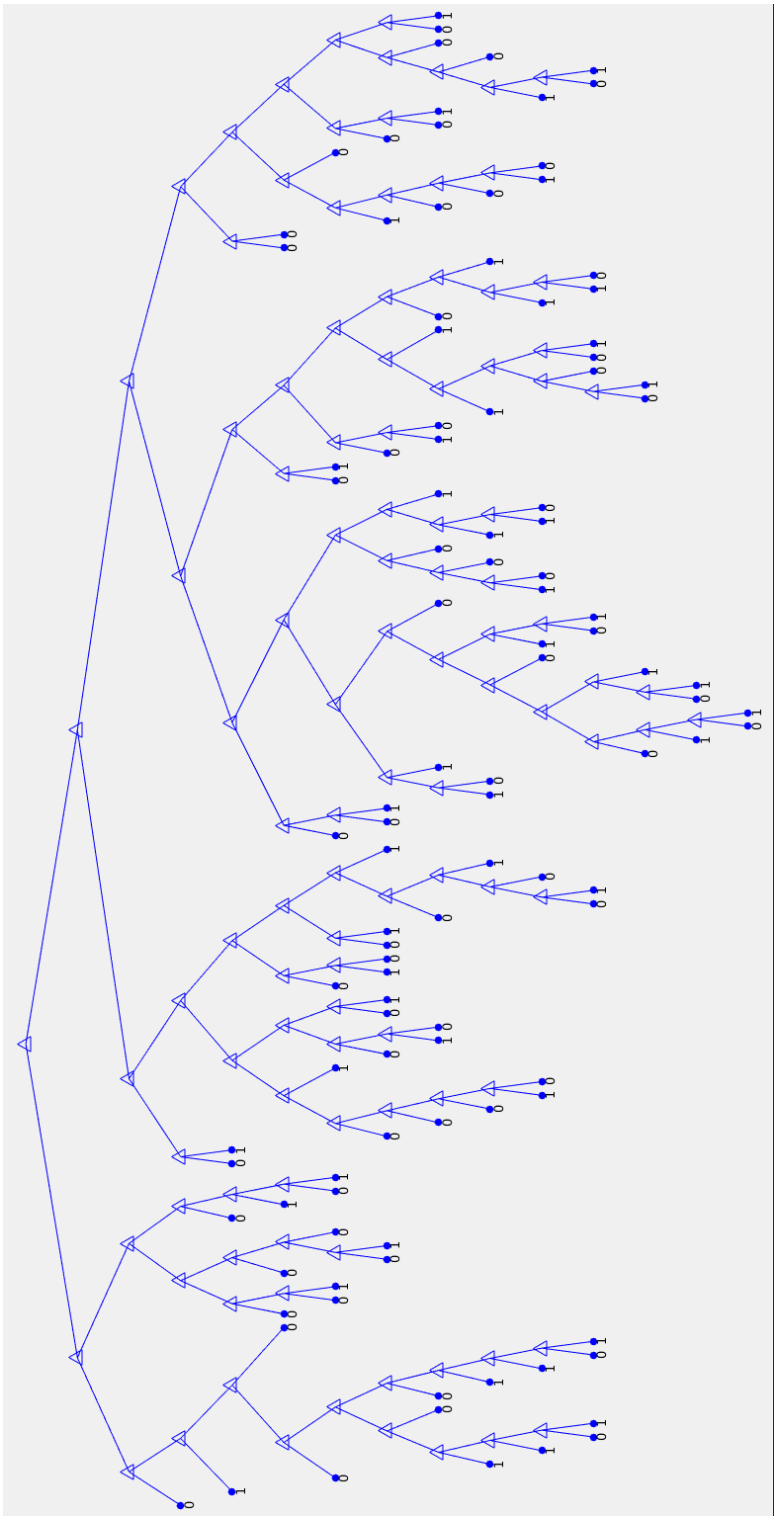

Figure 3.5: Classification tree grown for week 12 of 2016 with terminal nodes labeled as 0 for melted and 1 for survived.

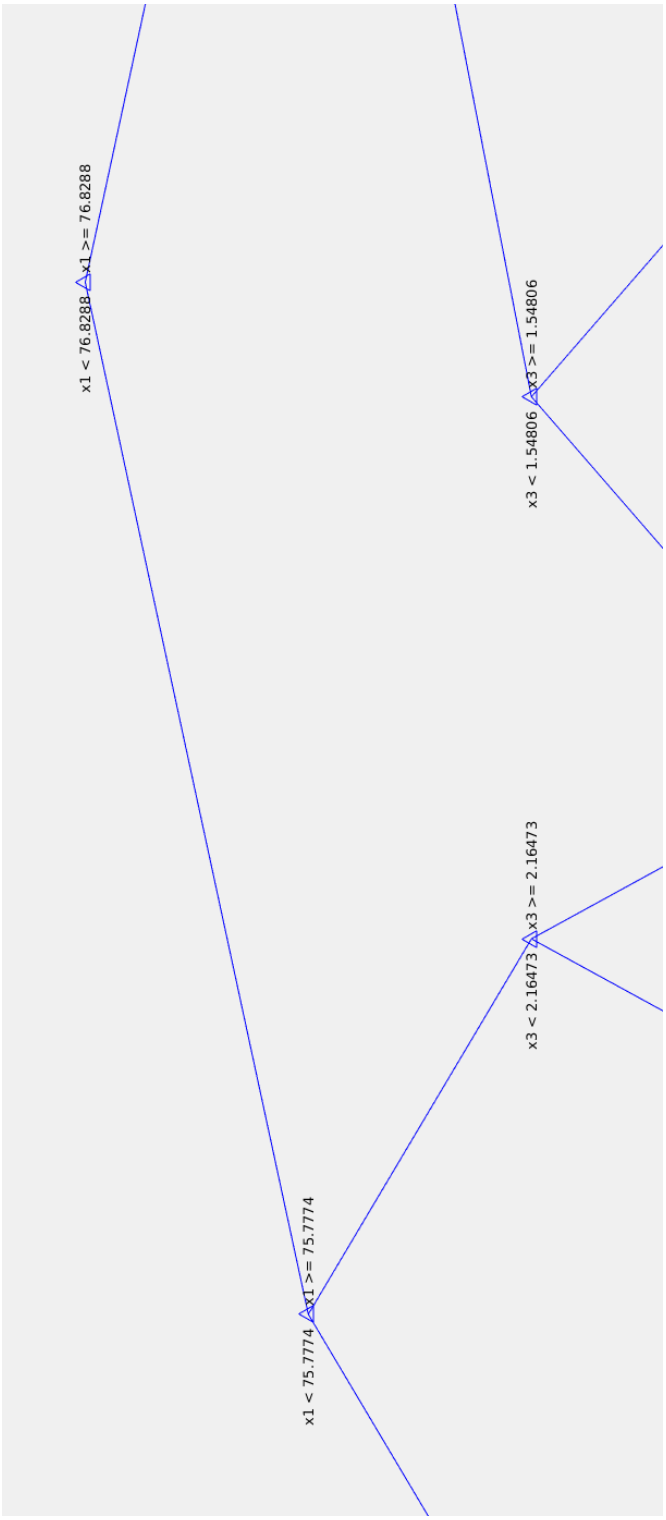

Figure 3.6: Example classification splits for week 12 of 2016. X1 corresponds to latitude, while X3 represents PIOMAS thickness.

Split criterion in growing classification trees can be defined in various ways, but they are based on node impurity. Two example measures of impurity are:

(1) Misclassification error:

$$1 - p_k(R_m) \tag{3.4}$$

where p_k is the class probability of region R_m as defined in Equation 3.2.

(2) The Gini index, which is a measure of total variance:

$$\sum_{k=1}^{K} p_k(R_m)(1 - p_k(R_m)) \tag{3.5}$$

Criterion such as the Gini index and misclassification error seek to greedily maximize the purity of a given split. Each subsequent split in the tree continues until a tree T_0 with T terminal nodes is grown. A final process called *pruning* seeks to form a subtree $T \subset T_0$ that minimizes:

$$\sum_{m=1}^{|T|} (1 - p_{cm}(R_m)) + \alpha T \tag{3.6}$$

where α is a tuning parameter that controls the size of the tree that is typically chosen through k-fold cross-validation. Individual leaves are removed starting with the "weakest" leaf at a time until the desired tree **T** is obtained.

Classification trees provide structures that are easily displayed graphically, and are considered interpretable in comparison with other methods. The computational complexity of generating a tree is low, which lends toward fast training and classification of large datasets. Classification trees suffer from lower predictive accuracy as compared to other methods however, and they are not considered portable. Small changes in the data utilized in training lead to large changes in the resulting structure of the tree. These drawbacks can be corrected for however, as is discussed in Section 3.4.

3.3.1 Methods: Binary Decision Tree

Binary tree-based classifiers are generated using the Matlab function fitctree() for this analysis. The impact of varying the portion of data used in training is assessed for 25, 50, and 75% of

available data. Additional cases assess the impact of including additional years of training data on the resulting classification accuracy.

3.3.2 Results: Binary Decision Tree

Survived/melted class classification accuracy peaks early in the year, with a maximum value occurring near week 12 (Figure 3.7). The addition of more data in training leads to a higher prediction accuracy in all cases, and leads to an increase in peak accuracy from 86% in the 25% of training data case to an accuracy of 88% in the 75% of training data case. These results lead to the use of week 12 and 75% of training data for the remainder of work with this classifier.

The impact of varying input training years on the binary tree-based classifier is assessed using predictions during week 12 for years 2012-2016 (Table 3.3). Test cases include the prior ten years, prior five years, and same year for the three test cases, with 75% of the available data being used for training. The inclusion of the prior ten years of data outperforms the prior five years of data for 2012-2014. In year 2015 and 2016 the prior five year cases are more accurate than the prior ten year case by 0.4% and 3.4% respectively. Training with the same year of data outperforms all cases with prior data for years 2012-2016.

Table 3.3: Binary tree-based classifier week 12 prediction accuracies for validation years 2012-2016 and various training years

Validation Year	Prior 10	Prior 5	Same Year
2012	76.2	74.7	88.3
2013	75.3	73.2	90.0
2014	73.2	69.2	85.9
2015	75.7	76.1	88.0
2016	66.0	67.4	82.1

3.4 Random Forest Approach

The binary tree-based classification approach described in Section 3.3 can be improved upon through the use of the bagging and random forest techniques [11,12,54,55,61,118]. These techniques

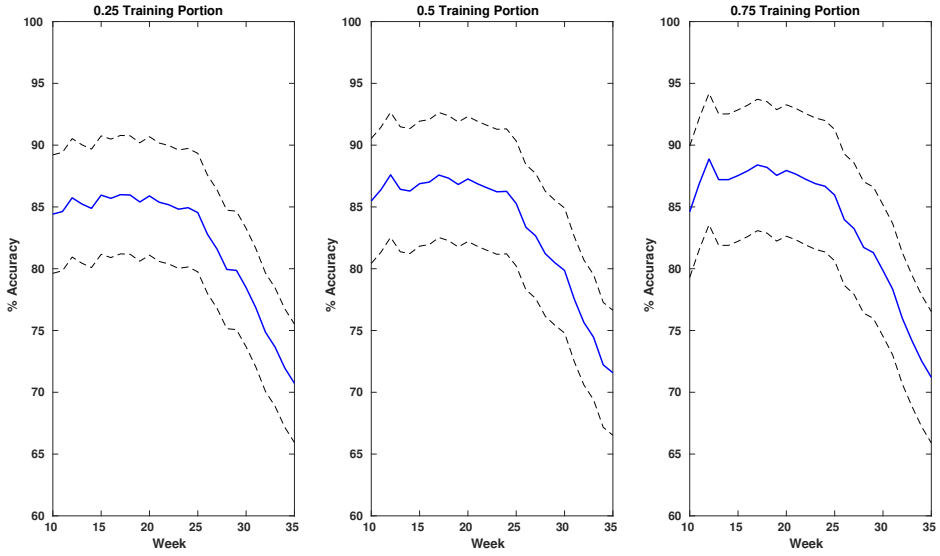

Figure 3.7: Mean binary tree-based classifier accuracy with varying training portions

help control the variance present in the binary tree-based classifier, and stabilize predictions made by the classifier [12, 92]. This is done through growing an ensemble of trees from the training data and combining the predictions of individual trees in the ensemble. Two common techniques that do this are known as bagging and random forests.

In the case of bagging (bootstrap aggregation), the algorithm averages predictions from classification trees over a collection of B bootstrap samples. Given training data $(x_i, y_i), i = 1, ..., n$, and $b = 1, ..., B$; n samples $(x_i^b, y_i^b), i = 1, ..., n$ are drawn from the data to fit a classification tree $f^{tree,b}$. After generating B trees from the data, the algorithm chooses either the class with the most votes:

$$f^{bag}(x) = maximum_{k=1,...,K} \sum_{b=1}^{B} p^{tree,b}(x) \tag{3.7}$$

or the class with the highest probability:

$$f^{bag}(x) = maximum_{k=1,...,K} \frac{1}{B} \sum_{b=1}^{B} p^{tree,b}(x) \tag{3.8}$$

The bagging technique works by reducing the probability of misclassification with an increase in B when the classifiers considered are independent. The classifiers are not independent in practice however, as they are fit using samples drawn from the same training set. This leads to poor performance in the case where the base classifier is poor, and improved performance when the base classifier performs well in terms of prediction error.

The random forest technique improves upon the bagging technique by de-correlating the trees in the ensemble. An ensemble of decision trees is constructed in the same manner as in bagging, but at each split in the constructed trees a random selection m_r of the predictors is chosen for consideration in the split. The most common number of predictors is $m_r \approx \sqrt{D}$, where D is the total number of predictors. In the case where all of the predictors are used ($m_r = D$) the random forest is made equivalent to bagging [54, 61, 93].

While bagging and random forests can increase the classification accuracy of the algorithm, they produce opaque results as compared to a single classification tree. Graphically representing the final decision tree is not possible in the ensemble case, which complicates interpretation of

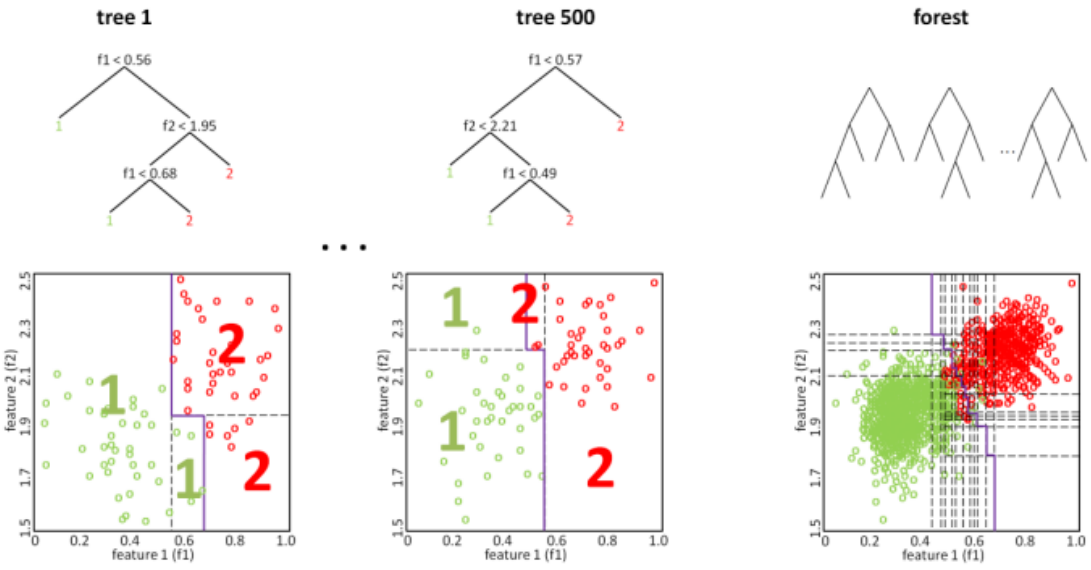

Figure 3.8: Two example trees with associated decision regions (left, center), and a forest ensemble with example member decision regions (right) [52]

the result. Additionally, the computational complexity increases from N splits to $N * B$ splits during training. In big data applications with lower accuracy requirements the binary classification tree may offer superior performance as compared to random forests and bagging depending on the project requirements.

3.4.1 Methods: Random Forest

The random forest technique is implemented in Matlab through the use of the treebagger() training function. The parameters of the random forest are selected through testing various combinations of forest size, minimum leaf size, and training data portions. Further work assesses the impact of varying the training input years on classification accuracy for years 2012-2016.

3.4.2 Results: Random Forest

The impact of varying the portion of data used in training is assessed using 25, 50, and 75% of available training data for all years 2002-2016 using a random forest of 100 trees (Figure 3.9). Peak melted/survived class prediction accuracies from 95-96% are attained in weeks 25-35 for all three test cases, with typical accuracy values from 92-95% throughout weeks 10-35. The classification accuracy is typically stable between weeks 28 and 32, so week 30 was chosen as the optimal prediction time for further testing of the classifier.

The effect of varying the number of trees utilized in the random forest classifier is also assessed (Table 3.4). Week 30 prediction accuracies are derived through the use of 50% of available training data for years 2002-2016 with varying numbers of trees in the random forest. Average accuracy gradually increases with the addition of more trees in the forest, and a peak accuracy near 96.7% is attained at and above a forest size of 100 trees.

Test cases are formulated to determine whether the inclusion of additional years of data during training has an impact on week 30 classification accuracy for the random forest implementation. A random forest of 100 trees is trained using 50% of available training data for the validation years of 2012-2016. The test cases incorporate the same year of data, the prior five years of data, and

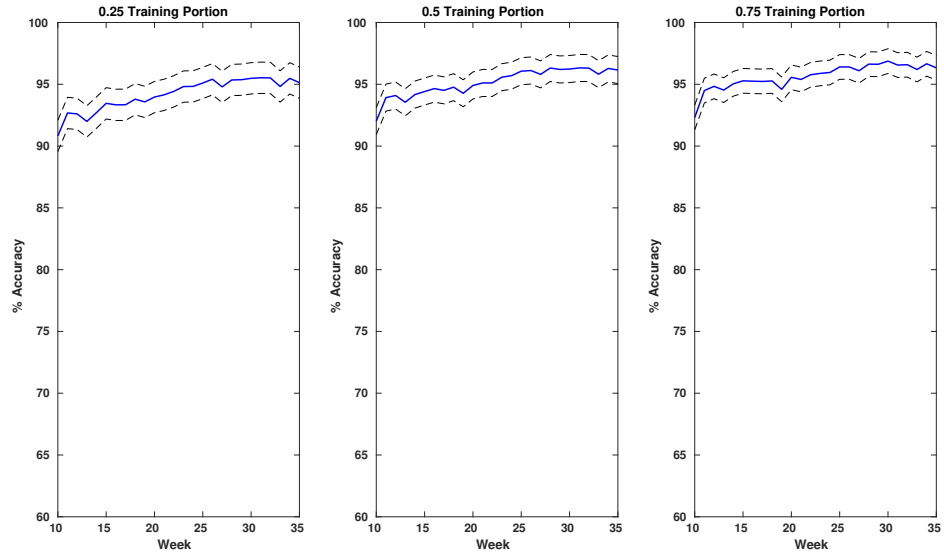

Figure 3.9: 2002-2016 mean random forest classification accuracy with varying training portions. Dashed lines indicate one standard deviation away from the main curve.

Table 3.4: Mean random forest classification accuracy for various numbers of trees in the forest

Number of Trees	**Mean Week 30 Accuracy (%)**
10	96.4
15	96.4
20	96.5
25	96.5
50	96.6
75	96.6
100	96.7
150	96.6
200	96.6
250	96.6

the prior ten years of data for the validation years (Table 3.5). The classification accuracy of the prior ten-year trained case outperforms the prior five-year trained case for years 2012-2014, with differences between 0.2% and 1.2%. In 2016 there is no difference in classification accuracy between the two cases, and in 2015 the prior five-year trained case accuracy exceeds the prior-ten case by 0.3%. The same-year accuracy exceeds both of the prior year inclusion cases in all of the study years by up to 16%.

3.5 Neural Networks

The neural network classifier seeks to learn nonlinear decision boundaries in a dataset [9]. The algorithm is inspired by biological neural networks, and is composed of a set of basic units that act as neurons in the network [3, 9, 45, 61]. A typical network receives a vector of inputs, feeds these inputs into units (activation), and continues feeding forward in the network until the inputs reach the output unit(s). Any number of layers can be used between the input and output nodes, and each layer can contain any number of nodes. The layers between the inputs and outputs are known as "hidden units/layers", and receive a net input a according to:

$$a = \sum \omega_i x_i \tag{3.9}$$

Table 3.5: Random forest week 30 prediction accuracies for validation years 2012-2016 and various training years

Validation Year	Prior 10	Prior 5	Same Year
2012	84.8	84.5	96.1
2013	84.5	83.4	94.0
2014	88.3	85.5	93.5
2015	92.7	93.0	97.2
2016	82.0	82.0	96.4

where ω_i represents a weight applied to the input value x_i. Each unit sums the inputs and weights to obtain a, and inputs the result to an activation function $f(a)$. The result is then fed forward in the neural network until outputs are determined at the output nodes [47, 61]. The neural network technique can exhibit many exotic structures and combinations of nodes, which provides a means for customization of the technique for specific applications.

3.5.1 Methods: Neural Network

Two feed-forward neural networks are implemented in order to compare the two techniques. The first is a pattern recognition neural network, and the second is a fitting neural network. The pattern recognition network is implemented through the use of the Matlab routine *patternnet()*, which produces a network with sigmoid hidden layers and a softmax output layer (Figure 3.10). The fitting network is implemented using the Matlab routine *fitnet()*, which produces a network with sigmoid hidden layers and a linear output layer. These networks are trained with a variety of training data portions, and several ranges of input years. Additional analysis determines an optimal number of layers and an optimal number of nodes in each layer. The networks are provided with an input vector of seven sea ice parameters, and they output a binary Nx2 vector indicating the melted/survived class membership of parcel $n \in N$. Training is carried out using the scaled conjugate gradient and Levenberg-Marquardt backpropogation algorithms [47, 112].

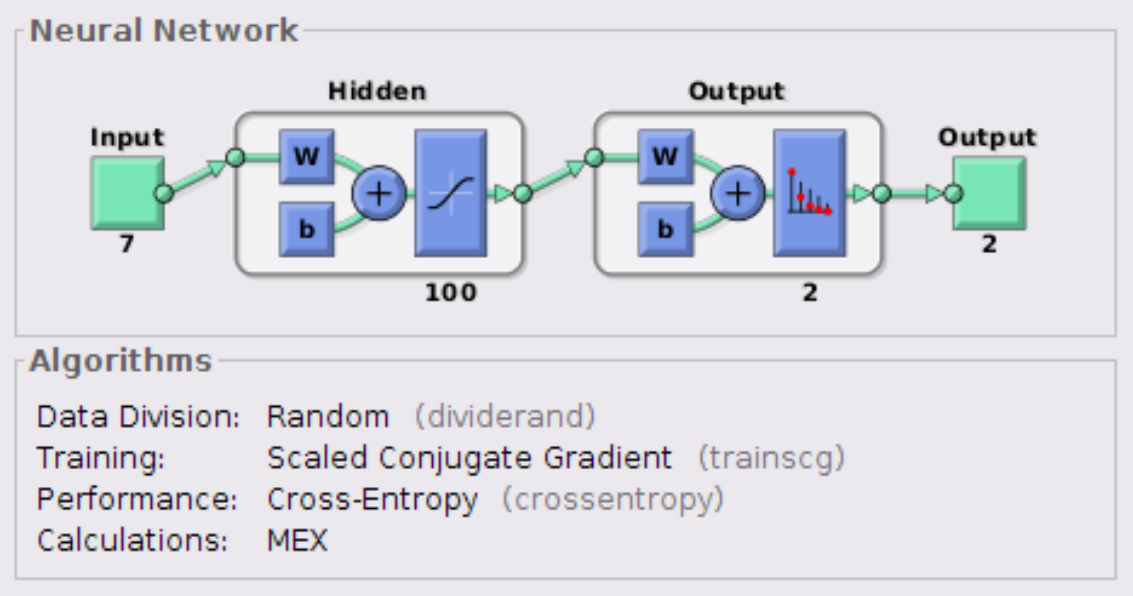

Figure 3.10: Two-layer pattern recognition neural network representation created using a Matlab network training tool

3.5.2 Results: Neural Network

The impact of varying the training data portion between 25%, 50%, and 75% is assessed for both the pattern and fitting neural network (Figure 3.11). We observe a less than 2% increase in survived/melted class prediction accuracy from the 25% case to the 75% case, with greater accuracy being found in the fitting network case. Accuracies from 80-85% are observed between weeks 10 and 20, with steadily increasing accuracy toward the end of the study period. Peak accuracies of 87% for the pattern network and 89% for the fitting network are attained near week 30.

The hidden layer size in both networks is varied to determine if there is an optimal number of hidden nodes to use during classification. Classification accuracy increases between 5 and 10 layers, but decreases with increasing numbers of nodes in the layer. This is likely due to there being sufficient freedom with a layer size of ten, and too much freedom in networks with larger layers. The optimal layer size of ten is utilized for the remainder of this analysis.

Table 3.6: Neural network peak classification accuracy for various hidden layer sizes

Hidden Layer Size	**Peak Accuracy (%)**
5	92.8
10	93.2
20	92.7
40	92.7
60	92.9
80	91.6
100	92.7

The effect of varying the number of hidden layers on classification accuracy is tested for both the pattern recognition network and the fitting network (Table 3.7). Four cases are tested for each method with a hidden layer size of ten and varying layer depths of 1-4 layers. The mean 2002-2016 accuracy at week 30 is assessed for each case in order to determine if there is an optimal combination. There are modest increases in accuracy with increasing layer size for the pattern net, with four 10-node layers being the optimal setup. The fitting network reaches peak accuracy with only two layers, and outperforms the peak pattern net results by 1%. This suggests that a

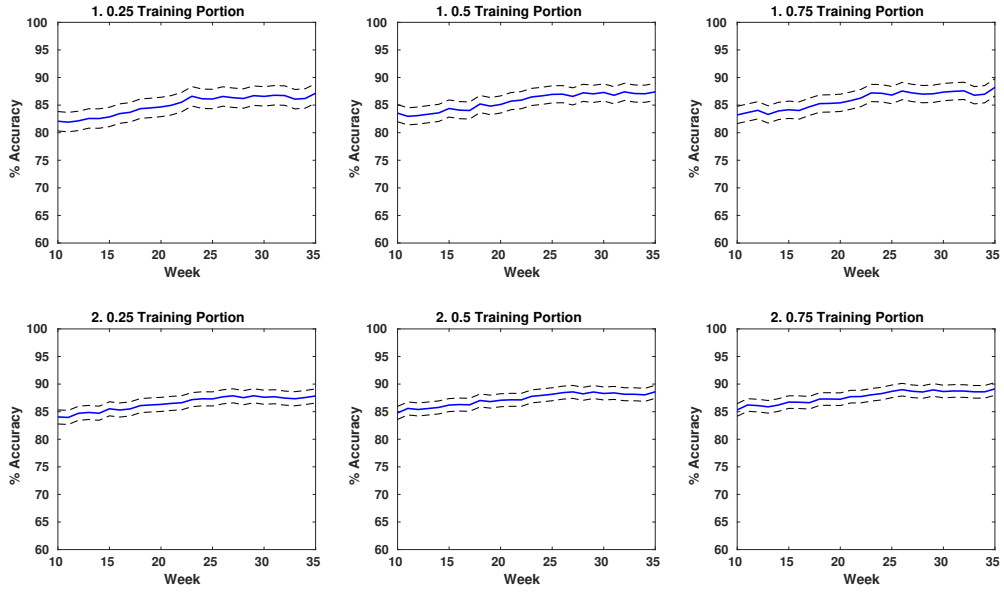

Figure 3.11: Mean pattern (1) and fitting (2) neural network accuracies with varying training portions

fitting network with two ten-node layers is the optimal tool for this analysis, and it is used in the remainder of this study.

Table 3.7: Neural network peak classification accuracy for various numbers of hidden layers of size 10

Network Type	**Number of Hidden Layers**	**Week 30 Accuracy (%)**
Pattern net	1	92.7
Pattern net	2	92.9
Pattern net	3	92.7
Pattern net	4	93.3
Fit net	1	93.9
Fit net	2	94.6
Fit net	3	94.5
Fit net	4	94.6

The impact of varying training years on the fitting network prediction accuracy is assessed for several cases: the prior ten years, prior five years, and same year of data (Table 3.8). This is done using 75% of available training data and a two-layer network. We observe little difference between the prior ten and prior five year cases outside of 2016. In 2016 the prior ten year case outperforms the prior five by 4%. In all years outside of 2015 the same-year accuracy is significantly greater than the prior year accuracies, while in 2015 the same-year accuracy exceeds the prior-year case by only 1%.

Table 3.8: Neural network week 30 prediction accuracies for validation years 2012-2016 and various training years

Validation Year	Prior 10	Prior 5	Same Year
2012	79.3	79.1	93.0
2013	80.1	81.4	90.0
2014	85.1	85.4	94.4
2015	93.0	93.1	94.7
2016	83.5	79.5	90.6

3.6 SIPN Example Case

An example use these classifiers for predicting September sea ice extent is shown using a random forest. This test case is performed in order to demonstrate the utility of the random forest in predicting September sea ice extent for efforts like the Sea Ice Prediction Network's Sea Ice Outlook. The results presented here are distinct from the previous results in that extent is being predicted in lieu of predicting survival of a summer melt season. The aim is to predict September ice extent in the first week of June, July, and August from 2008-2017. Work is carried out using a random forest of 100 trees that is trained using the prior five years of data as truth. The data consist of 25km EASE-Grid values of concentration, albedo, and ice thickness that are further described in Appendix C.

An example of the predicted extent maps obtained from this case study is shown in Figure 3.12. The spatial match accuracy of the random forest typically increases from June to August, with matches ranging from 75-91% agreement with truth without including land (Table 3.9). August predictions of extent are also within 0.5 million square kilometers of truth for most years. These results indicate that there is significant merit in the continued study of using this technique to form short-term predictions like those pursued by the SIPN team each summer. The results of this case study are comparable with other efforts to predict spatial maps, and are within 2% accuracy of other published work [51].

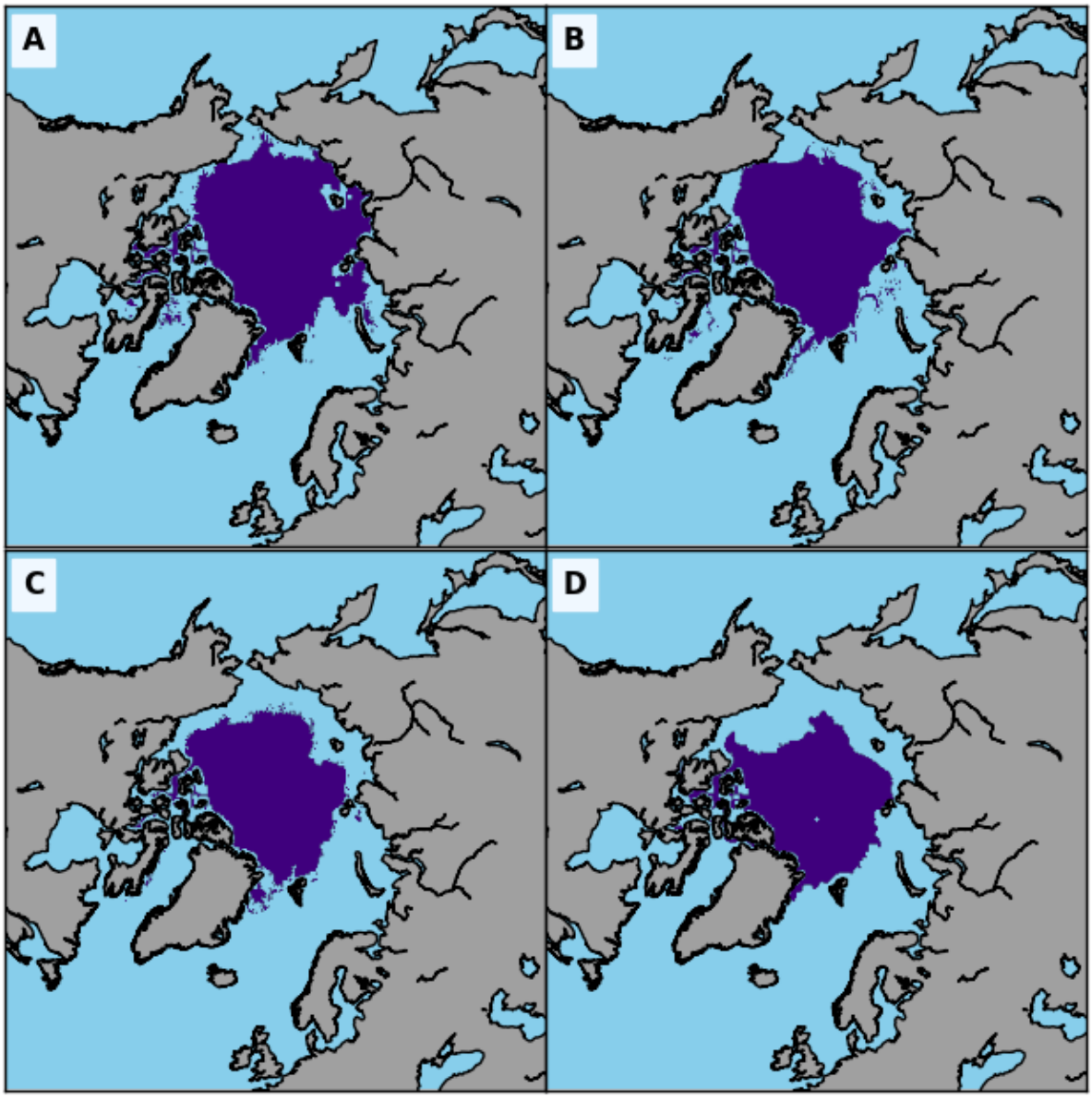

Figure 3.12: Predictions of September sea ice extent in 2017 made during (A) June, (B) July, (C) August. The true extent is shown in (D).

Table 3.9: Spatial match percentages between random forest predictions and truth for 2008-2017.

Year	June	July	August
2008	75.38	80.13	86.31
2009	82.32	87.94	89.49
2010	82.62	86.52	88.80
2011	81.23	87.36	91.17
2012	78.45	86.79	89.95
2013	84.89	82.73	89.90
2014	79.38	82.63	84.95
2015	80.64	84.96	87.87
2016	79.19	84.09	88.74
2017	76.60	84.02	90.26

3.7 Discussion

For the SVM classifier we observe that a quarter of the available data in the database is sufficient for training, and that peak performance near 80% accuracy is obtained using a gaussian kernel function. Computational cost, along with low accuracy, led to the abandonment of this classifier as a potential source of both prediction and inference. The method is also hampered by missing data, which is common during the melt season due to an increase in cloud cover [137]. The binary classification tree achieved peak accuracy in week 12 near 88% while utilizing 75% of available data, with decreasing accuracy as the melt season progressed. This technique was also hampered by the presence of missing data.

The random forest technique achieved peak accuracies up to 96% in weeks 25-35 with a forest of 100 trees while using 75% of available training data. Varying the number of trees in the forest served to increase accuracy initially, but the addition of more trees in the forest did not further increase classification accuracy. The random forest was capable of accurate predictions during the melt season due to its use of surrogate splits. A fitting neural network with two hidden layers of size ten achieved peak accuracies of 94% near week 30, and outperformed its pattern recognition network counterpart by up to 2% in some cases.

The addition of more years during training served to lower the accuracy in all three test cases

as compared to the same-year accuracy. The same-year case outperformed prior-ten and prior-five year cases by up to 15% in some experiments. We observe that the addition of more years during training does serve to increase classification accuracy in most cases however, as the prior-ten cases outperformed the prior-five cases in most tests. When predicting summer extent the same year of data isn't available as truth data, so this implies that problems of prediction would most benefit from inclusion of as many years of data as are available. Studies focusing on inference of predictor importance should be trained using the same year of data to maximize prediction accuracy.

We observe that the random forest and fitting neural network techniques are the optimal classifiers in this study. This is due to their high prediction accuracies and ability to work in the presence of missing data. Peak accuracies were achieved during the melt season, but the techniques yield high accuracies near week 20 when melt typically starts. Future work utilizing these data for forecasting summer ice extent should be refined by further testing and optimization of these techniques. Additional data sources such as those considered in [64, 65] and others may also serve to improve the prediction accuracy of these techniques.

It is important to note that the values considered in this chapter are weekly snapshots of conditions on the ice. Variables like downwelling radiation and surface temperature can exhibit large changes on a weekly basis, which makes these values poor predictors of whether a given parcel will survive a melt season. Future work using these techniques should address this issue by including cumulative values when available, such as total cumulative absorbed radiation. This would strengthen conclusions made using these data in prediction, and would likely increase predictive accuracy. Another variable that would provide context for future work is cumulative convergence/divergence. These values would provide an additional means through which changes in concentration and survival could be studied.

3.8 Summary

This chapter sought to demonstrate the implementation and testing of four discriminative classifiers with the aim of determining if there was an optimal technique for predicting sea ice

survival in the pan-Arctic domain. Further, it explored optimal weeks for prediction of survival, and sought to determine if the addition of more years of data would serve to increase classification accuracy.

Of the four techniques studied, we observe that the random forest and neural network techniques are the most optimal for prediction of summer melt, with prediction of melted/survived parcel class accuracies in the 80-95% for predictions made during the spring and summer. We have also obtained optimal tuning parameters for the neural network and random forest that are applied to the predictor importance estimates explored in Chapter 4. We find that peak survived/melted class prediction accuracies for the random forest and neural network were obtained from the start of the melt season to the end of melt, with peak accuracies later in the season. These techniques have proven to be stable in the presence of missing data, and we observe that the inclusion of training data from additional prior years is desirable.

Chapter 4

Relative Importance of Predictors in the Statistical Learning Analysis

This chapter details the methods through which pan-Arctic predictor importance values were determined from predicting whether parcels melted or survived, and the resulting predictor ranks. The work detailed herein utilizes the random forest and neural network techniques that were introduced in Chapter 3 and the data product described in Chapter 2. Predictor ranks for the neural network are determined through one-factor-at-a-time (OFAT) removal by using prediction accuracy as a benchmark. Random forest predictor ranks are determined through the use of the Gini impurity index, which provides readily available predictor ranks. These results are followed by a discussion of the predictor ranks obtained throughout the melt season, in which we determine that latitude, thickness, and albedo have the greatest impact on survival in the pan-Arctic case. The techniques used in this chapter are utilized as a part of the analysis presented in Chapter 5.

This chapter is partially adapted from *Tooth, M.; Tschudi, M.; and Matsuo, T., 2018* [160].

4.1 Introduction

Chapter 3 focuses on using statistical learning techniques for classification and prediction, and the optimal times and parameters for those predictions. Additional techniques enable the inference of predictor importance from several of the classifiers discussed. These inferred values of predictor importance can provide additional context for Arctic sea ice analyses through providing quantitative measures of the importance of specific variables in determining whether a given 12.5km by 12.5km EASE-Grid parcel will melt or survive during the summer melt season.

The random forest and fitting neural network techniques were chosen for use in this inference study due to their performance as outlined in Chapter 3. The random forest provides a means through which predictor importance can be determined through the use of the Gini index, while the neural network is studied through the use of OFAT analysis. The objective of this study is to determine if there is a predictor that accounts for the bulk of the decision boundary in the statistical learning techniques, and if there is another predictor that does not account for very much of the boundary. The study also seeks to determine the effect of the removal of the least and most important predictors on survival classification accuracy.

4.2 Predictor Importance in the Random Forest

Determination of predictor importance in the random forest is done through the use of the Gini index, which is a measure of the change in risk due to splits on each predictor considered in the forest. The Gini Impurity Index is defined as:

$$1 - \sum_i p_i(1 - p_i) \tag{4.1}$$

where classes $i \in [1, 2]$ represent the melted and survived classes, and p_i represents the observed fraction of data in class i at a node. Predictor importance is calculated at each split by taking the difference between the indices of the parent and child nodes. The resulting values are averaged for each split in the forest that involve the predictor in question to determine a Gini index for that predictor.

4.2.1 Methods: Predictor Importance in Random Forest

The Gini index is calculated for each predictor in a random forest of 100 trees from 2002-2016. A forest size of 100 trees was chosen based on the results presented in Chapter 3. The resulting values are used to compute predictor ranks in each year, and the mean predictor rank in the overall 2002-2016 case for each of the seven predictors is considered. Cases in which various factors are removed from training and testing are also investigated to determine if the removal of the most

or least important predictors had a strong impact on classification accuracy. These cases remove up to the three most important and three least important factors from the analysis, and use the resulting data to form predictions that are compared against a baseline case that includes all of the predictors. Additional Gini indices are computed for each week 10-35 of years 2002-2016. This forms a weekly set of ranks from which changes in predictor importance throughout a given year are investigated.

4.2.2 Results: Predictor Importance in Random Forest

Yearly predictor importance ranks are determined using the Gini index for 2002-2016 (Tables 4.1 & 4.2). Average predictor importance rank values are also computed, where a score of 1 indicates that the predictor is most important in a given year.

During week 20 the latitude and thickness of a sea ice parcel are ranked as the most important predictors in the forest, with average ranks of 1.4 and 2.0 respectively. Albedo is also highly ranked, with an average rank of 2.9. Parcel age, IST, and downwelling shortwave radiation are ranked lowest by the forest during this week. At week 30 parcel albedo and latitude are ranked as the most important predictors in the random forest, with an average rank of 1.6 and 1.9 respectively. Parcel thickness is also highly ranked, with an average predictor rank of 2.5. IST and age are the lowest ranked predictors in the forest, with ranks of 6.8 and 5.8 respectively.

Weekly mean predictor importance values for years 2002-2016 are also calculated (Figure 4.1). Latitude is typically ranked as the most important predictor between weeks 10 and 25. Parcel thickness is the second highest ranked predictor in weeks 10 to 25, with a brief period of being more highly ranked in the forest than latitude. Albedo exhibits a steadily rising prediction importance leading into week 25 as the melt season continues to degrade the ice surface, and it continues to be ranked within the top three predictors from week 25 to 35. While there is separation between IST, longwave, shortwave, and age at the start of the sample period, the importance of IST steadily declines throughout the study period as surface temperatures reach uniform melt temperature. The difference between longwave and shortwave decreases toward week 20, with longwave continuing to

be ranked higher than shortwave in weeks 20-35. The importance of parcel age is reported as being low consistently throughout the study period until IST falls below it in week 20.

Test cases that removed up to the most important and least important three variables are formed to examine the impact of removing those predictors on classification accuracy (Table 4.3). A baseline test with all predictors included in the forest is also performed, with a resulting week 30 accuracy of 96.2% for years 2002-2016. The removal of all predictors except albedo leads to a decrease of 3.2% from the baseline accuracy of 96.2%. The inclusion of the next most important predictor - latitude - leads to a further increase to 94.9% accuracy. The inclusion of the third most important predictor - thickness - yields an accuracy of 0.5% greater than in the baseline case.

In the case where all predictors are included excepting the lowest ranked predictor - IST - we observe an increase of 0.3% accuracy as compared to the baseline case. Further elimination of sea ice age increases the accuracy to 0.6% above baseline, and 0.1% above the case of including only the top three predictors. The last test case also removes downwelling longwave. This leads to a slight decrease in accuracy to 0.4% above baseline.

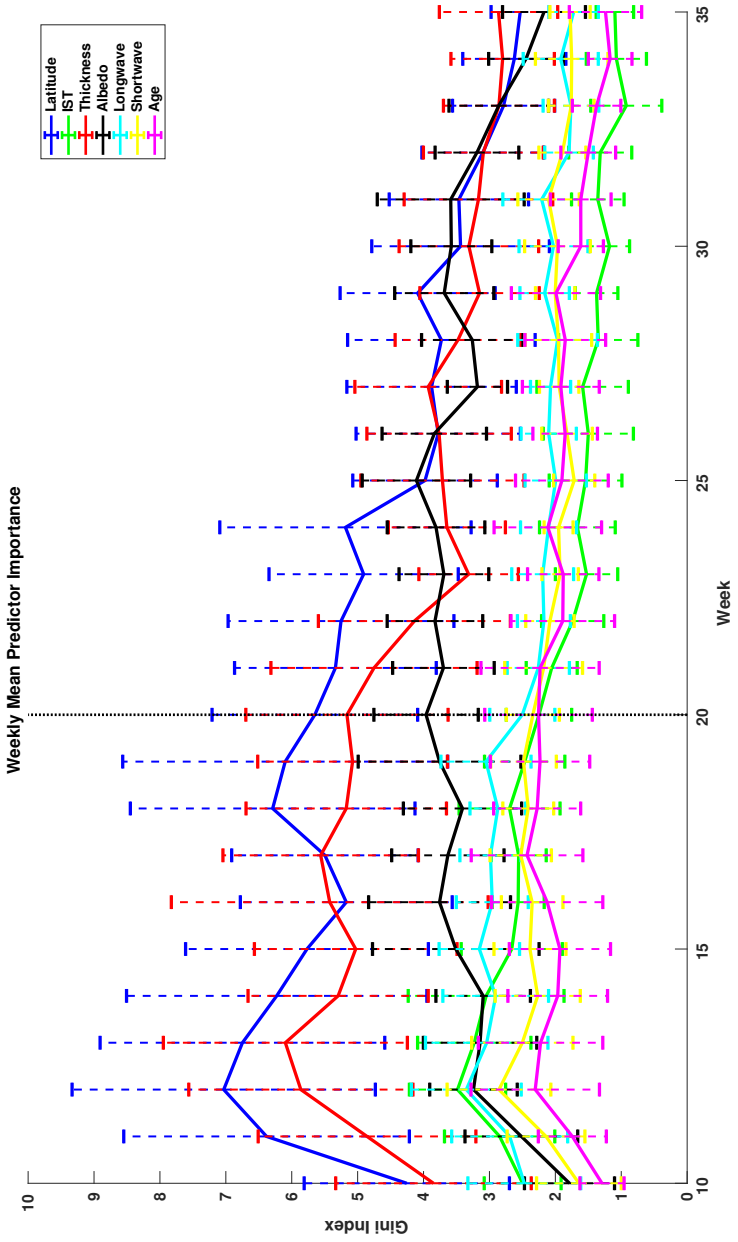

Figure 4.1: Weekly mean predictor importance in the random forest for years 2002-2016

Table 4.1: 2001-2016 pan-Arctic week 20 Gini index rank for predictors in the random forest classifier

Year	Latitude	IST	Thickness	Albedo	Longwave	Shortwave	Age
2002	1	2	3	4	5	6	7
2003	2	7	1	3	6	5	4
2004	1	6	3	2	5	4	7
2005	1	7	2	3	5	6	4
2006	2	7	1	3	6	5	4
2007	1	3	2	4	5	6	7
2008	1	6	2	3	4	7	5
2009	1	6	2	3	4	5	7
2010	1	7	2	3	5	6	4
2011	1	4	2	3	6	7	5
2012	1	6	2	3	5	4	7
2013	2	6	1	4	5	7	3
2014	2	7	4	1	5	3	6
2015	1	6	2	3	4	5	7
2016	2	4	1	3	5	6	7
Avg	**1.4**	**5.7**	**2.0**	**2.9**	**4.9**	**5.4**	**5.6**

Table 4.2: 2001-2016 pan-Arctic week 30 Gini index rank for predictors in the random forest classifier

Year	Latitude	IST	Thickness	Albedo	Longwave	Shortwave	Age
2002	2	4	3	1	7	5	6
2003	3	7	2	1	5	4	6
2004	2	7	3	1	5	4	6
2005	2	7	3	1	5	4	6
2006	1	7	3	2	6	4	5
2007	3	7	1	2	5	4	6
2008	2	7	3	1	5	4	6
2009	2	7	3	1	6	5	4
2010	2	7	1	3	5	4	6
2011	3	7	2	1	4	5	6
2012	1	7	3	2	6	5	4
2013	2	7	4	1	3	5	6
2014	1	7	2	3	4	5	6
2015	3	7	2	1	4	5	6
2016	1	6	2	3	5	4	7
Avg	**2.0**	**6.7**	**2.5**	**1.6**	**5**	**4.5**	**5.7**

Table 4.3: Table of variables included in random forest test cases and their mean 2002-2016 accuracies. 1 indicates that the variable was included in a test, while 0 indicates that it was excluded.

Case	**Latitude**	**IST**	**Thickness**	**Albedo**	**LW**	**SW**	**Age**	**Accuracy %**
1	1	1	1	1	1	1	1	96.2
2	1	0	0	0	0	0	0	93.0
3	1	0	0	1	0	0	0	94.9
4	1	0	1	1	0	0	0	96.7
5	1	0	1	1	1	1	1	96.5
6	1	0	1	1	1	1	0	96.8
7	1	0	1	1	0	1	0	96.6

4.3 Predictor Importance in the Neural Network

Predictor importance in the neural network was determined through the use of OFAT analysis [28]. This led to changes in prediction accuracy that were utilized as proxies for the readily available predictor importance measures that were shown for the random forest classifier.

4.3.1 Methods: Predictor Importance in the Neural Network

OFAT Test cases are formed that remove one of the seven predictors from consideration during each case. This work utilizes a fitting neural network with two ten-node layers that is trained on half of each year's available training data. A baseline prediction accuracy is determined utilizing all seven parameters for each year, and is used to compute changes in prediction accuracy with the removal of each factor. The algorithm then repeats the accuracy estimation for each year and every case of removing one factor from the analysis. Changes in prediction accuracy are computed, and ranks are formed from the parameters that have the highest negative impact on prediction accuracy.

Additional multi-factor test cases are also formed. These seek to determine the impact of removing the most and least important factors on classification accuracy for the neural network. Test cases are formed that utilize the ranks described above to remove up to the three most important and three least important factors from consideration during training and testing. The resulting accuracies for week 30 are computed, and are further discussed in the results and discussion.

Additional tests with the neural network seek to determine what changes in predictor importance occurred with time. Change in prediction accuracy is calculated for each OFAT case for weeks 10-35 of years 2002-2016.

4.3.2 Results: Predictor Importance in the Neural Network

OFAT analysis is utilized to determine yearly changes in week 20 & 30 accuracy for each predictor that is provided to the neural network (Tables 4.4 & 4.5). Average change in accuracy

is then calculated for each predictor, and a rank is assigned based on each predictor's impact on the overall classification accuracy. A rank of 1 indicates that the predictor is the most important predictor.

In week 20 we observe that latitude and thickness are ranked as the highest predictors, with net changes in accuracy of -1.09 and -0.69% respectively. Albedo is ranked third, with a change in accuracy of -0.62% that is similar to the change for thickness. Parcel age, downwelling shortwave, and IST are the lowest ranked predictors, with net positive changes in accuracy with their exclusion. At week 30 we observe that latitude is ranked as the highest predictor, with an average change in accuracy that is 1.5 times larger than the next highest predictors. Albedo and thickness are ranked as second and third respectively, but are very close in their average impact on the network's performance. IST is the next highest ranked predictor, with a change in accuracy that is nearly a tenth of the value of the prior two ranked predictors. Downwelling longwave was the last predictor to have the effect of improving mean accuracy, but the change is minimal at -0.01%. Downwelling shortwave and age are the two lowest ranked predictors, with average increases in accuracy with their exclusion.

Test cases that remove up to the most important and least important three variables are formed to examine the impact of removing those predictors on classification accuracy (Table 4.6). A baseline test with all predictors included in the neural network is also performed, with a resulting week 30 accuracy of 94.2% for years 2002-2016. Removing all predictors excepting the most important - latitude - results in a decrease of 6.4% as compared to the baseline. The addition of albedo increases accuracy to 1.2% below baseline. Running the network with the top three variables yields an accuracy that is 0.3% more accurate than the baseline case.

In the case where all predictors are included excepting the lowest ranked predictor - age - we observe an increase in accuracy of 0.1%. Further removal of the next lowest predictor - downwelling shortwave - yields an accuracy equal to the baseline case. The additional exclusion of downwelling longwave yields an accuracy that is 0.2% greater than the baseline case.

Weekly mean predictor importance values for years 2002-2016 are computed for the neural

network (Figure 4.2). OFAT removal of latitude results in the highest change in accuracy for all weeks 10-35. The next highest ranked predictors, thickness and albedo, follow similar trends throughout the study period, with typical changes in accuracy near 0.6-0.8%. All three of the top predictors are grouped similarly, and are well separated from the bottom four predictors.

IST, downwelling longwave, downwelling shortwave, and age exhibit similar behavior from weeks 10-35; with little change in rank between them throughout the study period. Age is typically ranked the lowest, with the network showing improvement in accuracy with its removal in all but three of the study weeks. IST, shortwave, and longwave fluctuate around the zero change in accuracy boundary, with little overall change in their impact on the classification accuracy. All three exhibit positive changes in accuracy during a significant portion of the study period, which indicates that their inclusion in training serves to confuse the neural network.

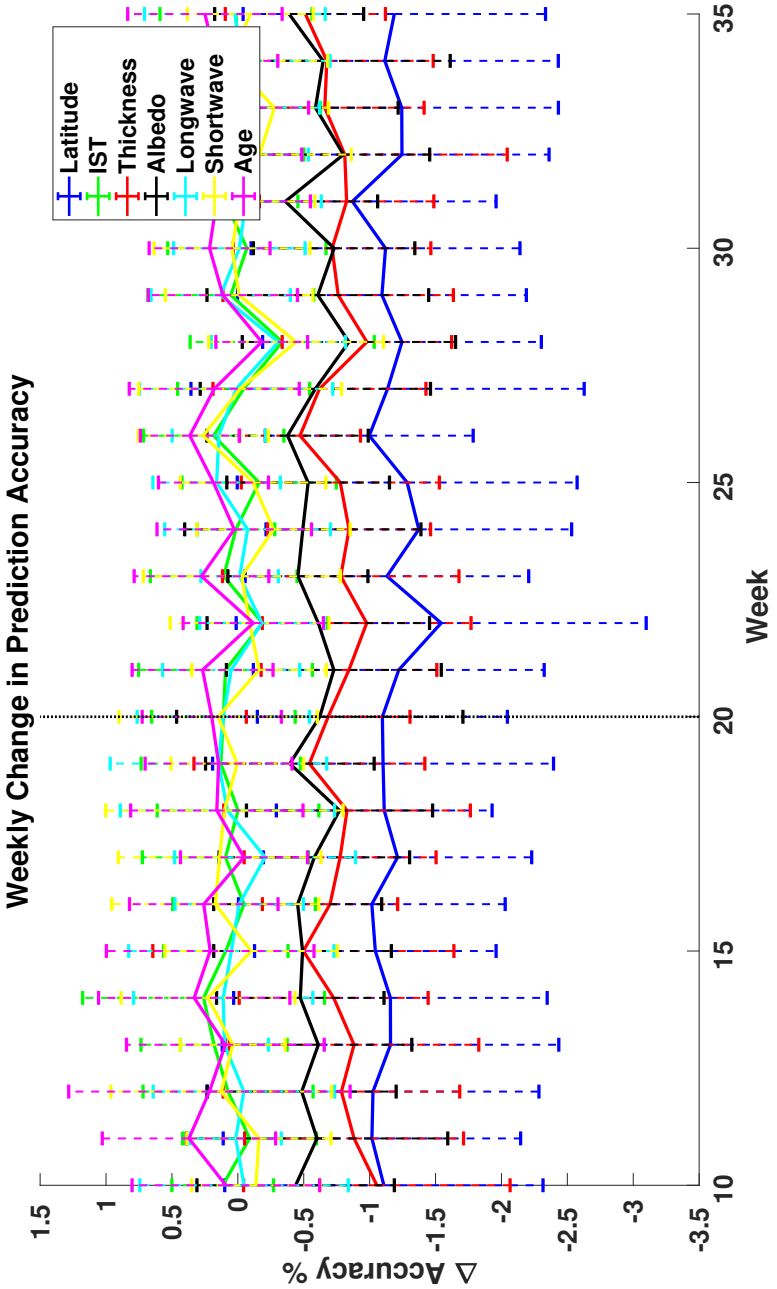

Figure 4.2: Weekly mean predictor importance in the neural network for years 2002-2016

Table 4.4: OFAT change in accuracy when removing specified variable during neural network prediction in week 20

Year	**Latitude**	**IST**	**Thickness**	**Albedo**	**LW**	**SW**	**Age**
2002	0.24	0.12	-0.06	-0.41	0.09	0.3	0.3
2003	-0.36	0	-0.54	-0.21	-0.12	-0.12	0.08
2004	-0.36	0.57	-0.08	0.05	0.15	0.16	0.57
2005	-0.53	0.79	0.21	-0.05	-0.29	-0.45	0.21
2006	-1.15	-0.04	-0.23	-0.48	-0.34	-0.36	-0.08
2007	-0.83	0.31	-1.76	0.35	1.85	0.92	1.3
2008	-1.3	0.02	-0.61	-1.38	-0.64	-1	-0.32
2009	-1.37	-0.36	-1.29	-1.3	-0.37	-0.53	0.24
2010	-1.7	-0.29	-0.75	-0.08	0.79	0.79	1.29
2011	-0.91	-0.48	-1.27	-1.41	-0.3	-0.32	-0.05
2012	-3.78	0.13	0.21	-1.37	0.5	0.76	0.06
2013	-1.25	0.11	-1.16	-2.84	-0.09	0	-0.22
2014	-2.05	-0.73	-1.41	-1.68	0	0.02	-0.57
2015	-0.36	0	-1.05	-0.34	-0.44	-0.04	-0.14
2016	-0.76	1.51	-0.48	1.81	0.87	2.09	0.29
Avg.	-1.09	0.109	-0.69	-0.62	0.111	0.15	0.19
Rank	1	5	2	3	4	6	7

Table 4.5: OFAT change in accuracy when removing specified variable during neural network prediction in week 30

Year	**Latitude**	**IST**	**Thickness**	**Albedo**	**LW**	**SW**	**Age**
2002	0	-0.09	-0.68	-0.18	-0.26	-0.2	0.11
2003	-0.73	-0.28	-0.51	-0.5	-0.47	-0.16	-0.37
2004	-0.74	-0.3	-0.45	-0.42	-0.45	-0.2	0.12
2005	-0.47	-0.03	0.34	-1.1	0.21	-0.97	0.71
2006	-0.92	0.72	0.08	0.1	0.07	-0.3	0.04
2007	-1.36	-1.28	-1.69	-0.84	-0.26	-1.21	-0.06
2008	-0.35	0.84	0.2	-0.22	0.73	0.54	0.89
2009	-0.7	-0.5	-0.32	-0.52	-0.04	0.77	0.4
2010	-1.52	0.52	-0.73	-1	0.98	0.62	0.77
2011	-0.35	-0.52	-1.04	-1.28	-0.17	0.19	0.23
2012	-3.03	0.46	-0.25	-1.87	-1.03	0.36	-0.11
2013	-1.51	-0.95	-1.29	-1.74	-0.04	0.22	-0.26
2014	-2.23	0.33	-0.93	-0.79	0.53	-0.18	-0.09
2015	0.25	-0.1	-0.92	-0.8	-0.05	0.35	-0.27
2016	-3.15	0.15	-2.53	0.24	0.08	0.83	1.09
Avg.	-1.12	-0.07	-0.72	-0.73	-0.01	0.04	0.21
Rank	1	4	3	2	5	6	7

Table 4.6: Table of variables included in neural network test cases and their mean 2002-2016 accuracies. 1 indicates that the variable was included in a test, while 0 indicates that it was excluded.

Case	**Latitude**	**IST**	**Thickness**	**Albedo**	**LW**	**SW**	**Age**	**Accuracy %**
1	1	1	1	1	1	1	1	94.2
2	1	0	0	0	0	0	0	87.8
3	1	0	0	1	0	0	0	93.0
4	1	0	1	1	0	0	0	94.5
5	1	1	1	1	1	1	0	94.3
6	1	1	1	1	1	0	0	94.2
7	1	1	1	1	0	0	0	94.4

4.4 Discussion

In the random forest results we observe that latitude, ice parcel thickness, and albedo are the highest ranked predictors for separating surviving and melted parcels. Latitude and thickness exhibit the greatest Gini indices prior to the onset of melt, and continue to be highly ranked throughout the remainder of the melt season. This is likely caused by more favorable surface conditions at higher latitudes and the greater mass of ice that must be ablated to melt a thicker parcel. The Gini index of albedo increases during the summer melt season to be nearly equal to the indices of latitude and thickness. This growth in the relative importance of albedo is a result of the sea ice albedo feedback, as parcels with lower albedos in the summer are susceptible to additional melt and continuation of the feedback cycle [27, 121].

We note that, while downwelling longwave and shortwave radiation are known to play an important role in melt onset [64, 65, 109, 137], their Gini indices and ranks are fairly low in comparison to other variables. While springtime radiative anomalies and related processes are well linked with melt onset [65], these effects are difficult for the random forest to discriminate. Parcels that melted during the summer had greater spring ISTs than those that survived summer melt, which leads to a higher Gini index in the spring than in the summer. The Gini index of IST in the summer is low due to both the melted and surviving populations of parcels being at melting temperature. While sea ice parcel age is recognized as being linked with ice thickness and survival [98, 170], we observe that the Gini index of parcel age is comparatively low. We do observe the expected link between age and ice thickness described in [170] in these data however (Figure 4.3). The low Gini index for age is likely a result of the thickness data accounting for the increased chance of survival for older ice.

In the neural network we observe that latitude has the greatest impact on prediction accuracy in the OFAT analysis for weeks 10-35. Thickness and albedo are the next highest ranked predictors, with similar changes in prediction accuracy throughout the year. The OFAT analysis appears to fail to sense seasonal variation, as each variable shows little change in $\Delta Accuracy$ throughout the

year. The order of ranks agrees well with the random forest ranks however, which suggests that latitude and thickness are still the main controlling factors in classifying a parcel as melted or surviving.

We observe that downwelling longwave/shortwave and IST have an equal impact on classification accuracy throughout the year in the neural network, with little seasonal change. The age of parcels serves to confuse the network, as removing it has a positive impact on classification accuracy. These predictors likely have a negative effect on the neural network accuracy due to the relatively small separation between the melted and surviving populations in the data space. There are larger separations in albedo, thickness, and latitude that provide more information for the network to make decisions based on, so they have a greater impact on the network's accuracy.

We note that the removal of the lowest ranked variables during training served to increase the predictive accuracy of both the random forest and neural network as compared to their baseline cases. Peak predictive accuracy was attained for the random forest by excluding the age and IST of parcels, which served to confuse the forest during classification. The neural network performed best when all but the top three predictors - latitude, thickness, and albedo - were removed. This suggests that predictive work using these methods should focus on these variables. Future research may benefit from testing other sources of the lower ranked data products, and other available surface parameters.

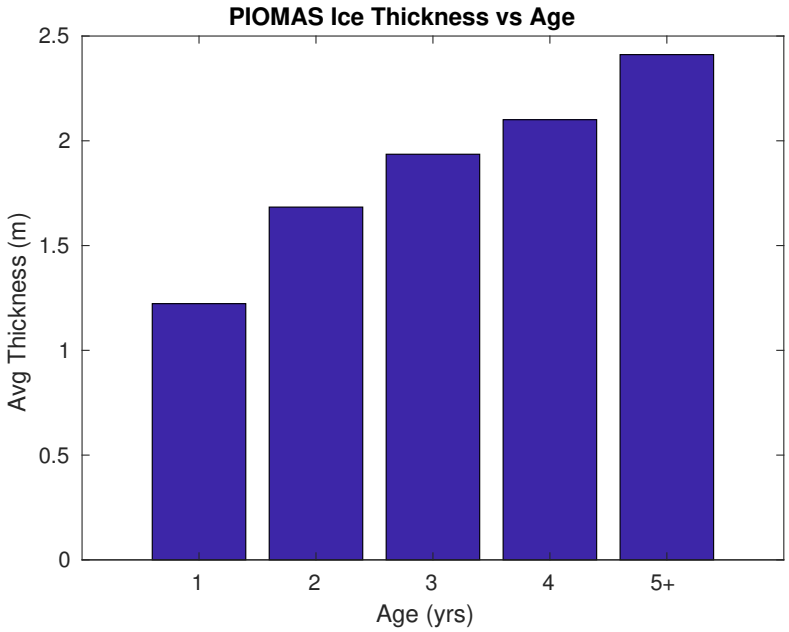

Figure 4.3: Mean PIOMAS thickness for parcels of different age classes

4.5 Summary

This chapter sought to use Gini index values derived from a random forest and OFAT neural network values to determine which predictors are most important in classifying sea ice parcels as surviving or melted at the end of the summer melt season. The work utilized the Lagrangian data product described in Chapter 2 and methods from Chapter 3 to study predictor ranks in these data.

While there is relative agreement between the random forest and neural network, the Gini indices derived from the random forest appear to be most suitable for determining predictor importance. The random forest was able to capture seasonal variation in these data that agree with other sea ice studies that were not present in the neural network results. The random forest method, and the readily available Gini indices derived from it, are valuable tools for future predictive work, operational planning, and investigations of predictor importance [66]. A random forest is used in Chapter 5 in order to compare these results with more traditional statistical methods in a case study of Beaufort Sea ice parcels from 2009-2016.

Chapter 5

Study of Beaufort Sea Ice Parcels from 2009-2016

Arctic sea ice extent has continued to decline in recent years, and the fractional coverage of multi-year sea ice has decreased significantly during this period [129]. The Beaufort Sea region has been the site of much of the loss of multi-year sea ice, and it continues to play a large role in the extinction of ice during the melt season. This chapter presents an analysis of the influence of satellite-derived ice surface temperature, ice thickness, albedo, and downwelling longwave/shortwave radiation as well as latitude and airborne snow depth estimates on the change in sea ice concentration in the Beaufort Sea from 2009 to 2016 using the Lagrangian tracking database described in Chapter 2. The predictor importance values described in Chapter 4 are also derived for these variables. Results from this analysis indicate that parcels that melt during summer in the Beaufort Sea reside at lower latitudes and have lower ice thickness at the beginning of the melt season in most cases. The influence of sea ice thickness and snow depth observed by IceBridge offers less conclusive results, with some years exhibiting higher thicknesses/depths for melted parcels. Parcels that melted along IceBridge tracks do exhibit lower latitudes and ice thicknesses, however, which indicates that earlier melt and breakup of ice may contribute to a greater likelihood of extinction of parcels in the summer.

This chapter is partially adapted from *Tooth M. and Tschudi M., 2018* [159].

5.1 Introduction

Arctic sea ice continues to decline in extent as a result of changes in global climate [129]. This reduction in ice contributes to changes in freshwater balances, the region's biosphere, energy input at the surface, and the future hardiness of the ice pack against climatic forcing [146]. A reduction of the extent of multi-year sea ice has coincided with this loss of total ice extent, which has led to weakening of the Arctic ice pack [99]. Further study of the major variables influencing these changes in the Arctic is of significant interest to governments, commercial entities, and regional stakeholders [33, 37, 104].

The Beaufort Sea region of the Arctic in particular has acted as a "sink" for sea ice over the past decade, and it accounts for a significant portion of the total areal loss of sea ice each year [57, 79, 127, 141, 145]. The region is the extinction location for many of the multi-year ice floes that advect from the Canadian Archipelago. The loss of multi-year sea ice further weakens the Arctic ice cover against future warming, as older ice floes are typically thicker and stronger than younger floes [170], which can help increase their odds of surviving the summer melt season. Given this region's relative importance in sea ice loss, it is valuable to explore the factors leading to the survival and/or extinction of ice parcels that reside there.

Through the use of Lagrangian tracking methods, the trajectory of sea ice parcels in the Arctic can be tracked and recorded. These sea ice parcel positions provide a means by which parcels that inhabit a specific region such as the Beaufort Sea can be sorted and analyzed. Further, coincident airborne and satellite data products can be tracked with these parcels over time.

We examine the influence of IST, ice thickness, surface albedo, downwelling longwave/shortwave radiation, and snow depth on the change in ice concentration in the Beaufort Sea from 2009 to 2016. Our analysis utilizes a Lagrangian tracking database developed in Chapter 2 that matches 12.5km by 12.5km EASE-Grid sea ice parcel locations with ice surface temperature, albedo, and downwelling longwave and shortwave radiation from satellite-based datasets, and ice thickness from a sea ice model [158]. Furthermore, we compare individual sea ice parcel ice thickness and snow

depth as estimated along flight tracks during NASA's Operation IceBridge [82]. The use of Lagrangian tracked data provides a means by which changes in the health of individual sea ice parcels can be observed over time, with additional initial conditions provided by the airborne observations.

5.2 Study Area and Data Sources

The Beaufort Sea study region mask utilized in this study is shown in Figure 5.1. This region contains roughly 6000 12.5 km EASE-Grid cells that formed the study area for this analysis. The landward boundaries of the study region were also conservatively masked in order to prevent the inclusion of parcels that made landfall in this analysis. Sea ice parcel positions were compared against the final Beaufort Sea region mask to determine weeks in which they were present during the study period.

This analysis utilizes six surface products derived from satellite measurements that are built into the Lagrangian tracking sea ice product described in Chapter 2 and [156, 158]. Additional airborne data from NASA's Operation IceBridge campaigns were also incorporated into the study [82]. Data from CULPIS-X and BESST were also considered, but they proved unsuitable for this analysis. Further information about these instruments is included in Appendix F.

5.2.1 Lagrangian Tracking Product

The database of pan-Arctic Lagrangian tracked ice parcel locations with ancillary sea ice property datasets described in Chapter 2 was utilized for this study [156, 158]. The database provides weekly sea ice parcel positions in the Arctic that were used to determine whether particular parcels inhabited the study region. Parcel positions are determined using EASE-Grid sea ice motion vectors [163,165]. This study utilized weekly snapshot values of EASE-Grid Sea Ice Motion Vectors [163, 165], MODIS Terra IST [50], SSM/I and SSMIS ice concentration [19], Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) ice thickness [136, 182], Extended AVHRR Polar Pathfinder (APP-X) albedo, APP-X shortwave up/downwelling radiation, and APP-X longwave up/downwelling radiation [70, 75]. Additional information about these data products and their

uncertainties is provided in Appendix C.

5.2.2 Operation IceBridge Data

Airborne data from spring NASA Operation IceBridge Campaigns were also incorporated into this analysis for years 2009-2016 [82]. The 40m level-4 airborne data were obtained from the National Snow and Ice Data Center (NSIDC), and are further described in their related documentation page [82]. The data include KT-19 infrared pyrometer IST values [78], snow depth estimates retrieved using the University of Kansas' snow radar [63, 86, 133], Digital Mapping System (DMS) derived open water concentrations [30], and Airborne Topographic Mapper (ATM) derived thickness and freeboard estimates [15, 31, 85]. Additional information about these data is provided in Appendix C.

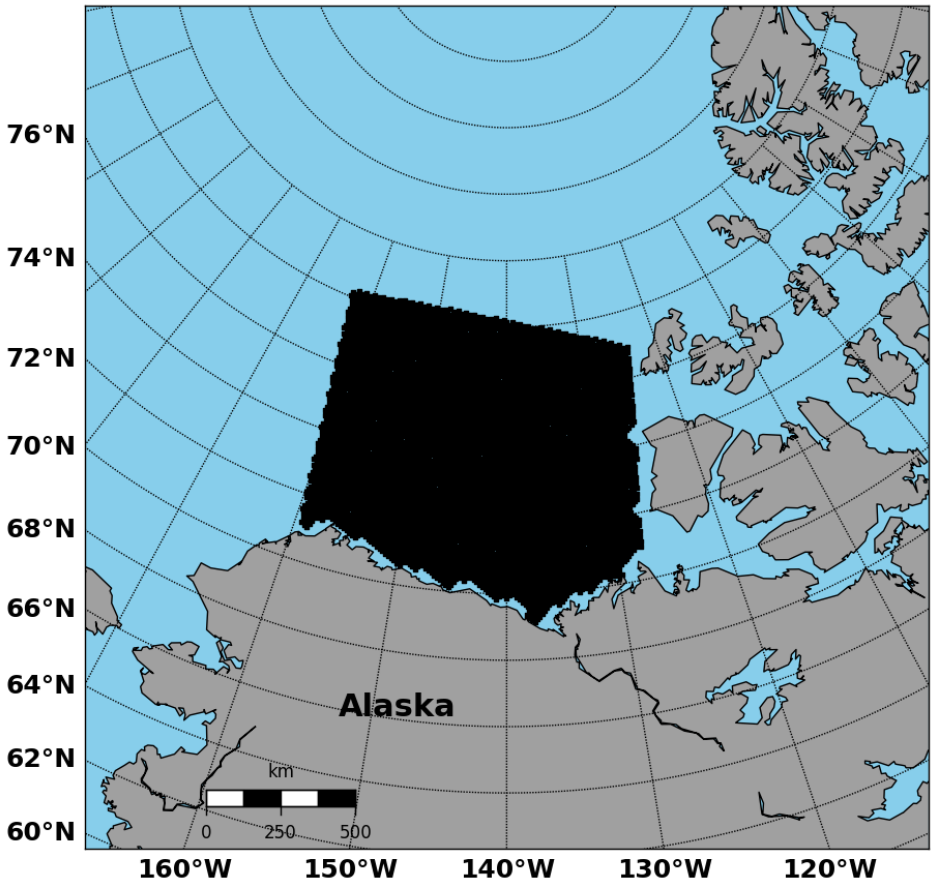

Figure 5.1: Beaufort Sea study region shaded in black.

5.3 Methodology

A set of sea ice parcel tracks for 2009–2016 was generated by searching the parcel database files for parcels that resided in the Beaufort Sea study region during any part of each analysis year. Parcels were sorted into two categories containing those that melted during the study year, and those that survived that year's melt season through the use of a melting threshold of 15% sea ice concentration. Additional checks removed parcels that were considered lost when they coincided with the shore of a land-mask, as these parcels could introduce potentially misleading results into the analysis. The result was a set of weekly observations for parcels that inhabited the Beaufort Sea study area during the years of study. An example of the behavior of the populations in 2009 demonstrates some of the general relationships between the melted and surviving populations present in this data set, such as lower average albedos and ice thickness in the melted parcels (Figure 5.2).

Weekly averages for values of IST, concentration, thickness, albedo, latitude, and shortwave/longwave energy input were then computed for both the melted and surviving populations of parcels, with sample sizes shown in Table 5.1. These values were used to determine Pearson Correlations (measures of the linear relationship) between major variables and change in sea ice concentration. The results of this branch of the analysis are further discussed in Section 5.4.

The Operation IceBridge airborne data were processed by converting the provided latitude and longitude coordinates to EASE-Grid coordinates through the use of the Python Basemap package. Once grid coordinates were obtained, the Operation IceBridge data points were compared against the Beaufort Sea region mask to determine which observations were obtained in the study region. These observations were binned and averaged for each cell with data, and were stored in separate files for use in generating statistics.

The IceBridge flights available in the Beaufort Sea region were sorted by week, and combined to form weekly observation files for each year of study (Table 5.2). The data used in this analysis included averages of snow depth, concentration, IST, and thickness for each overflown cell along the tracks shown in Figure 5.3. Parcels with coincident Operation IceBridge data during the weeks of

Table 5.1: Number of melted and surviving parcels analyzed for each year of study.

Year	**Surviving Parcels**	**Melted Parcels**	**Total**
2009	967	1204	2171
2010	94	1116	1210
2011	583	1776	2359
2012	394	1809	2203
2013	946	1263	2209
2014	982	934	1916
2015	636	1966	2602
2016	518	1421	1939

observation were separated into their own distinct melted and surviving populations for the second thrust of this analysis, which is further described in Section 5.5.

Table 5.2: Weeks of IceBridge data for years 2009–2016.

Year	Flights Over Beaufort Sea	Week(s) Covered
2009	2	14
2010	3	14 and 16
2011	2	11 and 12
2012	3	11 and 12
2013	2	12 and 13
2014	6	11 and 12
2015	5	13 and 14
2016	3	16 and 17

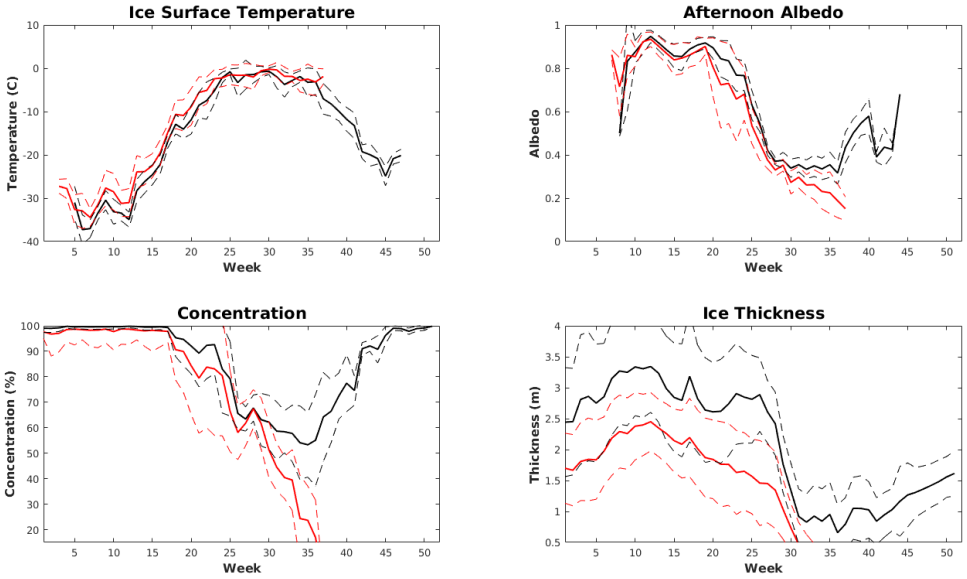

Figure 5.2: Example plots of four studied variables for melted (**red**) and surviving (**black**) parcels during 2009. Dotted lines indicate one standard deviation away from the main curve.

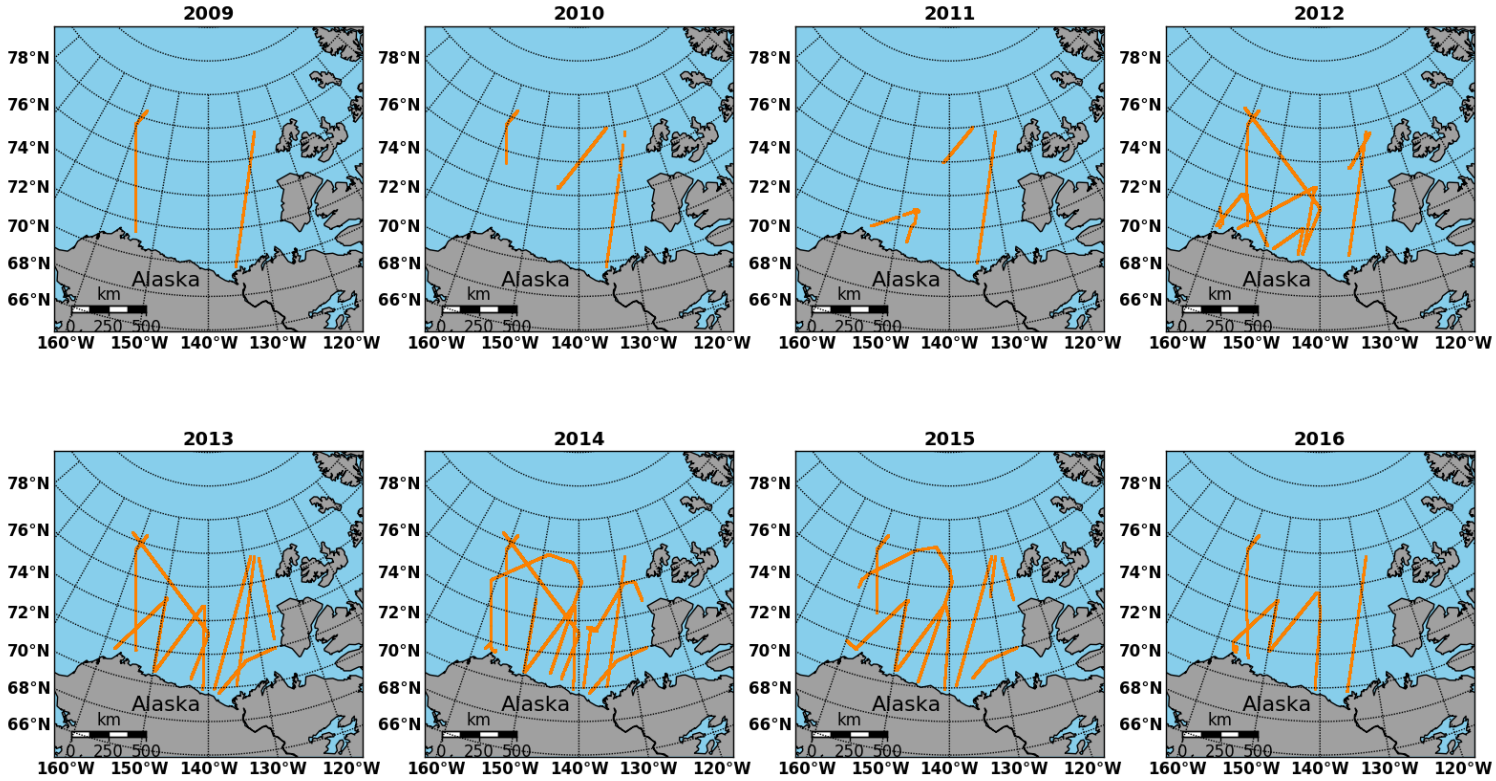

Figure 5.3: Operation IceBridge flight track sections utilized for 2009–2016.

5.4 Results of Analysis of Satellite Data

5.4.1 Parcel Ages

The percentages of parcels in each age class (1–5+ years) in the Beaufort Sea from 2009–2016 are shown in Figure 5.4. During years 2009 and 2010, the studied ice parcel population was comprised of approximately 30% old ice (5+ years) that melted. Twenty-five percent of the studied population in 2009 was older ice that survived the transit through the Beaufort Sea, while only 2% of parcels in 2010 were surviving older ice. Years 2011–2016 show a general trend of large losses of first-year sea ice parcels in the Beaufort Sea, with varying losses in the older age classes. In 2013 and 2014, there was an increase in surviving ice aged 5+ years (8% and 11%, respectively), but this increase was followed by an 8% increase in melted older ice during 2015. In 2016, the amount of surviving and melted older ice decreased, with a corresponding rise in melted and surviving young ice (1–2 years). Melted first-year ice made up the largest percent share of ice in all years excepting 2009, in which a large section of multi-year ice had advected into, and melted in, the Beaufort Sea. We observe that in most cases a higher fraction of multi-year ice survives over first-year ice.

The percent areal share of older ice in the Arctic has been declining over recent decades as measured by the EASE-Grid Sea Ice Age Product [99,164]. The extent of ice older than four years of age has decreased from 2.54 million km^2 to 0.13 million km^2 from March 1985 to March 2017 [129]. We observe significant loss of older (5+ years) ice during 2009 and 2010 in the Beaufort Sea, with a large reduction in ice that survived the transit through the region between the two years. While there was a slight recovery in surviving older ice during 2013 and 2014, the following years 2015 and 2016 returned to the general trend of more older ice being extinguished than surviving. The Beaufort Sea has accounted for much of the loss of older ice in previous years [99], and these data suggest that it has continued to play a role in the reduction of older ice in the Arctic due to the advection of older ice from the Canadian Archipelago into the Beaufort Sea.

5.4.2 Ice Surface Temperatures

Weekly mean ISTs obtained from MODIS data [50] for the melted and surviving parcels in this study are shown in Figure 5.5. Temperatures for the melted population of sea ice parcels exceeded those of the surviving population for every year of study by an average of 1.8 °C. This gap in mean temperatures persisted until the melt season, during which the ice surface is uniformly at its melting temperature. The warmer IST for melted parcels is likely due to the melted population having lower average latitudes, which typically exhibit higher spring temperatures due to greater downwelling shortwave flux and/or advection of warm air masses [131, 137].

5.4.3 Ice Thickness

Mean weekly sea ice thickness for the melted and surviving parcel populations were obtained from the the Pan-Arctic Ice-Ocean Modeling and Assimilation System (PIOMAS) for years 2009–2016 (Figure 5.6) [136, 177]. Parcels that survived a melt season were 0.5 m thicker on average than their melted counterparts in a given study year. Thicker sea ice parcels should be more likely to survive the melt season due to the larger mass of ice that must be melted to bring the sea ice parcel to the melted threshold of 15%, and their greater ability to resist breakup during collisions. In 2009, we observe an increase in mean thickness for melted parcels near the end of the study period. This is due to a smaller population of thicker parcels surviving at that point in the year.

5.4.4 Mean Latitudes

The mean latitudes of the sea ice parcels, which are shown in Figure 5.7, had a strong influence on the probability of survival in all years of study. Surviving parcels were typically 2–3° higher in latitude than melted parcels prior to the onset of melt. As the melt season progressed, the mean latitude of melted parcels progresses northward with the retreating ice edge, but this is a typical seasonal pattern of ice retreat.

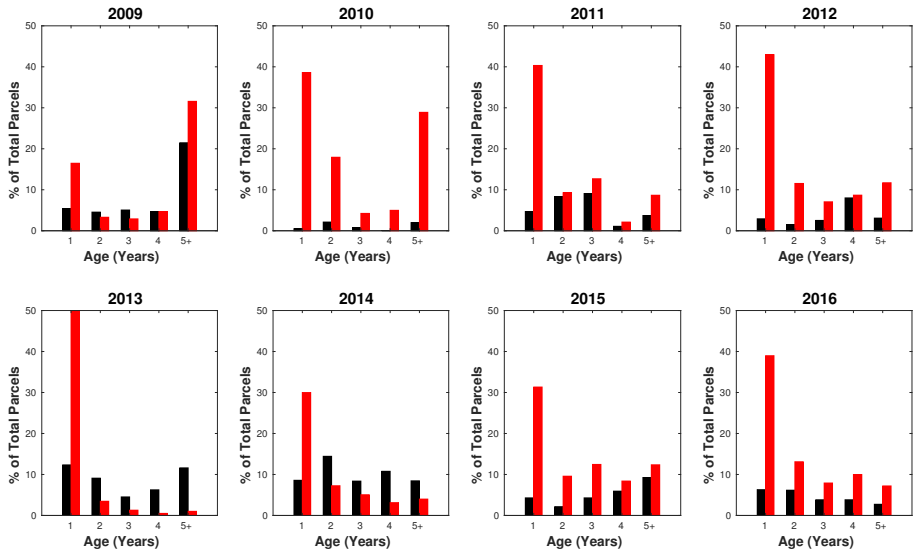

Figure 5.4: Percent of population in age categories 1–5+ years for surviving (**black**) and melted (**red**) ice parcels in the Beaufort Sea during 2009–2016.

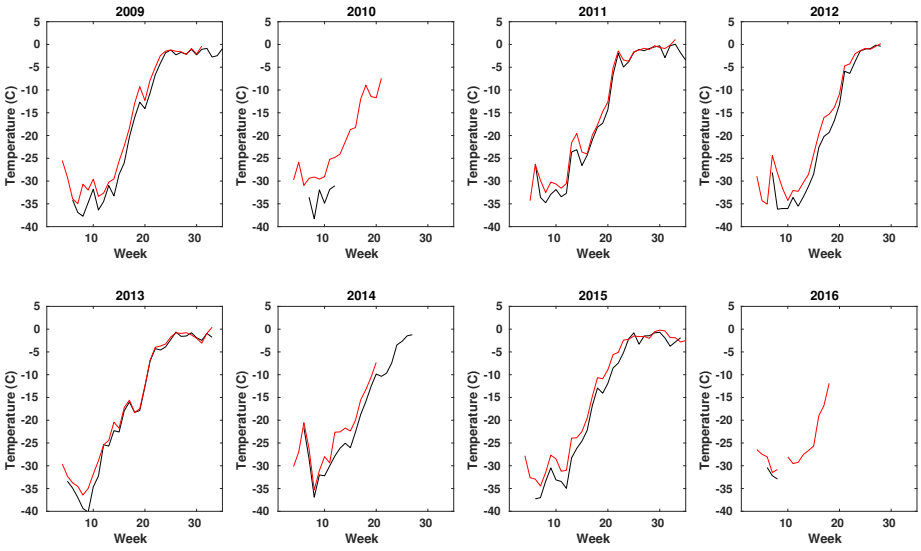

Figure 5.5: Ice surface temperature trends for melted (**red**) and surviving (**black**) parcels during 2009–2016.

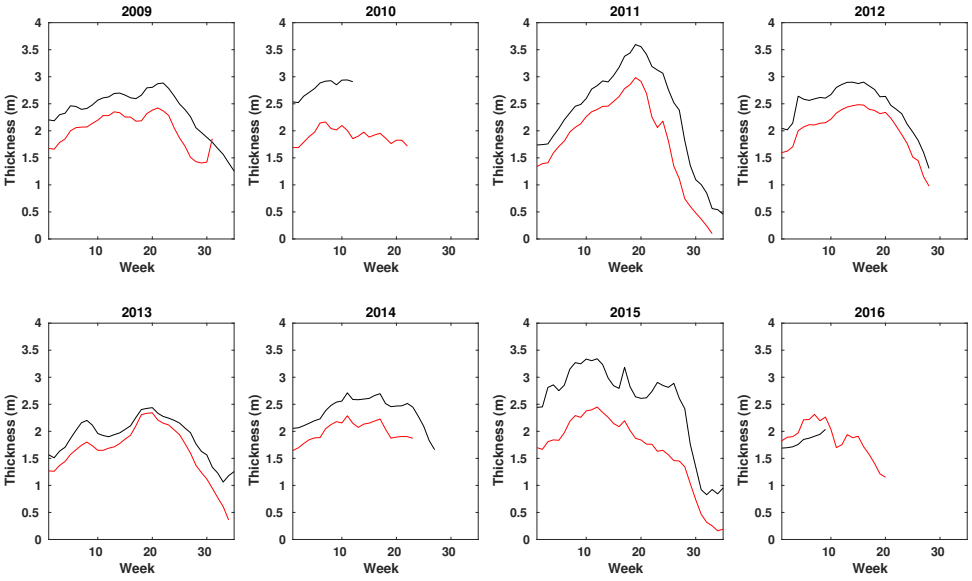

Figure 5.6: PIOMAS model ice thickness of melted (**red**) and surviving (**black**) sea ice parcels during 2009–2016.

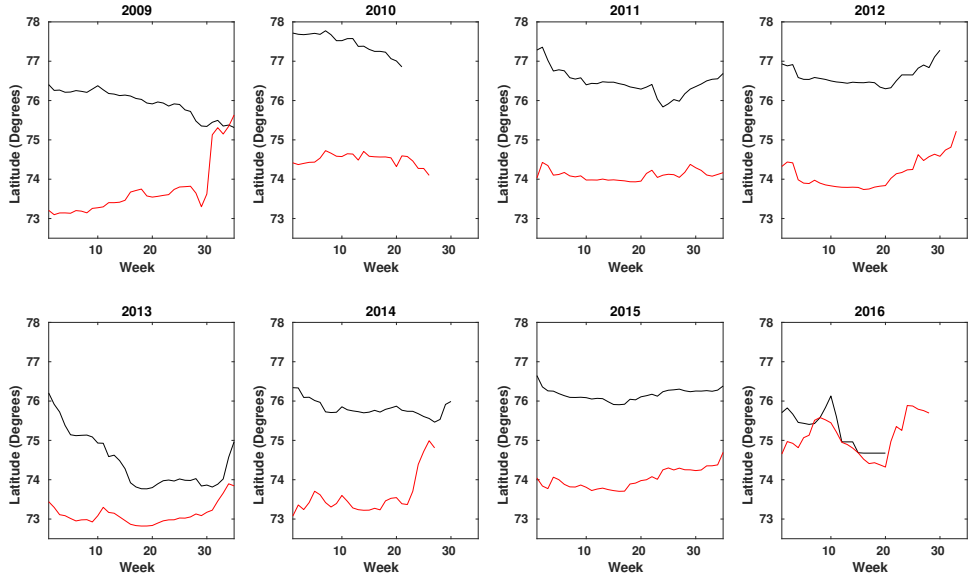

Figure 5.7: Mean latitudes of sea ice parcel populations during 2009–2016. Melted parcels are shown in **red**, surviving parcels are shown in **black**.

5.4.5 Downwelling Radiative Fluxes

Weekly snapshot values of downwelling shortwave and longwave radiative fluxes were obtained from the APP-x dataset [70]. Mean downwelling shortwave radiative fluxes for melted and surviving parcels were very similar for all years in the study period (Figure 5.8). The melted population typically received 10–15 W/m^2 more downwelling shortwave flux during the first part of the year, but after Week 20 the surviving population received up to 20 W/m^2 more downwelling flux. Mean downwelling longwave radiative fluxes were also very similar for both the surviving and melted parcels, with periods of up to 10 W/m^2 greater downwelling flux for the melted parcel population (Figure 5.9). The relationship of greater downwelling longwave flux for melted parcels typically held during later parts of the study years.

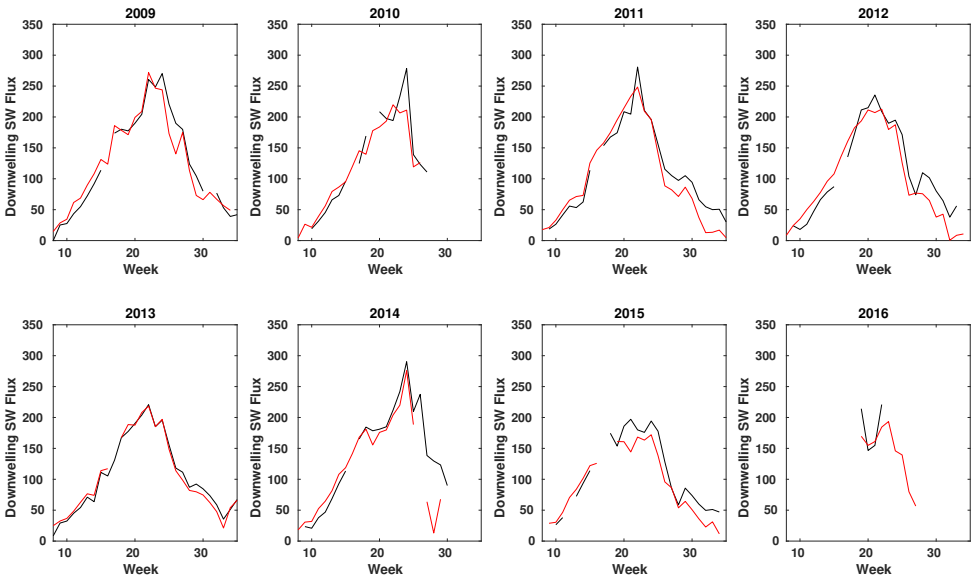

Figure 5.8: Mean downwelling shortwave radiation in W/m^2 for surviving (**black**) and melted (**red**) parcels during 2009–2016.

5.4.6 Surface Albedos

The mean APP-x [70] albedos of the melted and surviving parcels in this study are shown in Figure 5.10. The albedos in the melted populations were lower than the surviving parcel population

means, which led to greater energy input for equal radiative forcing at the surface in the melted population. The mean difference was typically 0.1, which represents a significant difference in available shortwave radiation during the melt season.

5.4.7 Pearson Correlations between Major Variables and Concentration Change

Pearson correlations were calculated to explore the linear relationship between the major thermodynamic variables discussed above, and the change in sea ice parcel concentration obtained from SSMI/SSMI-S [19] for the overall sample population in each year (Table 5.3). Due to the use of total change in parcel concentration as the variable to compare against, positive correlations imply that an increase in the variable will lead to an increase in melt rate, while negative correlations imply that an increase in the variable will lead to a decrease in melt rate. Table 5.3 values in black have high significance, values in blue have slightly lower significance, and orange values have low significance.

Table 5.3: Pearson correlations between studied variables and total change in concentration for 2009–2016. Values with $p \in (0.05, 0.3]$ are shown in blue. Values with $p \in (0.3, 1]$ are shown in orange

Variable	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Temperature	0.60	0.61	0.65	0.47	0.47	0.51	0.76	0.69
Latitude	0.39	0.31	0.52	0.29	0.24	0.38	−0.18	−0.66
Downwelling LW	0.68	0.71	0.77	0.72	0.53	0.77	0.84	0.95
Downwelling SW	−0.67	−0.71	−0.28	−0.26	−0.39	−0.15	−0.10	−0.82
Thickness	−0.75	−0.95	−0.64	−0.69	−0.74	−0.69	−0.64	−0.91
Albedo	−0.93	−0.91	−0.93	−0.85	−0.84	−0.84	−0.87	−0.89

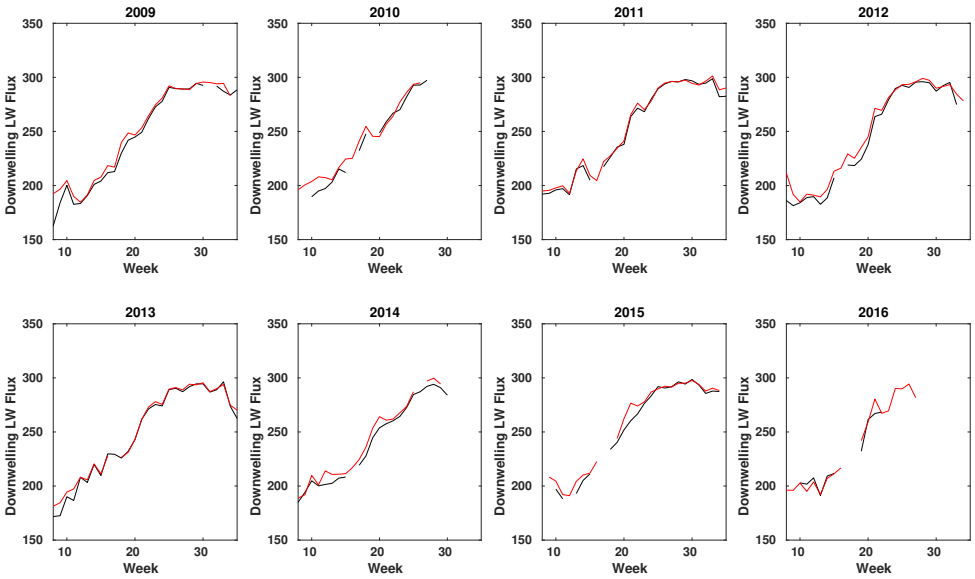

Figure 5.9: Mean downwelling longwave radiation in W/m^2 for surviving (**black**) and melted (**red**) parcels during 2009-2016

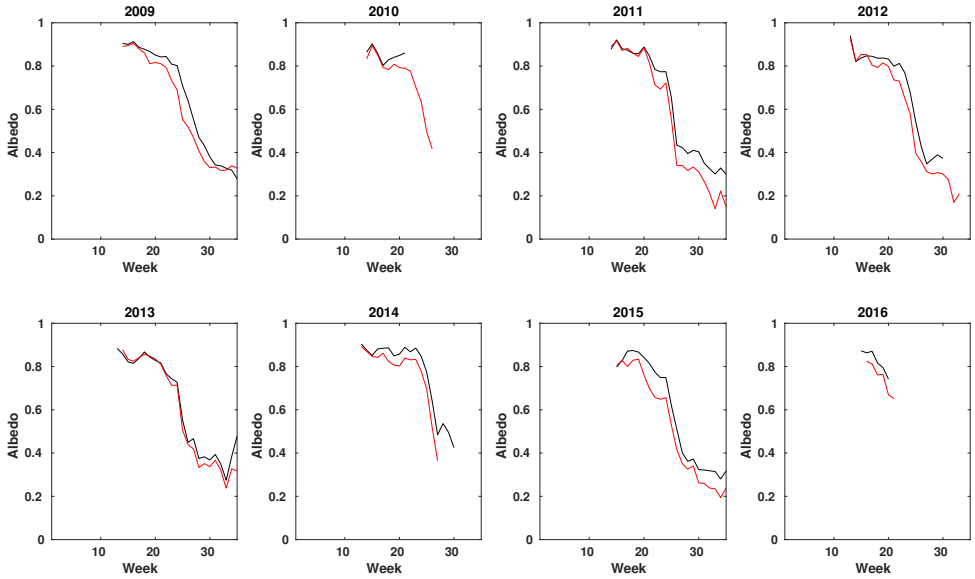

Figure 5.10: Mean albedo for melted (**red**) and surviving (**black**) parcels during 2009–2016.

5.5 Results for Parcels With Coincident Operation IceBridge Data

The Operation IceBridge data analyzed in this study were comprised of twenty-six separate flights over the Beaufort Sea study region during the analysis period. Files containing coincident data for a given week were combined to form weekly observation files for use in computing mean surface conditions for each year (Table 5.4). The year 2009 featured one week of data, while the years 2010–2016 each had two weeks of data to use as a point of comparison (Table 5.2). These data provided initial values for parcels prior to the onset of melt.

Ice surface temperature measurements from Operation IceBridge were available in the region of study for all observation years except 2009, 2011, and 2016 [78]. These data showed an average IST difference between melting and surviving populations of 5.8°C. The relationship of higher average temperatures for melted parcels as compared to surviving parcels held for all years in this analysis. This relationship is also present in the satellite IST means for this subset of parcels. These results match those described in the previous section, where the population of melted parcels endured higher surface temperatures than the surviving population during the spring. IST can change quickly from day to day however, so these results are limited due to their sparse sampling.

Snow depth was also considered in this analysis, as its higher albedo and negative heat flux contribution can help prevent the loss of an ice parcel by delaying melt [123]. Snow depths were greater in the surviving sea ice parcel population for all years except 2010 and 2016. Outside of years 2010 and 2016, the surviving parcels possessed an average of 5 cm greater snow depths than their melted counterparts.

Results from Operation IceBridge ice thickness estimates and PIOMAS model thicknesses yield contrasting results. In the IceBridge data, the melted parcels had an average of 27 cm greater mean ice thickness than surviving parcels in 2009, while melted parcels in 2010 and 2012 were an average of 4.5 cm thicker than surviving parcels. In 2014, the average thickness of surviving parcels exceeded that of melted parcels by 48 cm, and surviving parcels were an average of 10.5 cm thicker than melted parcels in 2015 and 2016. Coincident PIOMAS thickness data for these parcels

indicate that surviving parcels were thicker in every year by an average of 30 cm.

IceBridge concentration data for all years in which they were available indicate that sea ice parcels in both populations were near 100% sea ice concentration. SSMI sea ice concentrations along the flight tracks were also at or near 100%. While lower ice concentrations are possible in some regions during the spring, there weren't parcels lower then 100% present in the IceBridge estimates used in this study.

Table 5.4: Mean Operation IceBridge and satellite derived spring surface conditions of melted (**red**) and surviving (**black**) parcels

Variable	**Source**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Latitude (deg)	OIB	71.98	75.36	73.67	74.38	72.21	72.55	73.34	73.92
Latitude (deg)	OIB	75.73	77.08	76.05	76.78	74.07	75.34	76.27	76.08
IST (C)	OIB	−10.94		−33.48	−21.11	−21.16	−21.49		
IST (C)	OIB	−21.29		−35.69	−25.88	−26.24	−28.18		
IST (C)	MODIS	−29.23	−21.03	−29.64	−33.43	−22.91	−22.90	−22.19	−15.89
IST (C)	MODIS	−33.06	−21.78	−32.03	−34.89	−25.95	−27.01	−26.54	−18.11
Snow Depth (cm)	OIB	29	27	14	16	13	13	22	16
Snow Depth (cm)	OIB	33	24	16	18	25	18	28	11
Thickness (m)	OIB	2.66	2.99	2.00	2.32	1.13	1.957	2.59	2.36
Thickness (m)	OIB	2.39	2.95		2.27		2.44	3.48	2.46
Thickness (m)	PIOMAS	2.17	2.30	2.44	2.60	1.77	1.99	2.29	1.86
Thickness (m)	PIOMAS	2.46	2.51	2.71	2.89	1.80	2.51	3.00	2.00

5.6 Predictor Importance in the Beaufort Sea

Predictor importance ranks for the satellite-based data used in this study are obtained via the use of the Gini index procedure described in Chapter 4. A random forest of 100 trees is used to generate mean week 20 ranks for all years excepting 2010 and 2016, which do not have a large enough surviving population to produce Gini index ranks (Table 5.5).

The week 20 ranks in Table 4 represent the ranking of predictors during or near the onset of melt in the Beaufort Sea. Latitude and thickness are the two highest ranked parameters, with average ranks of 1.8 and 2.3 respectively. These are followed by the age of a sea ice parcel, with a rank of 3.8. Albedo, downwelling longwave, and downwelling shortwave are equivalent in rank during melt onset. IST, which is typically near the melting temperature at week 20, is ranked as the least important predictor.

Weekly values of the Gini index were also produced for weeks 10-35 (Figure 5.11). We observe that latitude and thickness have the highest Gini indices throughout the late spring and summer. These values steadily decline after melt onset, and are nearly equal to the Gini index of albedo later in the summer. The Gini index of albedo declines slightly as the melt season progresses, but it is ranked near thickness between weeks 25 and 35. The Gini index of age remains relatively constant throughout the year, with little change in value. The Gini indices of shortwave/longwave and IST decline steadily from week 20 to 35 as the melt season starts and progresses.

Table 5.5: Week 20 predictor importance in the Beaufort Sea

Year	**Latitude**	**IST**	**Thickness**	**Albedo**	**Longwave**	**Shortwave**	**Age**
2009	3	7	1	4	5	2	6
2010							
2011	1	7	3	5	6	2	4
2012	1	6	2	3	5	7	4
2013	1	7	4	5	3	6	2
2014	2	7	3	4	5	6	1
2015	3	2	1	7	4	5	6
2016							
Avg	**1.8**	**6.0**	**2.3**	**4.7**	**4.7**	**4.7**	**3.8**

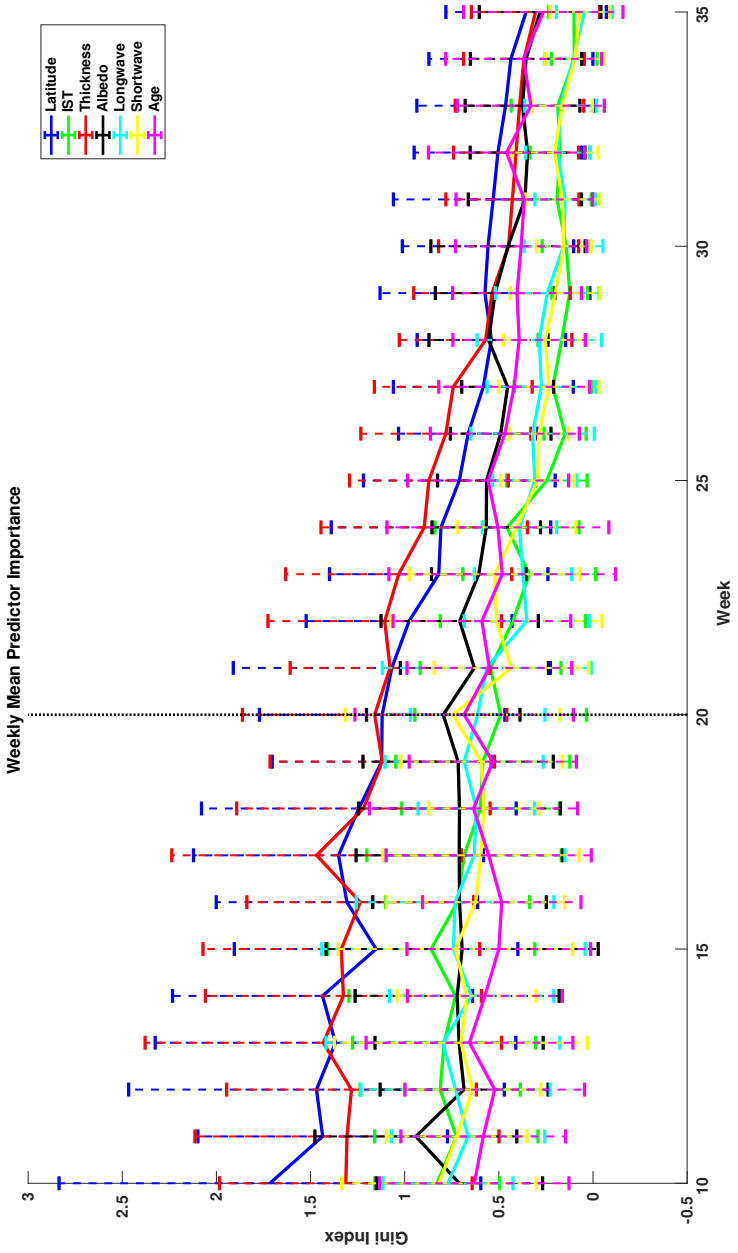

Figure 5.11: Weekly mean predictor importance in the Beaufort Sea

5.7 Discussion

For the satellite data studied from 2009–2016, we observe that albedo exhibited the strongest negative Pearson correlation with change in ice concentration. This is likely a result of the ice albedo feedback [27,121]. The downwelling longwave radiative flux had the highest positive Pearson correlation with change in ice concentration. These results agree well with other studies that have found that greater downwelling longwave radiation in the spring, and other linked processes, contribute to earlier melt onset and lower end of season ice extent [64, 65, 109]. We note that downwelling shortwave flux correlated negatively with change in ice concentration, but this is likely due to the observed seasonal decline in downwelling shortwave radiation (Figure 5.8) while melt continues at the end of the melt season [137]. Other studies have suggested that longwave radiation anomalies have a greater impact on melt onset, while shortwave radiation anomalies act as an amplifying feedback once melt has started [64].

We observe that melted parcels are located at lower latitudes and are thinner than surviving ice parcels prior to the onset of melt. These parcels are more likely to break up due to their lower ice thickness, which further contributes to their extinction due to increased lateral melt from the surrounding ocean [161]. We also observe that first year ice comprised the majority of melted ice in all years except 2009. The predominance of younger ice in the melted population is linked to the thickness of the ice, as older parcels are generally thicker than younger ones [170].

For the subset of parcels that had IceBridge data available during 2009–2016, we observe that spring snow depth may influence the survival of ice parcels in the Beaufort Sea, but these results are inconclusive. Sea ice parcels that survived a summer melt season had an average of 5 cm of additional snow cover when compared against those that did not survive the melt season. Snow depth data from two of the study years exhibited greater average snow depth for the melted parcel population however, so there isn't conclusive evidence that snow depth directly contributes to survival. This inconclusive result is likely due to the high uncertainty of the IceBridge snow depth estimates (Appendix C). The potential importance of snow is linked with our previous discussion

of albedo, as snow on the sea ice surface helps to delay melt onset by protecting the ice from further energy input through its higher albedo [122, 123]. This is limited to the summer melt season however, as snow can dampen sea ice growth during the fall and winter seasons [105].

We observe that IceBridge parcels had mixed results for parcel ice thickness throughout the study period, with half of the years exhibiting greater mean ice thickness in the melted population. The mean latitude of melted parcels was 2.5° lower than that of the surviving parcels, which may have contributed to the extinction of thicker parcels during years in which mean thicknesses in the melted population exceeded those of the survivors. PIOMAS model data for these parcels show that surviving parcels were thicker than melted parcels in all study years. The reported uncertainties for PIOMAS model thicknesses and IceBridge thickness estimates were similar for this study. The difference between the IceBridge and PIOMAS results may be a result of the smaller 40 m footprint of IceBridge observations as compared to the 12.5 km PIOMAS values. This difference in sampling size causes IceBridge tracks to report thicknesses for a much smaller region of ice as compared to the PIOMAS averages over entire EASE-Grid cells, which could lead to the mixed results that we observe for IceBridge thicknesses.

Gini index ranks obtained from a random forest for week 20 and weeks 10-35 also indicate that latitude and thickness are most linked to the survival of sea ice during the summer melt season prior to the onset of melt. The Gini index derived importance of these variables decreases as summer melt progresses, but continue to best separate the data through most of the year. The Gini index of albedo increases as summer melt continues as a result of the sea ice albedo feedback. We do not observe an increase in the Gini index of downwelling shortwave as [65] may indicate. This is likely a result of seasonal changes in downwelling shortwave radiation and decreased albedo during the summer. We also observe that downwelling longwave is more highly ranked than shortwave prior to melt onset, which agrees well with the discussion of longwave driving melt onset earlier in the year in [64, 65]. In contrast with the pan-Arctic results described in Chapter 4, we note that the Gini index derived rank of age is higher than downwelling shortwave/longwave and IST for much of the summer melt season. This suggests that older parcels in the Beaufort Sea are more likely to

survive a summer transit, and is likely a result of the large population of melted first year parcels observed in this study (Figure 5.4). The increased Gini index of ice age may also be linked to older parcels generally being thicker, and more likely to survive a summer melt season [170].

We note that, while this study has demonstrated the use of Lagrangian tracks to study sea ice parcel histories in a region of interest, its results are limited by several key factors. This analysis excludes ocean surface temperatures under the ice, which are an important factor during the melt season [124]. Dynamic variables such as surface winds and their impact on the deformation and further breakup of sea ice were also not considered. In addition, the variables considered in this work, such as snow cover and albedo, are not independent [6], and there are other factors that influence the extinction of ice in the Beaufort Sea. The results from this study do suggest that albedo, initial latitude, and thickness are the strongest predictors of ice survival for the variables studied. Results for snow are less conclusive, but snow on sea ice is an ongoing area of study [17, 87, 105, 123]. Future sources of snow data should help to further clarify its influence on the survival of sea ice parcels in the Beaufort Sea.

5.8 Summary

Major variables influencing the survival of sea ice in the Beaufort Sea during 2009–2016 have been considered through the combined use of the Lagrangian tracking database of sea ice parcels described in Chapter 2 and coincident Operation IceBridge airborne observations. The parcels have been sorted into melted and survived categories, and their weekly average values of IST, albedo, latitude, and downwelling longwave/shortwave radiation have been compared against weekly concentration change. Additional average values of snow depth, thickness, IST, and concentrations estimates from NASA's Operation IceBridge flights during the spring have also been compared.

The results of this study indicate that albedo appears to linearly correlate most with percent change in sea ice concentration, and thus survival, of sea ice parcels in the Beaufort Sea. Parcels that survived the summer melt had consistently higher initial albedos than those that melted. The Pearson correlation between albedo and change in concentration was found to have the strongest

negative correlation of the variables studied, while downwelling longwave radiative flux was most positively correlated with change in concentration.

We also observe that melted parcels were located at lower latitudes earlier in the year, and were thinner than their surviving counterparts. This led to earlier reduction in sea ice concentration that likely contributed to the extinction of these sea ice parcels. Melted parcels with coincident IceBridge data were located at lower latitudes than surviving parcels, but there were several years in which melted parcels had greater mean ice thickness. This is in contrast with PIOMAS model data along the flight tracks, which exhibit greater mean thicknesses for surviving parcels in all years. We conclude that the disagreement between the IceBridge and PIOMAS data is likely due to the limited footprint of the IceBridge observations, which can misrepresent the overall thickness of a 12.5 km EASE-Grid cell due to sampling limitations. Results from IceBridge snow radar estimates are inconclusive, as melted and surviving parcels did not exclusively exhibit higher or lower depths in all cases.

Gini indices derived for the satellite-derived data also indicate that latitude and thickness were valuable predictors for determining if sea ice parcels survived a melt season. These indices dropped as the melt season continued, and were nearly equal to the Gini index of albedo later in the summer. We also observe the relationships between downwelling longwave, downwelling shortwave, and melt onset described in [64, 65]. The Gini index results for parcel age are also interesting, as they may suggest that older parcels are more likely to survive summer melt in the Beaufort Sea.

Additional research with more sources of snow data throughout the spring are warranted. The inclusion of more thermodynamic and dynamic data sources to this analysis would also help provide context for possible breakup of ice and input of additional energy at the surface and from the underlying ocean. The use of statistical learning techniques on additional airborne data may also be of future interest for addressing these relationships.

Chapter 6

Conclusions

This thesis investigated the impact of remotely sensed parameters on sea ice survival during the summer melt season. The extensive work within this thesis lays the groundwork for future Arctic investigation through the use of statistical learning techniques, and enhances the community's understanding of the major factors that drive summer melt in the Beaufort Sea and throughout the Arctic Basin. From a data science standpoint, a new data product has been introduced that provides researchers a means through which parcel histories can be examined in future studies. From an Arctic science standpoint, the major factors impacting sea ice survival in the Beaufort Sea have been examined. The Arctic research community will also benefit from the exploration of underutilized statistical learning techniques that were used in this manuscript. The results from this work confirm findings from previous studies, and provide new areas of focus for future work. This investigation has addressed its overarching research goal by **assessing the impact of remotely sensed sea ice parameters on the survival of sea ice in the summer melt season**. In doing so, the following science questions were answered:

(1) **Which sets of remotely sensed data are both available and appropriate for studying sea ice survival in the Arctic?**

(2) **Does a particular statistical learning technique perform best in short-term prediction of sea ice survival during the summer melt season?**

(3) **How does the importance of predictors change during summer melt, and do**

different statistical learning techniques report different results?

(4) **Which predictors of sea ice survival in the Beaufort Sea during summer are ranked as most important, and how do these results compare to results from other statistical techniques?**

6.1 Major Findings

A data product combining Lagrangian tracks with coincident satellite data was developed in Chapter 2 to address the first research question. The database is comprised of weekly measurements of ice surface temperature, ice concentration, modeled ice thickness, convergence, downwelling shortwave/longwave radiation, and albedo on the 12.5km EASE-Grid. These data products are further described in Appendix C, and the grids they are provided on are described in Appendix D. Data were obtained from 2001-2016, and have been archived at the Pangaea Earth Science Data Publisher [156]. Additional information regarding the creation and formatting of this database is provided in its related publication [158]. The data contained in the parcel tracking database represent a set of remotely sensed properties, processes, and parameters that can be used to study Arctic sea ice survivability. The database does not contain products that describe the ocean surface, lower atmosphere, and cloud cover, but the database has been constructed in a way that allows for future inclusion of other sources of data. This database includes a set of available variables that can be used for the study of sea ice survival, and it formed the basis of the rest of the cases explored in this manuscript.

The short-term predictive capability of four discriminative classifiers was assessed in Chapter 3. Implementations of SVM, binary tree-based classification, a random forest, and a neural network were tested in cases where the classifiers were trained using the Lagrangian track data described in Chapter 2. These cases found that same-year prediction accuracy was greatest for the random forest and the neural network, with accuracies of 96% and 94% respectively. Further testing of the random forest determined that using 75% of available data during training was optimal, with

peak accuracies being obtained using forests with 75 or more trees. The fitting neural network was found to perform best using two ten-node layers while using 50% of available data during training.

Further tests of prediction using the random forest and neural network were performed with training data from prior years to demonstrate their utility in forecasting. These test cases included the prior five and prior ten years of data during training, and compared predictive accuracy with the values obtained from same-year training tests. We observe that training with the prior ten years of data outperforms the prior five case in most cases. This implies that work aimed at maximizing predictive accuracy would benefit from the addition of as much data as possible during training. The tests developed in Chapter 3 imply that, of the statistical learning techniques studied, the random forest and neural network perform the best in short-term prediction of sea ice survival during the summer melt season.

Chapter 4 determined predictor importance using both a random forest and a neural network for the entire Arctic domain. The importance values were determined through the use of OFAT analysis for the neural network, and the Gini index for the random forest. These techniques used the optimal parameters determined in Chapter 3 and were trained using the same year of data as the test year. Weekly importance values were calculated throughout the year, with particular focus on weeks 20 & 30 for years 2002-2016. Both the random forest and the neural network report that latitude, thickness, and albedo are the top three ranked predictors. The initial latitude of a sea ice parcel is best at predicting survival during the spring, and declines with the onset of melt. The parcel thickness follows a similar trend to latitude, and is ranked highly due to the greater mass of ice to melt in thicker parcels. The predictor importance of albedo grows with the onset of melt and continuation of the melt season due to worsening surface conditions and the sea ice albedo feedback. The random forest was more effective than the neural network in separating the importance of downwelling longwave during the spring and melt onset, and the growing relative importance of downwelling shortwave later in the melt season. In addition, the importance of IST declined after melt onset, when both melted and surviving parcels were uniformly at melting temperature. The neural network failed to capture these changes, and did not strongly distinguish

between the lowest-ranked predictors later in the melt season.

Further tests sought to determine if there was an optimal combination of predictors for peak classification accuracy in the random forest and neural network. The results of these tests show that the classifiers perform best when all but the top three predictors - latitude, thickness, and albedo - are removed during training and prediction. Additional analysis of the weekly change in predictor importance determined that the random forest was more capable of capturing seasonal changes in predictor importance, as its results contained more variation in rank during the melt season. As a result, the random forest was chosen for further use in Chapter 5.

Chapter 5 sought to determine which variables most influence sea ice survival in the Beaufort Sea domain through the use of both statistical learning techniques and other statistical techniques. This was accomplished by comparing satellite data with coincident IceBridge tracks in the Beaufort Sea from 2009-2016. Comparison of weekly satellite-derived data for melted and surviving parcels determined that the parcels that survived a given melt season were located at higher latitude and were thicker than their melted counterparts. Gini indices calculated using a random forest for week 20 also ranked latitude and thickness as the most important variables in predicting parcel survival. Pearson correlations between change in sea ice concentration and other parameters indicated that albedo had the strongest negative correlation with change in sea ice concentration. Downwelling longwave was found to have the most positive correlation with change in ice concentration. While the random forest ranked albedo as being less important for prediction in the spring, the Gini index values for albedo were closer to those of latitude and thickness during the melt season. Downwelling longwave was ranked higher in the forest than downwelling shortwave, IST, and age during the spring, but declined during the summer. These results suggest that there is relative agreement between the random forest and the other techniques used in Chapter 5.

For the subset of parcels with IceBridge data we observe mixed results from snow depth and ice thickness data. While melted parcels in the IceBridge subset were typically at lower latitudes than their surviving counterparts, the melted population exhibited higher spring ice thicknesses and snow depths in several years. Coincident PIOMAS data for these parcels agree with the broader

satellite results where surviving parcels have greater spring thicknesses. The difference between the IceBridge and satellite-based measurements is likely a result of the difference in their footprint sizes. The results of this study indicate that initial latitude and ice thickness prior to melt onset stand out as the most important predictors in determining summer melt season survival.

6.2 Potential Future Advancement

These results provide both data and techniques that may prove valuable in future Arctic studies. The Lagrangian data track product described in Chapter 2 provides the community with an accessible way to track sea ice parcel histories over time that can be used in a variety of applications. Through the use of the unique parcel IDs, the changing properties of parcels can be viewed in context throughout that parcel's lifetime in the Arctic. Additional work could investigate how summer melt conditions impact spring parcel properties during the next year. In particular, correspondence with a group of Arctic researchers indicates that there is interest in using these data for the tracking of polar bear dens. Further study of how summer conditions impact future spring conditions using IceBridge data has also been discussed with a researcher from Oregon State University and their colleagues at NOAA. A group at NASA Goddard has also expressed interest in using these data.

The statistical learning techniques described in Chapter 3 can be applied to short term predictive work that utilizes sea ice extent and other variables to estimate sea ice concentration at the end of a given melt season. Future studies could continue to refine work using random forests and neural networks to form predictions for efforts like the Sea Ice Prediction Network's Sea Ice Outlook [144]. Similar studies have shown promising results that may be improved through better neural network design or the incorporation of tree-based methods [21]. A researcher at NSIDC has also expressed interest in using the methods described in this thesis in a future proposal.

The predictor importance values explored in Chapter 4 and utilized in Chapter 5 are potentially useful for a broad set of research efforts. Work involving the comparison of several factors and the determination of remote sensing priorities could use these techniques to rank predictors. These

ranks can help indicate which parameters warrant further study while also assisting researchers with forming groups in datasets. The extension of these techniques to the multi-class case will also greatly expand their potential in future research work such as inter-comparisons of predictor ranks based on ice age.

Continued work should include cumulative values that represent the history of sea ice parcels in the Arctic. This is particularly valuable when discussing values like IST and surface radiation, which can vary greatly from week to week. The variables used throughout this thesis are not independent, but were used to demonstrate the utility of these techniques in studying Arctic data. Future work could benefit from the removal of latitude and further investigation of the processes and conditions that are dependent on it. Future studies would also benefit from exploring generative classifiers that could use statistical relationships that are present in the data. These types of techniques could produce more robust and interpretable predictions that would allow researchers to continue to refine and study these results in future work.

Bibliography

[1] Arctic Sea Ice News and Analysis. Available Online, Accessed 14-Feb-2018. https://nsidc.org/arcticseaicenews/2017/09/.

[2] S. Andersen, R. Tonboe, L. Kaleschke, G. Heygster, and L.T. Pedersen. Intercomparison of Passive Microwave Sea Ice Concentration Retrievals Over the High-Concentration Arctic Sea Ice. J. Geophys. Res., 112, C08004, 2007. doi:10.1029/2006JC003543.

[3] P. M. Atkinson and A. R. L. Tatnall. Introduction to Neural Networks in Remote Sensing. International Journal of Remote Sensing, 18(4):699–7069, 1997. doi:10.1080/014311697218700.

[4] M. Babiker, A. Korosov, and J. W. Park. Sea Ice Type Dynamics in the Arctic Based on Sentinel-1 Data. Proceedings of the 19th EGU General Assembly, page 17155, 2017.

[5] F.I. Badgley. Heat Balance at the Surface of the Arctic Ocean. Western Snow Conference, 1961.

[6] R Barry. The Parameterization of Surface Albedo for Sea Ice and its Snow Cover. Progress in Physical Geography, 20,1:63–79, 1996. doi:https://doi.org/10.1177/030913339602000104.

[7] M. S. Bartlett. Multivariate Analysis. Supplement to the Journal of the Royal Statistical Society, 9(2):176–197, 1947. doi:10.2307/2984113.

[8] L. Belgiu, M. Dragut. Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114:24–31, 2016. doi:10.1016/j.isprsjprs.2016.01.011.

[9] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1995.

[10] A.L. Boulesteix, S. Janitza, J. Kruppa, and I.R. Konig. Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics. WIREs Data Mining Knowl. Discov., 2:493–507, 2012. doi:10.1002/widm.1072.

[11] L. Breiman. Bagging Predictors. Machine Learning, 26(2):123–140, 1996. doi:https://doi.org/10.1007/BF00058655.

[12] L. Breiman. Random Forests. Machine Learning, 45(1):5–32, 2001. doi:https://doi.org/10.1023/A:1010933404324.

[13] L. Breiman, J. Friedman, C. Stone, and R. Olshen. Classification and Regression Trees. CRC Press, 1984.

[14] M.J. Brodzik, B. Billingsley, T. Haran, B. Raup, and M.H. Savoie. EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets. ISPRS Int. J. Geo-Inf., 1(1):32–45, 2012. doi:10.3390/ijgi1010032.

[15] K. M. Brunt, R.L. Hawley, E.R. Lutz, M. Studinger, J.G. Sonntag, M.A. Hofton, L.C. Andrews, and Neumann T.A. Assesment of NASA Airborne Laser Altimetry Data Using Ground-Based GPS Data Near Summit Station, Greenland. The Cryosphere, 11:681–692, 2017. doi:10.5194/tc-11-681-2017.

[16] J.J. Caroll and B.W. Fitch. Effects of Solar Elevation and Cloudiness on Snow Albedo at the South Pole. J. Geophys. Res., 86:5271–5276, 1981. doi:10.1029/JC086iC06p05271.

[17] K. Castro-Morales, R. Ricker, and R. Gerdes. Regional Distribution and Variability of Model-Simulated Arctic Snow on Sea Ice. Polar Science, 13:33–49, 2017. doi:https://doi.org/10.1016/j.polar.2017.05.003.

[18] D. Cavalieri. NASA Sea Ice Validation Program for the DMSP SMM/I: Final Report. NASA Technical Memorandum 104559; National Aeronautics and Space Administration: Washington, DC, USA, 1992.

[19] D. Cavalieri, C.L. Parkinson, P. Gloersen, and H.J. Zwally. Sea Ice Concentrations From Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center, 1996. doi:10.5067/8GQ8LZQVL0VL.

[20] D.J. Cavalieri and H.J. Zwally. Satellite Observations of Sea Ice. Advances in Space Research, 5(6):247–255, 1985. doi:10.1016/0273-1177(85)90328-X.

[21] J. Chi and H. Kim. Prediction of Arctic Sea Ice Concentration Using a Fully Data Driven Deep Neural Network. Remote Sensing, 9(12):1305, 2017. doi:10.3390/rs9121305.

[22] J. Comiso. A Rapidly Declining Perennial Sea Ice Cover in the Arctic. Geophys. Res. Lett., 29(20), 2002. doi:10.1029/2002GL015650.

[23] J.C. Comiso and R. Kwok. Surface and Radiative Characteristics of the Summer Arctic Sea Ice Cover from Multisensor Satellite Observations. J. Geophys. Res., 101:28397–28416, 1996. doi:10.1029/96JC02816.

[24] J.C. Comiso and K. Steffen. Studies of Antarctic Sea Ice Concentrations from Satellite Data and Their Applications. J. Geophys. Res., 106:31361–31385, 2001. doi:10.1029/2001JC000823.

[25] C. Cortes and V. Vapnik. Support-Vector Networks. Machine Learning, 20(3):273–297, 1995.

[26] R.I. Crocker, J. Maslanik, J. Alder, S.E. Palo, U.C. Herzfeld, and W.J. Emery. A Sensor Package for Ice Surface Observations Using Small Unmanned Aircraft Systems. IEEE Trans. Geosci. and Rem. Sens., 50(4), 2012. doi:10.1109/TGRS.2011.2167339.

[27] J.A. Curry and J.L. Schramm. Sea Ice-Albedo Climate Feedback Mechanism. Journal of Climate, 8:240–247, 1995. doi:10.1175/1520-0442(1995)008¡0240:SIACFM¿2.0.CO;2.

[28] D. Cuthbert. One-at-a-Time Plans. Journal of the American Statistical Association, 68(342):353–360, 1972.

[29] M. S. Dawson, A. K. Fung, and M. T. Manry. Sea Ice Classification Using Fast Learning Neural Networks. Proceedings of the 12th Annual International Geoscience and Remote Sensing Symposium, 2 (A93-47551 20-43):1070–1071, 1992.

[30] R. Dominguez. IceBridge DMS L1B Geolocated and Orthorectified Images, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center DAAC, 2010.

[31] Y. Donghui, J.P. Harbeck, S.S. Manizade, N.T. Kurtz, M. Studinger, and M. Hofton. Arctic Sea Ice Freeboard Retrieval with Waveform Characteristics for NASA's Airborne Topographic Mapper (ATM) and Land, Vegetation, and Ice Sensor (LVIS). IEEE Trans. Geoscience and Remote Sensing, 53(3):1403–1410, 2015. doi:10.1109/TGRS.2014.2339737.

[32] C. Drue and G. Heinemann. High-Resolution Maps of the Sea-Ice Concentration from MODIS Satellite Data. Geophys. Res. Lett., 31:L20403, 2004. doi:10.1029/2004GL020808.

[33] C. Ebinger and E. Zambetakis. The Geopolitics of Arctic Melt. International Affairs, Vol. 85, Iss. 6, 1215-1232, 2009. doi:10.1111/j.1468-2346.2009.00858.x.

[34] B. Elder. Ice Mass Balance Data Buoy Page. Available Online, Accessed 25-May-2017. http://imb-crrel-dartmouth.org/imb.crrel/buoysum.htm.

[35] W. Emery, W.S. Good, W. Tandy, M.A. Izaguirre, and P.J. Minnett. A Microbolometer Airborne Calibrated Infrared Radiometer: The Ball Experimental Sea Surface Temperature (BESST) Radiometer. IEEE Trans. on Geoscience and Remote Sensing, Vol. 52, Iss. 12, 2014. doi:10.1109/TGRS.2014.2318683.

[36] W.J. Emery, C. Fowler, and J. Maslanik. Arctic Sea Ice Concentration from Special Sensor Microwave Imager and Advanced Very High Resolution Radiometer Satellite Data. J. Geophys. Res., 99:329–342, 1994. doi:10.1029/94JC01413.

[37] C. Emmerson and G. Lahn. Lloyd's Arctic Opening: Opportunity and Risk in the High North. Available Online, Accessed 8-Jan-2018. https://www.lloyds.com/ /media/files/news

[38] S. Farrell, N. Kurtz, L.N. Connor, B.C. Elder, C. Leuschen, T. Markus, D.C. McAdoo, B. Panzer, J. Richter-Menge, and J.G. Sonntag. A First Assessment of IceBridge Snow and Ice Thickness Data over Arctic Sea Ice. IEEE Trans. on Geoscience and Remote Sensing, 50(6):2098-2111, 2012. doi:10.1109/TGRS.2011.2170843.

[39] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. Annals of Human Genetics, 7(2), 1936. doi:10.1111/j.1469-1809.1936.tb02137.x.

[40] R. A. Fisher. The Precision of Discriminant Functions. Annals of Human Genetics, 10(1):422–429, 1940. doi:10.1111/j.1469-1809.1940.tb02264.x.

[41] D. Flocco, D. Schroeder, D.L. Feltham, and E.C. Hunke. Impact of Melt Ponds on Arctic Sea Ice Simulations from 1990 to 2007. J. Geophys. Res. Oceans, 117(C9):C09032, 2012. doi:10.1029/2012JC008195.

[42] C. Fowler, W.J. Emery, and J. Maslanik. Satellite-Derived Evolution of Arctic Sea Ice Age: October 1978 to March 2003. Geoscience and Remote Sensing Letters, IEEE, 1(2); 72-74, 2004. doi:10.1109/LGRS.2004.824741.

[43] K.E. Frey, G.W.K. Moore, L.W. Cooper, and J.M. Grebmeier. Divergent Patterns of Recent Sea Ice Cover Across the Bering, Chukchi, and Beaufort Seas of the Pacific Arctic Region. Progress in Oceanography, 136:32–49, 2015. doi:10.1016/j.pocean.2015.05.009.

[44] R.J. Galley, D. Babb, M. Ogi, B.G.T. Else, N.X. Geilfus, O. Crabeck, D.G. Barber, and S. Rysgaard. Replacement of Multiyear Sea Ice and Changes in the Open Water Season Duration in the Beaufort Sea Since 2004. J. Geophys. Res. Oceans, 121(3):1806–1823, 2016. doi:10.1002/2015JC011583.

[45] M. W. Gardnera and S. R. Dorlinga. Artificial Neural Networks (the multilayer perceptron) - A Review of Applications in the Atmospheric Sciences. Atmospheric Environment, 32(14-15):2627–2636, 1998. doi:10.1016/S1352-2310(97)00447-0.

[46] P. Gloerson and W.J. Campbell. Recent Variations in Arctic and Antarctic Sea-Ice Covers. Nature, 352:33–36, 1991. doi:10.1038/352033a0.

[47] M. T. Hagan and M. B. Menhaj. Training Feedforward Networks With the Marquardt Algorithm. IEEE Trans. on Neural Networks, 5(6):989–993, 1994.

[48] D. K. Hall, S.V. Nghiem, Rigor I.G., and J.A. Miller. Uncertainties of Temperature Measurements on Snow-Covered Land and Sea Ice From In Situ and MODIS Data During BROMEX. J. Appl. Meteor. Climatol., 54, 966-978, 2015. doi:10.1175/JAMC-D-14-0175.1.

[49] D.K. Hall, J.R. Key, K.A. Casey, G.A. Riggs, and D.J. Cavalieri. Sea ice surface temperature product from MODIS. IEEE Trans. Geosci. and Rem. Sens., 42:1076–1087, 2004. doi:10.1109/TGRS.2004.825587.

[50] D.K. Hall and G.A. Riggs. MODIS/Terra Sea Ice Extent and IST Daily L3 Global 4km EASE-Grid Day, Version 6. Boulder, Colorado USA:NASA DAAC at the National Snow and Ice Data Center, 2015. doi:10.5067/MODIS/MOD29E1D.006.

[51] H. Han, J. Im, M. Kim, S. Sim, J. Kim, D.J. Kim, and S.H. Kang. Retreival of Melt Ponds on Arctic Multiyear Sea Ice in Summer from TerraSAR-X Dual-Polarization Data Using Machine Learning Approaches: A Case Study in the Chukchi Sea with Mid-Incidence Angle Data. Remote Sensing, 8(1):57, 2016. doi:10.3390/rs8010057.

[52] M. Hanselmann, U. Kothe, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R. Heeren, and F.A. Hamprecht. Towards Digital Staining Using Imaging Mass Spectrometry and Random Forests. Technical Report, 2009. doi:10.1021/pr900253y.

[53] T. Hastie and R. Tibshirani. Generalized Additive Models: Some Applications. Journal of the American Statistical Association, 82(398):371–386, 1987.

[54] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning, 2nd Ed. Springer, 2017.

[55] Tin Kam Ho. Random Decision Forests. Proceedings of the Third International Conference on Document Analysis and Recognition. Montreal, Quebec. IEEE, 1:278–282, 1995. doi:https://doi.org/10.1109

[56] R. Hock, J.K. Hutchings, and M. Lehning. Grand Challenges in Cryospheric Sciences: Toward Better Predictability of Glaciers, Snow, and Sea Ice. Front. Earth Sci., 5:64, 2017. doi:10.3389/feart.2017.00064.

[57] S. Howell, M. Brady, C. Derksen, and R. Kelly. Recent Changes in Sea Ice Area Flux Through the Beaufort Sea During the Summer. J. Geophys. Res. Oceans, 121:2569–2672, 2016. doi:10.1002/2015JC011464.

[58] N. Ivanova, L.T. Pedersen, R.T. Tonboe, S. Kern, G. Heygster, T. Lavergne, A. Sorensen, R. Saldo, G. Dybkjoer, L. Brucker, and M. Shokr. Inter-comparison and Evaluation of Sea Ice Algorithms: Towards Further Identification of Challenges and Optimal Approach Using Passive Microwave Observations. The Cryosphere, 9, 1797-1817, 2015. doi:10.5194/tc-9-1797-2015.

[59] J.D. Jacobs, R.G. Barry, and R.L. Weaver. Fast Ice Characteristics with Special Reference to the Eastern Canadian Arctic. Polar Record, 17:521–536, 1975. doi:10.1017/S0032247400032484.

[60] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1):4–37, 2000. doi:10.1109/34.824819.

[61] G. James, D. Witten, T. Hastie, and R. Tibshirani. An Introduction to Statistical Learning with Applications to R. Springer, 2015.

[62] O.M. Johannessen, E.V. Shalina, and M.W. Miles. Satellite Evidence for an Arctic Sea Ice Cover in Transformation. Science, 286(5446):1937–1939, 1999. doi:10.1126/science.286.5446.1937.

[63] P. Kanagaratnam, T. Markus, V. Lytle, B. Heavey, P. Jansen, G. Prescott, and S.P. Gogineni. Ultrawideband Radar Measurements of Thickness of Snow Over Sea Ice. IEEE Trans. on Geoscience and Remote Sensing, 45(9):2715–2724, 2007. doi:10.1109/TGRS.2007.900673.

[64] M. Kapsch, R. Graversen, and M. Tjernstrom. Springtime Atmospheric Energy Transport and the Control of Arctic Summer Sea-Ice Extent. Nature Climate Change, 3:744–748, 2013. doi:10.1038/nclimate1884.

[65] M. Kapsch, R.G. Graversen, M. Tjernstrom, and R. Brintanja. The Effect of Downwelling Longwave and Shortwave Radiation on Arctic Summer Sea Ice. Journal of Climate, 29:1143–1159, 2016. doi:10.1175/JCLI-D-15-0238.1.

[66] M.L. Kapsch, R.G. Graversen, T. Economou, and M. Tjernstrom. The Importance of Spring Atmospheric Conditions for Predictions of the Arctic Summer Sea Ice Extent. Geophys. Res. Lett., 41(14):5288–5296, 2014. doi:10.1002/2014GL060826.

[67] V.M. Kattsov, V.E. Ryabinin, J.E. Overland, M.C. Serreze, M. Visbeck, J.E. Walsh, W. Meier, and X. Zhang. Arctic Sea Ice Change: A Grand Challenge of Climate Science. J. of Glaciology, 56(200):1115–1121, 2010. doi:10.3189/002214311796406176.

[68] J.E. Kay, T. L'Ecuyer, A. Gettelman, G. Stephens, and C. O'Dell. The Contribution of Cloud and Radiation Anomalies to the 2007 Arctic Sea Ice Extent Minimum. Geophys. Res. Lett., 35(8):L08503, 2008. doi:10.1029/2008GL033451.

[69] J. Key. The Cloud and Surface Parameter Retrieval (CASPR) System for Polar AVHRR Data User's Guide. Space Science and Engineering Center, Univ. of WI, Madison, 2001.

[70] J. Key. The AVHRR Polar Pathfinder Climate Data Records. Remote Sensing, 8:167, 2016. doi:10.3390/rs8030167.

[71] J. Key and M. Haefliger. Arctic Ice Surface Temperature Retrieval from AVHRR Thermal Channels. J. Geophys. Res. Atmospheres, 97(D5):5885–5893, 1992. doi:10.1029/92JD00348.

[72] J. Key and A. Schweiger. Tools for Atmospheric Radiative Transfer: Streamer and FluxNet. Computers and Geosciences, 24(5), 443-451, 1998. doi:10.1016/S0098-3004(97)00130-1.

[73] J. Key and A. Schweiger. Tools for Atmospheric Radiative Transfer: Streamer and FluxNet. Computational Geoscience, 24:443–451, 1998. doi:10.1016/S0098-3004(97)00130-1.

[74] J. Key and X. Wang. Extended AVHRR Polar Pathfinder (APP-X) Climate Algorithm Theoretical Basis Document. NOAA National Climatic Data Center, 2015. CDRP-ATBD-0573.

[75] J. Key, X. Wang, and Y. Liu. NOAA Climate Data Record of AVHRR Polar Pathfinder Extended (APP-X), Version 1, Revision 1. NOAA National Climatic Data Center, 2014. doi:10.7289/V5MK69W6.

[76] J. Key, X. Wang, J.C. Stroeve, and C. Fowler. Estimating the Cloudy-Sky Albedo of Sea Ice and Snow from Space. Journal of Geophysical Research Atmos., 106:12489–12497, 2001. doi:10.1029/2001JD900069.

[77] R.M. Koerner. Weather and Ice Observations of the British Trans-Arctic Expedition 1968-1969. Weather, 25:218–228, 1970. doi:10.1002/j.1477-8696.1970.tb03266.x.

[78] W. B. Krabill and E. Buzay. IceBridge KT19 IR Surface Temperature, Version 1. Boulder, Colorado USA. NASA National Snow and Ice Data Center DAAC, 2012. doi:http://dx.doi.org/10.5067/I883KXU7ZO8O.

[79] R. Krishfield, A. Proshutinsky, K. Tateyama, W.J. Williams, E.C. Carmack, F.A. McLaughlin, and M.L. Timmermans. Deterioration of Perennial Sea Ice in the Beaufort Gyre from 2003 to 2012 and its Impact on the Oceanic Freshwater Cycle. J. Geophys. Res. Oceans, 119, 1271-1305, 2014. doi:10.1002/2013JC008999.

[80] N. Kurtz, S.L. Farrell, M. Studinger, N. Gallin, J.P. Harbeck, R. Lindsay, V.D. Onana, B. Panzer, and J.G. Sonntag. Sea Ice Thickness, Freeboard, and Snow Depth Products from Operation IceBridge Airborne Data. The Cryosphere 7:1035-1056, 2013. doi:10.5194/tc-7-1035-2013.

[81] N. Kurtz and J. Harbeck. Operation IceBridge Sea Ice Freeboard, Snow Depth, and Thickness Data Products Manual, Version 2 Processing. Available Online, Accessed 10-Jan-2018. https://nsidc.org/sites/nsidc.org/files/files/data/icebridge/IDCSI4-icebridge-products-manual-Version2-June-2015.pdf.

[82] N. Kurtz, M. Studinger, J. Harbeck, V. Onana, and D. Yi. IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center, 2015. doi:10.5067/G519SHCKWQV6.

[83] R. Kwok and G.F. Cunningham. Contribution of Melt in the Beaufort Sea to the Decline in Arctic Multiyear Sea Ice Coverage: 1993-2009. Geophys. Res. lett., 37(20):L20501, 2010. doi:10.1029/2010GL044678.

[84] R. Kwok and G.F. Cunningham. Variability of Arctic Sea Ice Thickness and Volume from CryoSat-2. Phil. Trans. R. Soc. A, 373(2045), 2015. doi:10.1098/rsta.2014.0157.

[85] R. Kwok, G.F. Cunningham, S.S. Manizade, and W.B. Krabill. Arctic Sea Ice Freeboard from IceBridge Acquisitions in 2009: Estimates and Comparisons With ICESat. Journal of Geophysical Research, 117:CO2018, 2012. doi:10.1029/2011JC007654.

[86] R. Kwok, C. Leuschen, B. Panzer, A. Patel, N. Kurtz, T. Markus, B. Holt, and P. Gogineni. Radar Surveys of Snow Depth Over Arctic Sea Ice During Operation IceBridge. AGU Fall Meeting Abstracts, C21D-02, 2010.

[87] R. Kwok and T. Markus. Potential Basin-Scale Estimates of Arctic Snow Depth With Sea Ice Freeboards from CryoSat-2 and ICESat-2: An Exploratory Analysis. Adv. Space Res., 2017. doi:https://doi.org/10.1016/j.asr.2017.09.007.

[88] R. Kwok, B. Panzer, C. Leuschen, S. Pang, T. Markus, B. Holt, and S. Gogineni. Airborne Surveys of Snow Depth Over Arctic Sea Ice. Journal of Geophysical Research Oceans, 116:C11018, 2011. doi:10.1029/2011JC007371.

[89] R. Kwok and N. Untersteiner. The Thinning of Arctic Sea Ice. Physics Today, 64:36–41, 2011. doi:10.1063/1.3580491.

[90] S. Lee, J. Im, J. Kim, M. Kim, M. Shin, H.C. Kim, and L.J. Quackenbush. Arctic Sea Ice Thickness Estimation from CryoSat-2 Satellite Data Using Machine Learning-Based Lead Detection. Remote Sensing, 8(9):698, 2016. doi:10.3390/rs8090698.

[91] R. Lindsay and A. Schweiger. Arctic Sea Ice Thickness Loss Determined Using Subsurface, Aircraft, and Satellite Measurements. The Cryosphere, 9:269–283, 2015. doi:10.5194/tc-9-269-2015.

[92] W. Y. Loh. Regression Trees with Unbiased Variable Selection and Interaction Detection. Statistica sinica, 12:361–386, 2002.

[93] W. Y. Loh and Y. S. Shih. Split selection methods for classification trees. Statistica sinica, 7:815–840, 1997.

[94] A. Lubin. Linear and Non-Linear Discriminating Functions. British Journal of Mathematical and Statistical Psychology, 3(2):90–104, 1950. doi:10.1111/j.2044-8317.1950.tb00287.x.

[95] T. Markus, D.J. Cavalieri, M. Tschudi, and A. Ivanoff. Comparison of Aerial Video and Landsat 7 Data Over Ponded Sea Ice. Rem. Sens. of Env., 86:458–469, 2003. doi:10.1016/S0034-4257(03)00124-X.

[96] T. Martin, M. Steele, and J. Zhang. Seasonality and Long-Term Trend of Arctic Ocean Surface Stress in a Model. J. Geophys. Res. Oceans, 119(3):1723–1738, 2014. doi:10.1002/2013JC009425.

[97] J. Maslanik. Investigations of Spatial and Temporal Variability of Ocean and Ice Conditions in and Near the Marginal Ice Zone: The "Marginal Ice Zone Observations and Processes EXperiment" (MIZOPEX). NASA Grant NNX11AN57G Annual Report, Year 2, 2013.

[98] J. Maslanik, C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W.J. Emery. A Younger, Thinner Arctic Ice Cover: Increased Potential for Rapid, Extensive Sea-Ice Loss. The Cryosphere, 34(24):L24501, 2007. doi:10.1029/2007GL032043.

[99] J. Maslanik, J. Stroeve, C. Fowler, and W.J. Emery. Distribution and Trends in Arctic Sea ice Age Through Spring 2011. Geophys. Res. Lett., 38(L13502), 2011. doi:10.1029/2011GL047735.

[100] J.A. Maslanik, M.C. Serreze, and R.G. Barry. Recent Decreases in Arctic Summer Ice Cover and Linkages to Atmospheric Circulation Anomalies. Geophys. Res. Lett., 23(13):1677–1680, 1996. doi:10.1029/96GL01426.

[101] P. Mather and B. Tso. Classification Methods for Remotely Sensed Data. CRC press, 2016.

[102] W.N. Meier, D. Gallaher, and G.G. Campbell. New Estimates of Arctic and Antarctic Sea ice Extent During September 1964 From Recovered Nimbus 1 Satellite Imagery. The Cryosphere, 7:699–705, 2013. doi:10.5194/tc-7-699-2013.

[103] W.N. Meier, G.K. Hovelsrud, B.E.H. Oort, J.R. Key, K.M. Kovacs, C. Michel, C. Haas, M.A. Granskog, S. Gerland, D.K. Perovich, A. Makshtas, and J.D. Reist. Arctic Sea Ice in Transformation: A Review of Recent Observed Changes and Impacts on Biology and Human Activity. Rev. of Geophys., 52(3):185–217, 2014. doi:10.1002/2013RG000431.

[104] N. Melia, K. Haines, and E. Hawkins. Sea Ice Decline and 21st Century Trans-Arctic Shipping Routes. Geophys. Res. Lett., 43:9720–9728, 2016. doi:10.1002/2016GL069315.

[105] I. Merkouriadi, B. Cheng, R.M. Graham, A. Rosel, and M.A. Granskog. Critical Role of Snow on Sea Ice Growth in the Atlantic Sector of the Arctic Ocean. Geophys. Res. Lett., In Press., 2017. doi:10.1002/2017GL075494.

[106] X. Miao, H. Xie, S.F. Ackley, D.K. Perovich, and C. Ke. Object-Based Detection of Arctic Sea Ice and Melt Ponds Using High Spatial Resolution Aerial Photographs. Cold Regions Science and Technology, 119:211–222, 2015. doi:10.1016/j.coldregions.2015.06.014.

[107] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers. Fisher discriminant analysis with kernels. Neural Networks for Signal Processing IX. Proceedings of the 1999 IEEE Signal Processing Society Workshop., pages 41–48, 1999.

[108] J. Morrison. Polar Science Center Seasonal Ice Zone Reconnaisance Surveys (SIZRS) Website. Available Online, Accessed 21-August-2017. http://psc.apl.uw.edu/research/projects/sizrs/.

[109] J. Mortin, G. Svensseon, R.G. Graversen, M.L. Kapsch, J. Stroeve, and L.N. Boisvert. Melt Onset Over Arctic Sea Ice Controlled by Atmospheric Moisture Transport. Geophysical Research Letters, 43:6636–6642, 2016. doi:10.1002/2016GL069330.

[110] G. Mountrakis, Im. J., and C. Ogole. Support Vector Machines in Remote Sensing: A Review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3):247–259, 2011. doi:10.1016/j.isprsjprs.2010.11.001.

[111] A. Mozaffari, K.A. Scott, S. Chenouri, and N.L. Azad. A Modular Ridge Randomized Neural Network With Differential Evolutionary Distributor Applied to the Estimation of Sea Ice Thickness. Soft Computing, 21(16):4635–4659, 2017. doi:10.1007/s00500-016-2074-5.

[112] Martin F. Mller. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks, 6(4):525–533, 1993.

[113] J. A. Nelder and R. J. Baker. Generalized Linear Models. John Wiley and Sons, 1972.

[114] S.V. Nghiem, I.G. Rigor, D.K. Perovich, P. Clemente-Colon, J.W. Weatherly, and G. Neumann. Rapid Reduction of Arctic Perennial Sea Ice. Geophys. Res. Lett., 34(19):L19504, 2007. doi:10.1029/2007GL031138.

[115] D. Notz. How Well Must Climate Models Agree with Observations? Phil. Trans. R. Soc. A, 373(2052), 2015. doi:10.1098/rsta.2014.0164.

[116] M. Ortiz, J.C. Pinales, H.C. Graber, J. Wilkinson, and B. Lund. Quantifying the Evolution of Melt Ponds in the Marginal Ice Zone Using High Resolution Optical Imagery and Neural Netowrks. American Geophysical Union Ocean Sciences Meeting 2016, pages HE21A–03, 2017.

[117] J.E. Overland and M. Wang. Large Scale Atmospheric Circulation Changes are Associated With the Recent Loss of Arctic Sea Ice. Tellus A, 62(1):1–9, 2010. doi:10.1111/j.1600-0870.2009.00421.x.

[118] M. Pal. Random Forest Classifier for Remote Sensing Classification. International Journal of Remote Sensing, 26(1):217–222, 2005. doi:https://doi.org/10.1080/01431160412331269698.

[119] C.L. Parkinson, D.J. Cavalieri, P. Gloerson, H.J. Zwally, and J.C. Comiso. Arctic Sea Ice Extents, Areas, and Trends 1978-1996. J. of Geophys. Res. Oceans, 104(C9):20837–20856, 1999. doi:10.1029/1999JC900082.

[120] C.L. Parkinson and N. E. DiGirolamo. New Visualizations Highlight New Information on the Contrasting Arctic and Antarctic Sea-Ice Trends Since the late 1970s. Remote Sensing of the Environment, 183(15):198–204, 2016. doi:10.1016/j.rse.2016.05.020.

[121] D. Perovich, B. Light, H. Eicken, K.F. Jones, K. Runciman, and S.V. Nghiem. Increasing Solar Heating of the Arctic Ocean and Adjacent Seas, 1979-2005: Attribution and Role in the Ice-Albedo Feedback. Geophys. Res. Lett., 34:L19505, 2007. doi:10.1029/2007GL031480.

[122] D. Perovich and C. Polashenski. Albedo Evolution of Seasonal Arctic Sea Ice. Geophys. Res. Lett., 39:L08501, 2012. doi:10.1029/2012GL051432.

[123] D. Perovich, C. Polashenski, A. Arntsen, and C. Stwertka. Anatomy of a Late Spring Snowfall on Sea Ice. Geophys. Res. Lett., 44:2802–2809, 2017. doi:10.1002/2016GL071470.

[124] D. Perovich and J. Richter-Menge. Regional Variability in Sea Ice Melt in a Changing Arctic. Pilos. Trans. R. Soc. A Math. Phys. Eng. Sci., 373, 2015. doi:10.1098/rsta.2014.0165.

[125] D. Perovich, J.A. Richter-Menge, B. Elder, K. Claffey, and C. Polashenski. Observing and Understanding Climate Change: Monitoring the Mass Balance, Motion, and Thickness of Arctic Sea Ice. Available Online, Accessed Jun 2017. http://imb-crrel-dartmouth.org.

[126] D.K. Perovich, K.F. Jones, B. Light, and H. Eicken. Solar Partitioning in a Changing Arctic Sea-Ice Cover. Annals of Glaciology, 52(57):192–196, 2011. doi:10.3189/172756411795931543.

[127] A. Proshutinsky, D. Dukhovskoy, M.L. Timmermans, R. Krishfield, and J.L. Bamber. Arctic Circulation Regimes. Phil. Trans. R. Soc. A, 373:20140160, 2015. doi:10.1098/rsta.2014.0160.

[128] W.G. Rees and S.P. James. Angular Variation of the Infrared Emissivity of Ice and Water Surfaces. International Journal of Remote Sensing, 13:2873–2886, 1992. doi:10.1080/01431169208904088.

[129] J. Richter-Menge, J.E. Overland, J.T. Mathis, and E. Osborne. Arctic Report Card 2017. Available Online: `http://www.arctic.noaa.gov/Report-Card`. Accessed 8-Jan-2018.

[130] I. Rigor. IABP Drifting Buoy, Pressure, Temperature, Position, and Interpolated Ice Velocity. Compiled by the Polar Science Center, APL, Univ. of Washington, Seattle in Assoc. with NSIDC, Boulder CO., 2002. doi:10.7265/N53X84K7.

[131] I. Rigor, R. Colony, and S. Martin. Variations in Surface Air Temperature Observations in the Arctic, 1979-1997. Journal of Climate, 13:896–914, 2000. doi:10.1175/1520-0442(2000)013¡0896:VISATO¿2.0.CO;2.

[132] I. Rigor and J. Wallace. Variations in the Age of Arctic Sea Ice and Summer Sea Ice Extent. Geophysical Research Letters, 31(L09401), 2004. doi:10.1029/2004GL019492.

[133] F. Rodriguez-Morales, S. Gogineni, C.J. leuschen, J.D. Paden, J. Li, C.S. Lewis, B. Panzer, D. Alvestegui, A. Patel, K. Byers, R. Crowe, K. Player, R.D. Hale, E.J. Arnold, L. Smith, C.M. Gifford, D. Braaten, and C. Panton. Advanced Multifrequency Radar Instrumentation for Polar Research. IEEE Trans. on Geoscience and Remote Sensing, 52(5):2824–2842, 2014. doi:10.1109/TGRS.2013.2266415.

[134] A. Rosel and L. Kaleschke. Exceptional Melt Pond Occurence in the Years 2007 and 2011 on the Arctic Sea Ice Revealed from MODIS Satellite Data. J. Geophys. Res. Oceans, 117(C5):C05018, 2012. doi:10.1029/2011JC007869.

[135] T. Scambos, T. Haran, and R. Massom. Validation of AVHRR and MODIS Ice Surface Temperature Products Using In Situ Radiometers. Annals of Glaciology, 44, 345-351, 2006. doi:10.3189/172756406781811457.

[136] A. Schweiger, R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok. Uncertainty in Modeled Arctic Sea Ice Volume. J. Geophys. Res., 2011. doi:10.1029/2011JC007084.

[137] M. Serreze and R. Barry. The Arctic Climate System. Cambridge University Press, 2014.

[138] X. Shen, J. Zhang, X. Zhang, J. Meng, and C. Ke. Sea Ice Classification Using Cryosat-2 Altimeter Data by Optimal Classifier-Feature Assembly. IEEE Geoscience and Remote Sensing Letters, 14(11):1948–1952, 2017. doi:10.1109/LGRS.2017.2743339.

[139] C. A. Shuman, D.K. Hall, N.E. DiGirolamo, T.K. Mefford, and M.J. Schnaubelt. Comparison of Near-Surface Air Temperatures and MODIS Ice Surface Temperatures at Summit, Greenland (2008-2013). J. of App. Meteorology and Climatology, 53(9), 2171-2180, 2014. doi:10.1175/JAMC-D-14-0023.1.

[140] H. W. Sorenson. Least-Squares Estimation: from Gauss to Kalman. IEEE Spectrum, 7(7):63–68, 1970. doi:10.1109/MSPEC.1970.5213471.

[141] M. Steele, S. Dickinson, J. Zhang, and R.W. Lindsay. Seasonal Ice Loss in the Beaufort Sea: Toward Synchrony and Prediction. J. Geophys. Res. Oceans, 120:1118–1132, 2015. doi:10.1002/2014JC010247.

[142] S. M. Stigler. Gauss and the Invention of Least Squares. The Annals of Statistics, 9(3):465–474, 1981.

[143] G. Stossmeister. SHEBA Data Archive. Available Online, Accessed 22-May-2017. https://www.eol.ucar.edu/projects/sheba/.

[144] J. Stroeve, H. Goessling, and F. Massonnet. Sea Ice Prediction Network 2017 Post-Season Report. Available Online, Accessed 12-Feb-2018. https://www.arcus.org/sipn/sea-ice-outlook/2017/post-season.

[145] J. Stroeve, M.C. Serreze, M.M. Holland, J.E. Kay, J. Maslanik, and A.P. Barrett. The Arctic's Rapidly Shrinking Sea Ice Cover: A Research Synthesis. Climactic Change, Vol. 110, Iss. 3-4, pp1005-1027, 2012. doi:10.1007/s10584-011-0101-1.

[146] J. C. Stroeve, T. Markus, L. Boisvert, J. Miller, and A. Barrett. Changes in Arctic Melt Season and Implications for Sea Ice Loss. Geophys. Res. Lett., 41:1216–1225, 2014. doi:10.1002/2013GL05895.

[147] J. C. Stroeve, J. Maslanik, M.C. Serreze, I. Rigor, W. Meier, and C. Fowler. Sea Ice Response to an Extreme Negative Phase of the Arctic Oscillation During Winter 2009/2010. Geophys. Res. Lett., 38:L02502, 2011. doi:10.1029/2010GL045662.

[148] M. Sturm, D.K. Perovich, and M.C. Serreze. Meltdown in the North. Scientific American, 289:60–67, 2003. doi:10.1038/scientificamerican1003-60.

[149] J. Suykens and J. Vandewalle. Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 9(3):293–300, 1999. doi:https://doi.org/10.1023/A:1018628609742.

[150] S. Szanyi, J.V. Lukovich, D.G. Barber, and G. Haller. Persistent Artifacts in the NSIDC Ice Motion Data Set and Their Implications for Analysis. J. Geophys. Res. Lett., 43, 10800-10807, 2016. doi:10.1002/2016GL069799.

[151] D.N. Thomas. Sea Ice, 3rd Ed. John Wiley and Sons, 2017.

[152] R.A. Thomas, C. Deser, and Sun L. The Role of Ocean Heat Transport in the Global Climate Response to Projected Arctic Sea Ice Loss. J. of Climate, 29:6841–6859, 2016. doi:10.1175/JCLI-D-15-0651.1.

[153] A. S. Thorndike and R. Colony. Sea Ice Motion in Response to Geostrophic Winds. J. Geophys. Res., (87)C8, 5845-5852, 2010. doi:10.1029/JC087iC08p05845.

[154] M. Tooth. Airborne Remote Sensing of Sea Surface Temperature Using the Ball Experimental Sea Surface Temperature (BESST) Radiometer With a Discussion of the 2013 Marginal Ice Zone Observation and Processes EXperiment. AGU Fall Meeting Posters, 2014.

[155] M. Tooth and M. Tschudi. Quantifying Sea ice Advection Through Key "Gates" in the Arctic Using the Sea Ice Motion and Age Data Products at the University of Colorado Boulder With Applications to Studying Changes in the Arctic Ice Pack. AGU Fall Meeting Posters, 2015.

[156] M. Tooth and M. Tschudi. EASE-Grid Sea Ice Parcel Tracks With Ancillary Satellite Data Products from 2001 to 2016. Pangaea Data Publisher for Earth and Data Science, 2016. doi:10.1594/PANGAEA.871925.

[157] M. Tooth and M. Tschudi. Tracking Sea Ice Parcel Histories: Analyzing the Processes Leading to the Decline of Arctic Sea ice From 1996-2015 by Linking Sea Ice Motion With Ancillary Data Products. AGU Fall Meeting Presentation, SC21E, 2016.

[158] M. Tooth and M. Tschudi. A Database of Weekly Sea Ice Parcel Tracks Derived from Lagrangian Motion Data with Ancillary Data Products. Data, 2, 25, 2017. doi:10.3390/data2030025.

[159] M. Tooth and M. Tschudi. Investigating Arctic Sea Ice Survivability in the Beaufort Sea. Remote Sensing, 10(2):267, 2018. doi:10.3390/rs10020267.

[160] M. Tooth, M. Tschudi, and T. Matsuo. Using Tree-Based Methods to Investigate Arctic Sea Ice Survival. In Preparation for Submission, 2018.

[161] M. Tsamados, D. Feltham, A. Petty, D. Schroeder, and D. Flocco. Processess Controlling Surface, Bottom, and Lateral Melt of Arctic Sea Ice in a State of the Art Sea Ice Model. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 373(2052):247–259, 2015. doi:10.1098/rsta.2014.0167.

[162] M. Tschudi. Acquisition of Ice Thickness and Ice Surface Characteristics in the Seasonal Ice Zone by CULPIS-X During the US Coast Guard's Arctic Domain Awareness Program. ONR Grant N00014-12-1-0233 Annual Report, 2015.

[163] M. Tschudi, C. Fowler, and J. Maslanik. Tracking the Movement and Changing Surface Characteristics of Arctic Sea Ice. IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens., Vol 3 Issue 4, 536-540, 2010. doi:10.1109/JSTARS.2010.2048305.

[164] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. EASE-Grid Sea Ice Age. Boulder, Colorado USA: National Snow and Ice Data Center, 2016. doi:10.5067/PFSVFZA9Y85G.

[165] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. Polar Pathfinder Daily 25km EASE-Grid Sea Ice Motion Vectors. Version 3. Boulder, Colorado USA: National Snow and Ice Data Center, 2016. doi:10.5067/O57VAIT2AYYY.

[166] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. Ice Motion from AVHRR. Available Online, Accessed Jul 2017. `https://nsidc.org/data/docs/daac/nsidc0116_icemotion/avhrr.html#accuracy`.

[167] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. Ice Motion from IABP Buoys. Available Online, Accessed Jul 2017. https://nsidc.org/data/docs/daac/nsidc0116_icemotion/buoy.html#accuracy.

[168] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. Ice Motion from NCEP/NCAR Winds. Available Online, Accessed Jul 2017. https://nsidc.org/data/docs/daac/nsidc0116_icemotion/winds.html#accuracy.

[169] M. Tschudi, C. Fowler, J. Maslanik, S. Stewart, and W. Meier. Ice Motion from Passive Microwave: SMMR, SSM/I, SSMIS, and AMSR-E. Available Online, Accessed Jul 2017. https://nsidc.org/data/docs/daac/nsidc0116_icemotion/smmr_ssmi.html#accuracy.

[170] M. Tschudi, J. Stroeve, and S. Stewart. Relating the Age of Arctic Sea Ice to its Thickness, as Measured during NASA's ICESat and IceBridge Campaigns. Remote Sens., 8(6):457, 2016. doi:10.3390/rs8060457.

[171] M. Tschudi, M. Tooth, and R. Barton-Grimley. Acquisition of Airborne Sea Ice Remote Sensing Data with CULPIS-X: an Instrument Mounted on a Coast Guard C-130. AGU Fall Meeting Presentation, SC43D, 2016.

[172] M. Tschudi, M. Tooth, W. Meier, and S. Stewart. There Goes the Sea Ice: Following Arctic Sea Ice Parcels and Their Properties. AGU Fall Meeting Posters, C21G-1186, 2017.

[173] M.A. Tschudi, J.A. Maslanik, and D.K. Perovich. Derivation of Melt Pond Coverage on Arctic Sea Ice Using MODIS Observations. Rem. Sens. Env., 112:2605–2614, 2008. doi:10.1016/j.rse.2007.12.009.

[174] N. Untersteiner. On the Mass and Heat Budget of Arctic Sea Ice. Meteorology and Atmospheric Physics, 12(2):151–182, 1961. doi:10.1007/BF02247491.

[175] N. Untersteiner. Remarks on the Cooling Power in Polar Regions. US-IGY Drifting Station Alpha Arctic Ocean 1957-1958, 13:81, 1965.

[176] P. Wadhams. Ice in the Ocean. Gordon and Breach Science Publishers, 1972.

[177] X. Wang, J. Key, R. Kwok, and J. Zhang. Comparison of Arctic Sea Ice Thickness from Satellites, Aircraft, and PIOMAS Data. Remote Sens., 8(9):713, 2016. doi:10.3390/rs8090713.

[178] S.G. Warren. Optical Properties of Snow. Rev. Geophys. Space Phys., 2:67–89, 1982. doi:10.1029/RG020i001p00067.

[179] W. Weeks and W. D. Hibler. On Sea Ice. University of Alaska Press, 2010.

[180] G. Whiteman, C. Hope, and P. Wadhams. Climate Science: Vast Costs of Arctic Change. Nature, 499:401–403, 2013. doi:10.1038/499401a.

[181] D. Yi, J.P. Harbeck, S.S. Manizade, N.T. Kurtz, M. Studinger, and M. Hofton. Arctic Sea Ice Freeboard Retreival With Waveform Characteristics for NASA's Airborne Topographic Mapper (ATM) and Land, Vegetation, and Ice Sensor (LVIS). IEEE Trans. on Geoscience and Remote Sensing, 53(3):1403-1410, 2015. doi:10.1109/TGRS.2014.2339737.

[182] J. L. Zhang and D. A. Rothrock. Modeling Global Sea Ice With a Thickness and Enthalpy Distribution Model in Generalized Curvilinear Coordinates. Mon. Weather Rev., 131, 845-861, 2003. doi:10.1175/1520-0493(2003)131¡0845:MGSIWA¿2.0.CO;2.

Appendix A

Author's Publications and Presentations

A.1 Publications

The work described in this manuscript has been the subject of the following peer-reviewed publications led by the author:

(1) Tooth, M. and Tschudi, M. A Database of Weekly Sea Ice Parcel Tracks Derived from Lagrangian Motion Data with Ancillary Data Products. *Data*, 2, 25, 2017. [158]

(2) Tooth, M. and Tschudi, M. Investigating Arctic Sea Ice Survival in the Beaufort Sea. *Remote Sensing*, 10(2):267, 2018. [159]

(3) Tooth, M., Tschudi, M., and Matsuo, T. Using Tree-Based Methods to Investigate Arctic Sea Ice Survival. *In Preparation for Submission*, 2018. [160]

A.2 Presentations

The work described in this manuscript has been presented by the author as part of the following talks, listed in descending chronological order:

(1) Tooth, M. The Application of Machine Learning Techniques to Studying Arctic Sea Ice Survival. *CU ATOC Forum*, 2018.

(2) Tooth, M. The Application of Statistical Learning Techniques to Studying Arctic Sea Ice Survival. *NSIDC CPP Invited Talk*, 2018.

(3) Tooth, M. Bridging the Gap Between Satellite and Airborne Measurements of Sea Ice. *CU Aerospace Engineering Department Researchpalooza*, 2017.

(4) Tooth, M. Studying Sea Ice: Using LIDAR and Buoys for Arctic Science. *CU ATOC Forum*, 2017.

(5) Tooth, M. and Tschudi, M. Tracking Sea Ice Parcel Histories: Analyzing the Processes Leading to the Decline of Arctic Sea Ice From 1996-2015 by Linking Sea Ice Motion With Ancillary Data Products. *AGU Fall Meeting*, C21E-01, 2016.

(6) Tooth, M. Tracking Arctic Sea Ice Parcel Histories. *CU Aerospace Engineering Department Researchpalooza*, 2016.

(7) Tooth, M. Witnessing the Decline of Arctic Sea Ice: How Airborne and Satellite Remote Sensing Techniques Help us Understand the Current and Future States of the Arctic Sea Ice Cover. *CU STEMinar Talks*, 2016.

(8) Tooth, M. Tracking Ice Parcels and Co-Located Data Products in the Arctic Using the Sea Ice Motion and Age Data Products at the University of Colorado, Boulder with Applications to Studying Changes in the Arctic Ice Pack. *INSTAAR 46th International Arctic Workshop*, 2016.

(9) Tooth, M. The Ball Experimental Sea Surface Temperature Radiometer. *IceBridge Science Team Meeting*, 2016.

Appendix B

Acronyms

AGU	American Geophysical Union
APP-x	The Extended AVHRR Polar Pathfinder Product
ATM	Airborne Topographic Mapper
AVHRR	Advanced Very High Resolution Radiometer
BESST	Ball Experimental Sea Surface Radiometer
CCAR	Colorado Center for Astrodynamics Research
CDR	Climate Data Record Program
CSV	Comma Separated Value
CU	University of Colorado Boulder
CULPIS-X	CU-Laser Profiling Instrument - Extended
DAAC	Distributed Active Archive Center
DMS	Digital Mapping Sensor
EASE	Equal Area Scalable Earth Grid
GOCC	Generalized Orthogonal Curvilinear Coordinate

GPS Global Positioning System

IABP International Arctic Buoy Programme

IC Ice Concentration

IMU Inertial Measurement System

IR Infra-Red

IST Ice Surface Temperature

LDA Linear Discriminant Analysis

LIDAR Light Detection and Ranging

LW Longwave Radiation

MIZ Marginal Ice Zone

MODIS Moderate Resolution Imaging Spectroradiometer

MST Mountain Standard Time

NOAA National Oceanic and Atmospheric Administration

NN Neural Network

NSIDC National Snow and Ice Data Center

ONR Office of Naval Research

PIOMAS Pan-Arctic Ice Ocean Modeling and Assimilation System

RF Random Forest

SHEBA Surface Heat Budget of the Arctic Ocean (Campaign)

SIPN Sea Ice Prediction Network

SIZ Seasonal Ice Zone

SIZRS Seasonal Ice Zone Reconnaissance Survey

SSMI Special Sensor Microwave Imager

SSMI/S Special Sensor Microwave Imager Sounder

SVM Support Vector Machine

SW Shortwave Radiation

UAS Unmanned Aerial Systems, Commonly referred to as UAVs

ULS Universal Laser Systems

USCG United States Coast Guard

VIIRS Visible Infrared Imaging Radiometer Suite

Appendix C

Ancillary Data Product Descriptions

This appendix provides additional background on the ancillary data products utilized in the Lagrangian tracking database discussed in Chapter 2 and utilized throughout this study in other chapters. Discussions of the sources of each product are included in each section along with relevant background and error papers. Further discussion of the grids that these products are provided on, and the conversion between them, is included in Appendix D.

C.1 EASE-Grid Sea Ice Motion

C.1.1 Product Description

Polar Pathfinder equal area scalable Earth (EASE) Grid sea ice motion vectors produced at the University of Colorado Boulder are used in this product, and they serve as the foundation of determining ice parcel positions for each week. The data are hosted at NSIDC, and are further described on their documentation page [165]. The data are provided on a 25km EASE-Grid, and are incorporated into this database's 12.5km grid through the use of a two-dimensional linear interpolation scheme. The product is based on ice motion products derived from satellite measurements, buoy drift tracks, and NCEP/NCAR wind data. Weekly mean motion fields are generated by merging motion fields from the various sources by using the expected accuracy of each source field as the weight during the merging process. The database described in this paper utilizes the weekly mean motions computed using those merged fields, as the weekly mean data reduce the contribution from noise present in each individual source. The individual motion field sources have RMS errors that

range from 1-6cm/s [153, 166, 167, 168, 169], which yield annual displacement errors of 50-100km in some cases [163]. Further discussion of the error in the merged motion vector product can be found in [163] and the product's documentation page [165].

C.1.2 Processing Steps

Ice motion vector files are obtained via HTTPS from the NSIDC website [165]. These weekly files are in a space-delimited ASCII text format, and are named with the convention of icemotion.grid.week.YYYY.WW.n.v3.bin. The weekly data files are are processed into .CSV files prior to their use in the parcel tracking program. The conversion from .bin files to .CSV files is accomplished through the use of a Python script that:

(1) Loads the U and V fields from the original .bin file

(2) Scales the motion fields

(3) Masks the values using a pre-made landmask file

(4) Saves two separate .CSV files for the U and V fields

The result is two .CSV files containing the 361x361 grids of the U and V vector fields. These files are saved with the naming convention of motion_YYYY_WW_vec.csv, where vec represents u or v for the respective files. These motion files were the original data utilized in parcel tracking runs, but additional 722x722 files are also generated for use in later versions of the program. These files are generated by loading the 361x361 grid motion data files into a Matlab script that uses a two-dimensional linear interpolation scheme to produce a 722x722 grid for each U and V field [158]. These data are saved in a separate folder under the same naming convention as the other .CSV files, and are used in current runs of the sea ice parcel tracking product.

C.2 EASE-Grid Sea Ice Age

C.2.1 Product Description

EASE-Grid Sea Ice Age fields produced at the University of Colorado are also incorporated into the parcel database, and serve as context for analyses of sea ice survivability. The data are hosted at NSIDC, and are further described in their associated documentation page [164]. The sea ice age data are based on the NSIDC sea ice motion product [165]. Sea ice age is calculated by initially assigning all ice as first-year ice and then, as the model spins up over five simulated years, adding one year of age to each ice parcel that survives the summer melt. The data are available on a 12.5km EASE-Grid through the NSIDC from 1984-2016 [164]. It is important to note that when ice of different ages occupy the same EASE-Grid cell the cell is assigned the age of the oldest ice. Open water cells in this product are those that contain 15% sea ice concentration or less. Additional information about the product can be found in its associated NSIDC documentation page [164] and its related publications [42, 99, 132, 163].

C.2.2 Processing Steps

Ice age files are obtained from NSIDC via HTTPS [164]. The data files are provided as weekly binary files with a specified format that are named with the convention of iceage.grid.week.YYYY.WW.n.v3.bin. Once the binary data files are obtained from NSIDC they are processed into usable .CSV files using a Python script that loads the age field, masks the field using a land mask, and saves the field using the naming convention of age_YYYY_WW.csv. The resulting files contain 722x722 EASE-grid sea ice age data that are incorporated into the Lagrangian tracking database.

C.3 SSMI/SSMI-S Sea Ice Concentration Product

C.3.1 Product Description

SSM/I & SSMIS sea ice concentration values [19] are utilized by the product to determine when parcels are considered melted, and to search for new parcels later in the year. Changes in concentration are additionally useful for tracking the degradation of ice health during the melt season. The data are provided on a 25km Polar-Stereographic grid, and are re-gridded to fit a 25km EASE-Grid before use in this product. Each 25km EASE-Grid cell is then quartered into 12.5km EASE-Grid cells with equal values in order to fit the 12.5km EASE-Grid. The concentration data are less accurate in the presence of thin sea ice, melt ponds, and near the ice edge where some ocean and atmospheric effects can be mistaken for sea ice. The seasonality of these error-producing conditions lends to a $+/-5\%$ accuracy during the winter and a $+/-15\%$ accuracy during the summer [19]. Further discussion of the error and performance of the product can be found in [2,58]. Concentrations in this combined database range from 15% to 100%, with other values masked to 9999.0 and the pole-hole masked to 9998.0.

C.3.2 Processing Steps

The sea ice concentration data are provided in formatted binary files, and are obtained from NSIDC through their HTTPS system [19]. These daily concentration files contain the polar stereographic gridded data, and are named with the convention of nt_YYYYMMDD.SSS.v1.1_n.bin; where SSS represents the source of the data contained in the file. Once the files are obtained from NSIDC, the data are processed into .CSV files using a Python script that:

(1) Loads the 304x448 polar stereographic data

(2) Scales the data

(3) Utilizes Basemap to find usable grid locations

(4) Masks land and the pole-hole

(5) Performs averaging to fill the converted grid's gaps

(6) Saves the files in a .CSV format

The data are saved as .CSV files with the naming convention of YYYY_DOY_concentration.csv. These files contain 361x361 grids of ice concentration on a 25km EASE-Grid that utilize the 9999.0 and 9999.8 masks for land and the pole-hole respectively. The data are re-gridded to the 12.5km EASE-Grid by splitting each 25km grid cell into four equal valued cells [158]. These files are used by the sea ice tracking program, and are averaged into weekly means during runtime.

C.4 PIOMAS Sea Ice Thicknesses

C.4.1 Product Description

The PIOMAS ice thickness model data are incorporated into the product to provide volume loss estimates, and to enable tracking of changes in ice health. The data are provided in a curvilinear coordinate system that is re-gridded to the 12.5km EASE-Grid prior to use in this database. The product is validated using submarine, mooring, and satellite observations in order to compare its model output to available data sources. An estimated RMS difference between the model output and independent submarine tracks is stated as $0.78m$. The model is known to overestimate thin ice thickness and underestimate thick ice thickness [136]. The product is provided on a stretched generalized orthogonal curvilinear coordinate (GOCC) grid with a displaced north pole that is located in Greenland. The mean resolution of the grid is stated as 4–5°, with the greatest resolutions and accuracies being found in the Greenland Sea, Baffin Bay, and the Eastern Canadian Archipelago [182]. Further information about the product and error measurements can be found in [136, 182].

C.4.2 Processing Steps

PIOMAS data are provided in formatted binary files obtained from the University of Washington Polar Science Center website [136]. The files contain GOCC model thickness data for each day of a year, and are named with the convention of hiday.HYYYY. The data are re-gridded to the EASE-Grid and saved in .CSV file format through the use of a Python script that:

(1) Loads and parses scale factors from the original file

(2) Applies several scaling factors to the 365 daily data slices

(3) Performs weekly averaging of the thickness data

(4) Utilizes Basemap to find usable grid locations

(5) Masks the data using a landmask

(6) Performs averaging to fill the converted grid's gaps

(7) Saves the data in a .CSV format

The resulting files contain weekly thickness values on the 722x722 EASE-Grid, and are named with the convention of YYYY_wkWW.csv. These files are directly loaded into the parcel tracking program during runs, and do not require further processing or averaging.

C.5 MODIS Ice Surface Temperatures

C.5.1 Product Description

ISTs from the MODIS instrument aboard NASA's Terra satellite are incorporated in this parcel database. The temperatures indicate when individual ice parcels are at melting temperature, and are provided as context for changes in other variables that are tracked. MODIS ISTs are provided through the National Snow and Ice Data Center (NSIDC) on a 4 km EASE-Grid [50], and are re-gridded to fit the 12.5 km EASE-Grid via a two-dimensional interpolation scheme [158].

Comparisons with other IST sources and in-situ measurements have yielded estimated accuracies of 1–3 Kelvin under ideal conditions [48]. The accuracy of the IST data degrades in the presence of clouds and water vapor, which lends to seasonally lower accuracy during the summer [135, 139]. Valid ranges for this data product are $210K$ to $313.2K$, pole-hole points are masked to 9999.8, and all other values are masked to 9999.0. It is important to note that while other products may be available outside of the 2001 to 2016 date range, these data form a complete year starting in 2001. This limits the current version of the parcel database to years including and after 2001 due to its use of these data.

C.5.2 Processing Steps

Ice surface temperature data are obtained from the NSIDC via HTTPS [50]. Each daily file is provided in HDF-EOS format with the naming convention of MOD[PID].A[YYYY][DDD].[VVV].[yyyy][ddd][hhmmss].hdf; where PID represents the product ID, VVV represents the version number, and the lower-case dates represent the processing date and time. The files are converted to .CSV format through the use of a Matlab script that:

(1) Loads the HDF file and obtains the northern IST field

(2) Applies a scaling factor to the data

(3) Crops the data field to fit our analysis region

(4) Utilizes a two-dimensional linear interpolation scheme to obtain a new grid

(5) Saves the new grid to a .CSV format file

The resulting files contain daily IST data on the 722x722 EASE-Grid named with the convention of YYYY_DOY.csv. These files are read and averaged to obtain weekly values during runtime.

C.6 APP-x Atmospheric Products

C.6.1 Product Description

Broadband albedo, downwelling shortwave, and downwelling longwave are obtained from the AVHRR Polar Pathfinder-Extended (APP-x) dataset [70, 75]. The data are provided on a 25 km EASE-Grid that is re-gridded to a 12.5 km grid by splitting each 25 km grid cell into four equal-valued cells [158]. Albedo is corrected for clouds, and is derived through the procedures described in [76]. Comparisons of APP-x albedos with SHEBA measurements yielded a bias of -0.05 and an RMSE of 0.1 [70]. Downwelling shortwave and longwave fluxes at the surface are computed using FluxNet: a neural network that was trained to simulate a radiative model [73]. Comparisons of downwelling shortwave radiation with SHEBA campaign measurements obtained a bias of $9.8\frac{W}{m^2}$ and an RMSE of $34.4\frac{W}{m^2}$. Comparisons between downwelling longwave and Surface Heat Budget of the Arctic Ocean (SHEBA) Campaign values yielded a bias of $2.1\frac{W}{m^2}$ and an RMSE of $22.4\frac{W}{m^2}$ [70].

C.6.2 Processing Steps

APP-x data are obtained from the NOAA Climate Data Record Program site via direct download [75]. The data are provided as NetCDF files with the naming convention of Polar-APP-X_VerNumRevNum_Nhem_TTTT_dYYYYMMDD_cyyyymmdd.nc; where lowercase dates represent the date the data were processed and released on the CDR site. The data are re-formatted and split into separate .CSV files through the use of a Matlab script that:

(1) Loads the .nc file data

(2) Parses the shortwave, longwave, and albedo variables

(3) Saves the fields as separate .CSV files

The resulting individual files contain daily values of each variable with the naming convention of YYYY_DOY_TTTT_var.csv, where var represents a coded name for each of the five variables tracked. These data are averaged into weekly mean values during runtime.

C.7 IceBridge Airborne Data

C.7.1 Product Description

Airborne data from spring NASA Operation IceBridge Campaigns were also incorporated into this analysis for years 2009–2016 [82]. The level-4 airborne data were obtained from NSIDC on a 40 m length scale, and are further described in their related documentation page and publication [80, 82]. The data include KT-19 infrared pyrometer IST values [78], snow depth estimates retrieved using the University of Kansas' snow radar [63, 88, 133], Digital Mapping System (DMS) derived open water concentrations [30], and Airborne Topographic Mapper (ATM) derived thickness estimates [15, 31, 85].

IceBridge thickness data rely on ATM estimates of freeboard and retrieval of snow depth from radar measurements [80, 82]. Freeboard measurements are further refined through the use of DMS measurements to detect leads that are used as surface tie points [80], and final thickness is determined through the use of the hydrostatic balance equation [82]. The mean uncertainty of the thickness data used in this study was 0.82m, and was obtained by averaging the provided uncertainty in each measurement utilized in this analysis. This value is highly variable however, as it depends on several factors that can change in flight [81]. The uncertainty in snow depth estimates is cited as 5.7–5.8 cm [38, 80, 181].

C.7.2 Processing Steps

IceBridge data files were obtained for all available flight lines in the Arctic from 2009-2016. Examples of a flight line and associated snow thickness data are shown in Figure C.2. The data are downloaded through the HTTPS system, and are named with the convention of IDCSI4_YYYYMMDD.txt. A Python script was developed to sort these data into files that contain flight line data from major regions of the Arctic (Figure C.1). The 40m observations were averaged for each 12.5km EASE-Grid cell to obtain mean value of thickness, snow depth, concentration, and surface temperature for each cell. These EASE-Grid IceBridge files were combined into weekly files

for each region for use in this manuscript.

C.8 Sea Ice Convergence Fields from EASE-Grid Sea Ice Motion Data

Convergence values for the u and v EASE-Grid axes are produced using the weekly motion vectors described in Section C.1. In order to compute a convergence value in each direction, an individual EASE-Grid cell's motion vector components are compared to the motion vector components of directly adjacent cells. The boundary convergence calculations are performed using the equations C.1 & C.2 for the v direction, and equations C.3 & C.4 for the u direction:

$$C_{top} = V - V_{above} \tag{C.1}$$

$$C_{bottom} = V_{below} - V \tag{C.2}$$

$$C_{right} = U - U_{right} \tag{C.3}$$

$$C_{left} = U_{left} - U, \tag{C.4}$$

where C_{top} & C_{bottom} represent the convergence between the parcel of interest and the parcels above and below it, and C_{left} & C_{right} represent the convergence between the parcel of interest and the parcels to the left and right of it. The U and V terms represent the x and y grid velocity components of each parcel respectively.

The location of each of these equation elements, along with a representation of the parcels used in the calculation, are shown in Figure C.3. In this grid system V vectors represent the vertical component and are positive-upward, while U vectors represent the horizontal component and are positive-right. It is important to note that the u and v coordinates are 0,0 at the top left corner of the grid, which lends to the V vectors being inverted with respect to the grid coordinates. Total convergence values for the cell in the u and v directions are then derived by summing the obtained boundary values on each axis:

$$C_v = C_{top} + C_{bottom} = V_{below} - V_{above} \tag{C.5}$$

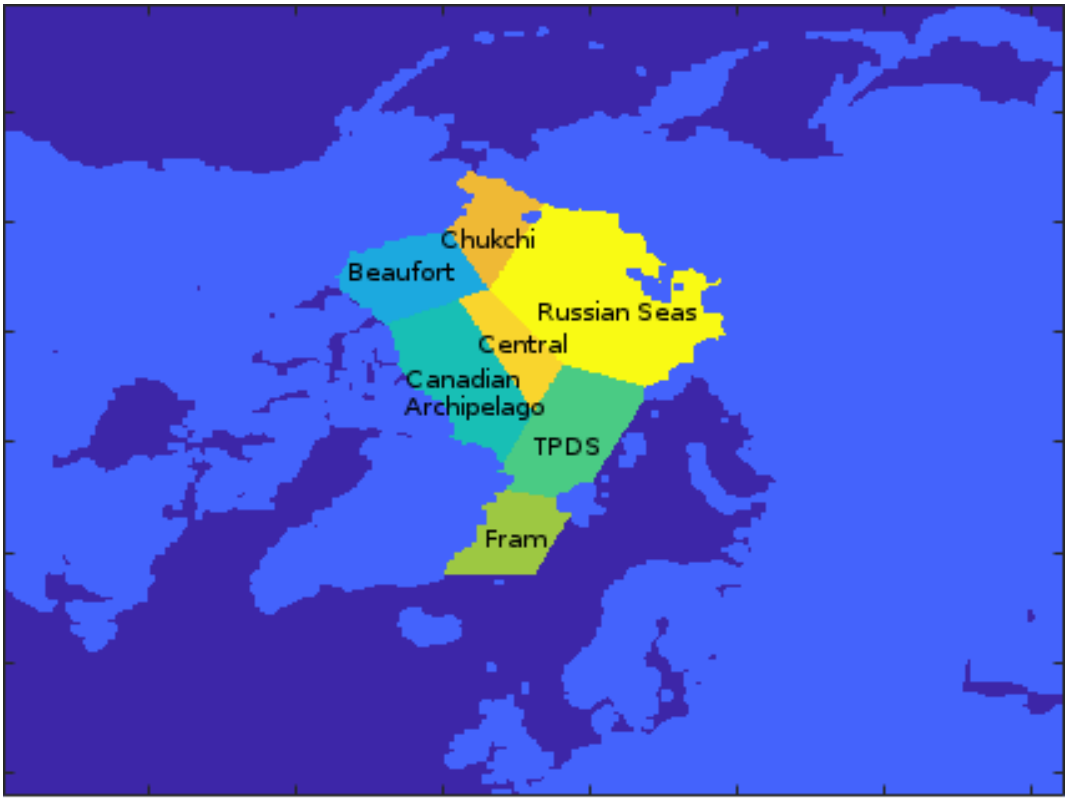

Figure C.1: Arctic regions used for IceBridge data

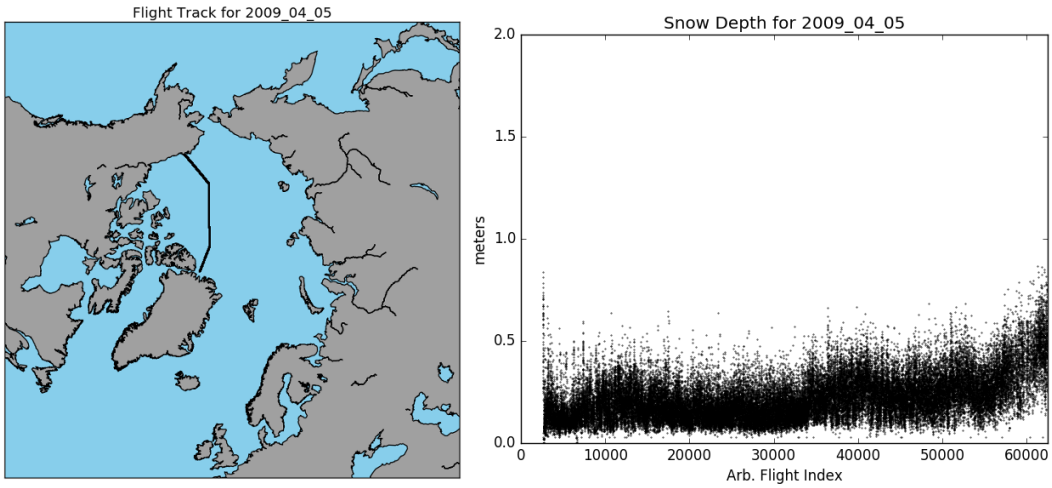

Figure C.2: IceBridge flight track from April 5^{th}, 2009 with corresponding snow thickness data.

$$C_u = C_{left} + C_{right} = U_{left} - U_{right} \tag{C.6}$$

The resulting numbers represent the speed at which a parcel is approaching its neighboring parcels, where larger positive values of C represent more convergent motion. This number serves as an indicator of greater potential for ice collision and deformation due to this convergence. Negative values of C indicate that the neighboring parcels are receding, and that other processes such as lead formation may be more likely for that parcel. These data are provided to aid researchers in determining where relationships like these may occur.

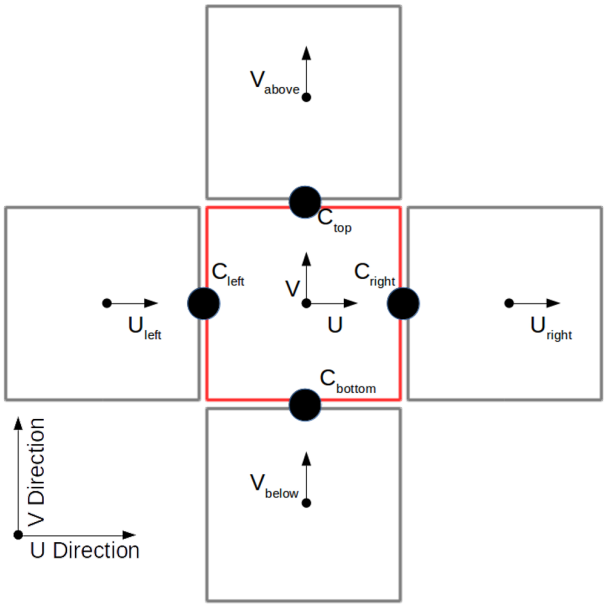

Figure C.3: EASE-Grid cell with vector components used to compute convergence values shown

Appendix D

Map Coordinate Systems

This appendix describes the grids on which the products discussed in Appendix C are provided. These include the EASE-Grid, Polar Stereographic Grid, and Generalized Orthogonal Curvilinear Coordinate Grid. The procedure through which data products were re-gridded to fit the EASE-Grid is also provided, along with a brief discussion of filling gaps in data caused by warping during conversion.

D.1 EASE-Grid 2.0

The Equal Area Scalable Earth (EASE) Grid 2.0 grid format was defined in 2011, and is the follow-on to the NSIDC's EASE-Grid 1.0 format [14] (Figure D.1). The map projection has been updated to rely on the WGS84 ellipsoid, which offers improvement over the spherical Earth ellipsoid used in EASE-Grid 1.0. The pole is located at the intersection of four cells in EASE-Grid 2.0, and the 12.5km and 25km grids are 722x722 and 361x361 cells in dimension respectively. Many of the data products used in this work are natively hosted on EASE-Grid. Products obtained on other grids were converted to the EASE-Grid through the use of the Python Basemap package.

D.2 Polar Stereographic Grid

The Polar Stereographic grid is a projection that specifies a grid tangent to the surface of Earth at 70°, and is defined using an Earth radius of 6378.273km and an eccentricity of 0.081816153 (Figure D.2. The data products used from the Polar Stereographic grid were provided at 25km

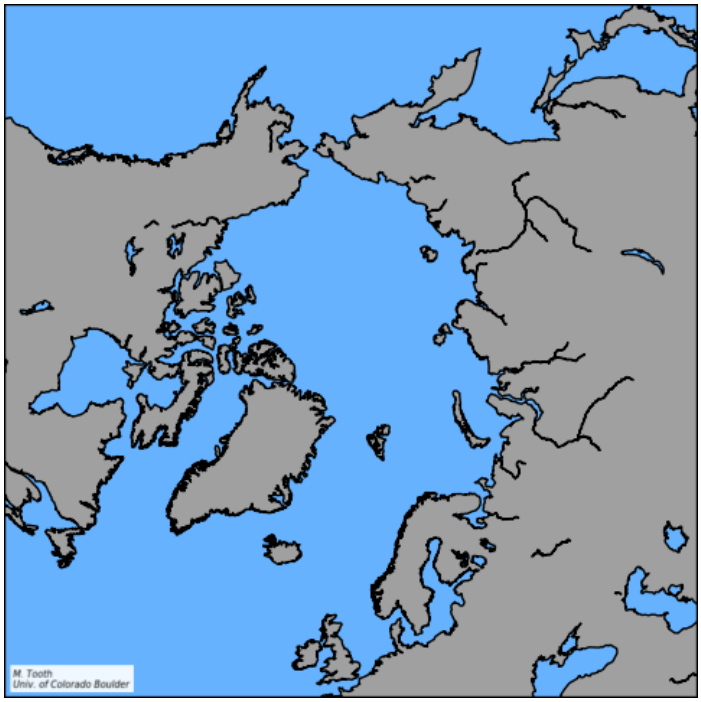

Figure D.1: EASE-Grid 2.0 map of the Arctic, generated using the Python Basemap package

resolution (304x448 cells). Data provided on the Polar Stereographic Grid were converted using latitude/longitude coordinates provided by NSIDC through the use of the Python Basemap package.

D.3 Generalized Orthogonal Curvilinear Coordinate Grid System

PIOMAS sea ice thickness data are provided in a 120x360 generalized orthogonal curvilinear coordinate grid system (GOCC) that covers latitudes 45°-90° (Figure D.3). Cells in the grid vary in size, with the greatest resolutions and stated accuracies being found near the artificial pole over Greenland. PIOMAS data were converted to EASE-Grid through the use of latitude/longitude files that were provided with the data [136].

D.4 Converting Between Grid Coordinates

The EASE-Grid was chosen as a common grid for use in this manuscript, as the sea ice motion data product that formed the basis of this work is provided on a 25km EASE-Grid [165]. Conversion between the various source grids to the EASE-Grid was carried out through the use of the Python Basemap package, and the procedures used to do so are detailed further in the following sections.

D.4.1 Converting Grids With Coordinate Locations

Conversion to EASE-Grid utilized the Basemap parameters described in Table D.1. EASE-Grid coordinates were generated with the map instance in Table D.1 using the provided latitude/longitude coordinates for each non-EASE-Grid product (Figures D.4 & D.5). Once map coordinates were obtained the products were converted to EASE-Grid prior to use in generating the Lagrangian tracking database described in Chapter 2.

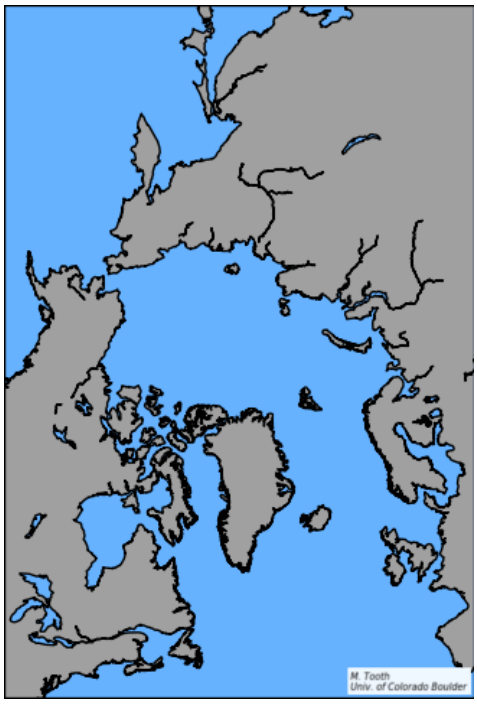

Figure D.2: Polar Stereographic map of the Arctic, generated using Basemap.

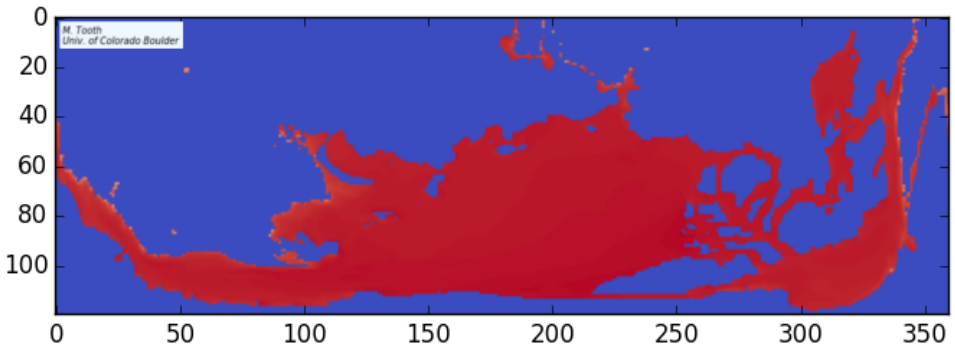

Figure D.3: Representation of the PIOMAS GOCC grid

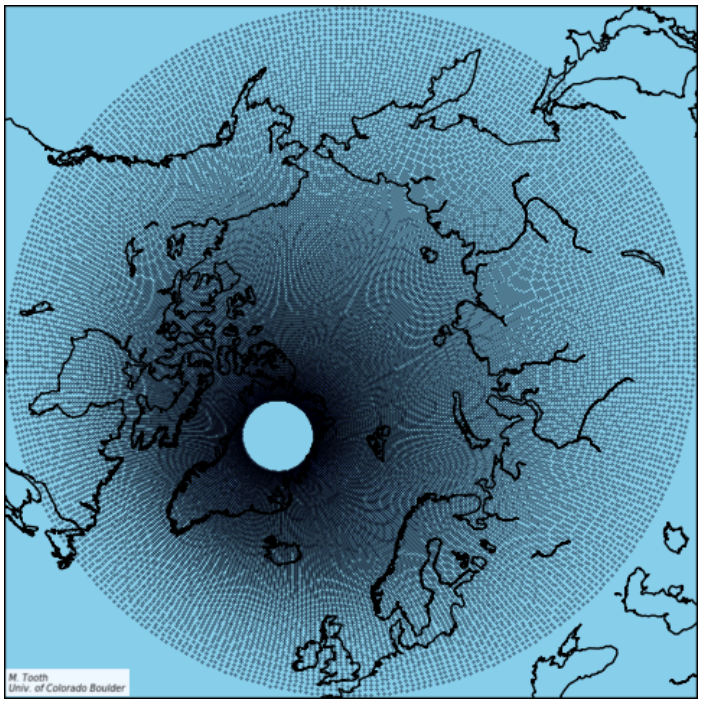

Figure D.4: PIOMAS model coordinates plotted on EASE-Grid

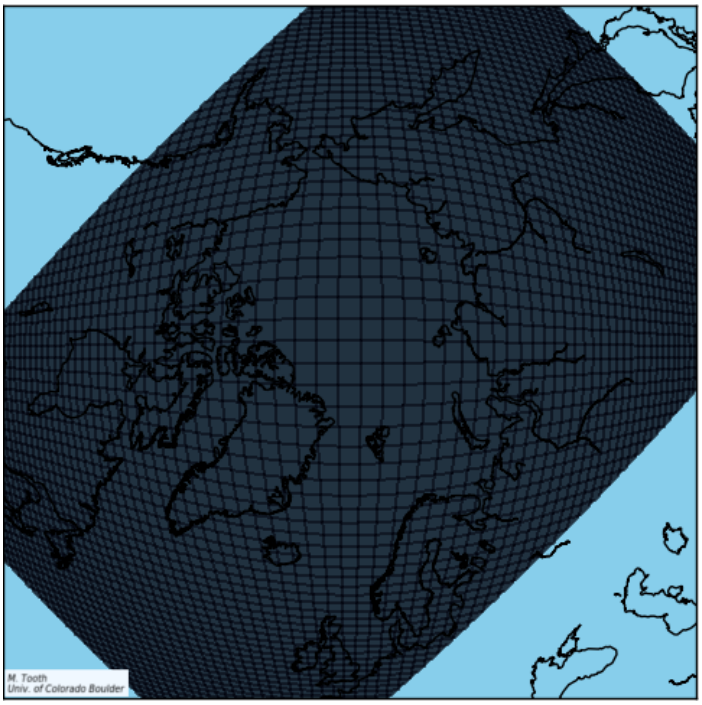

Figure D.5: Polar Stereographic coordinates plotted on EASE-Grid

Table D.1: Basemap instance parameters for grid conversions

Variable	**2016**
ellps	WGS84
resolution	l
projection	laea
lat_ts	90
lat_0	90
lon_0	0
width	9025000
height	9025000
area_thresh	10000

D.4.2 Filling Empty Grid Cells After Converting Coordinates

Warping of grid coordinates during conversion to EASE-Grid yielded empty cells in some data products that were not present in their native grids (Figure D.6). Averaging was applied to the converted grid data to rectify this issue during which empty cells with at least three adjacent cells with data were filled with the average of their neighboring cells (Figure D.7).

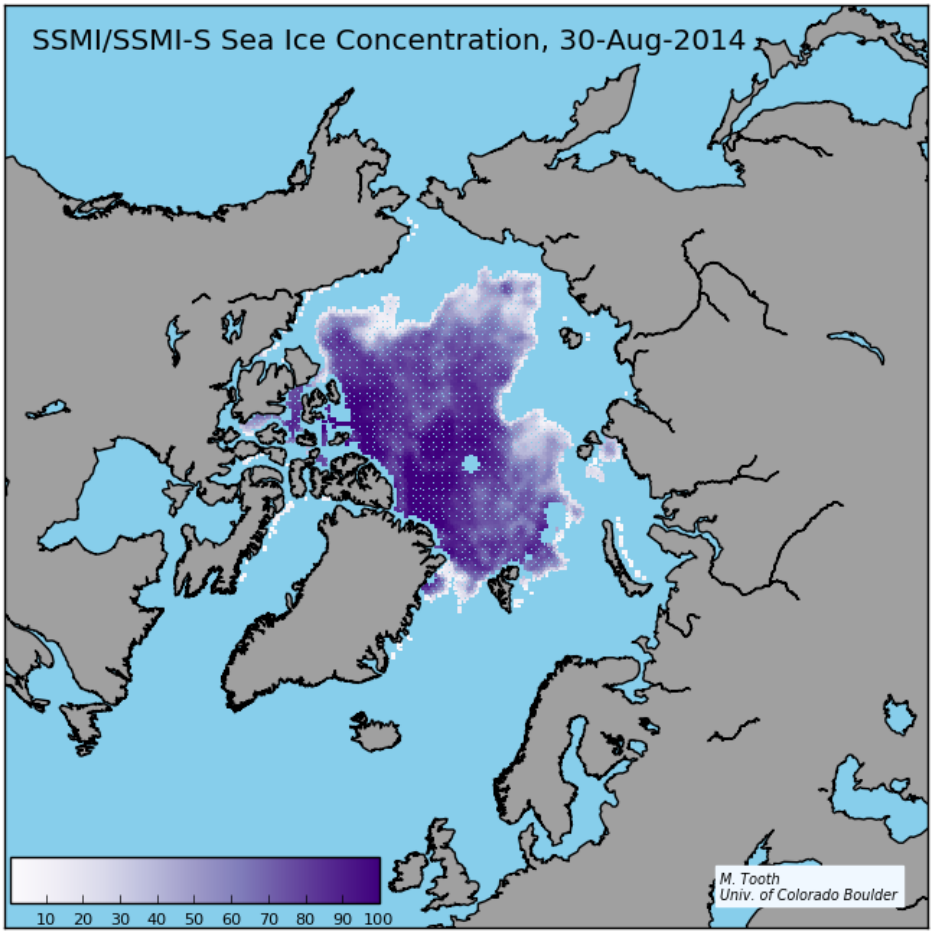

Figure D.6: Un-filled sea ice concentration field for August 30^{th}, 2014

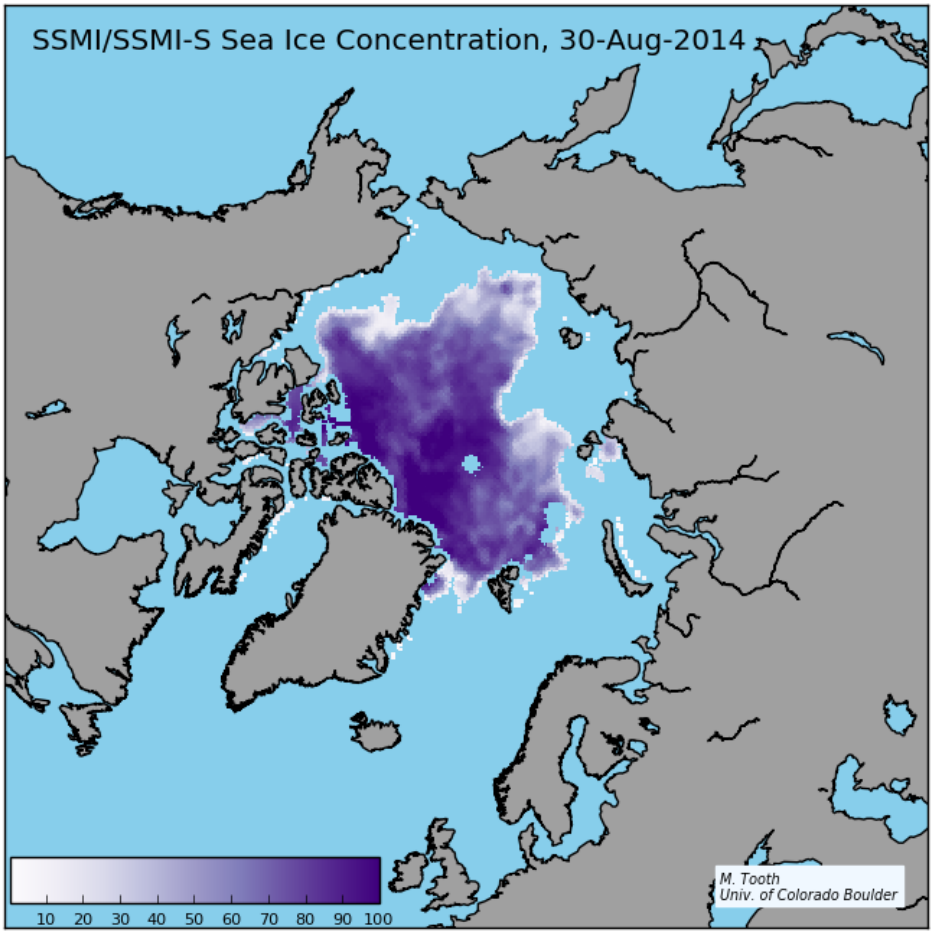

Figure D.7: Filled sea ice concentration field for August 30^{th}, 2014

Appendix E

Additional Information on Tracking Product

E.1 History of Product Development

The Lagrangian tracking product described in Chapter 2 has gone through many iterations from the start of development to its present release state. The initial code was developed in Fall 2015 as a standalone Matlab script. This initial release included EASE-Grid sea ice motion vectors, MODIS ice surface temperatures, and SSMI/SSMI-S ice concentration data that was used to present preliminary results at AGU in Fall 2015 [155]. The version history of this Matlab program branch is included below for the reader's reference:

- Version 1.0
    - Formed main program loop
    - Created structure to store weekly parcel positions
- Version 1.2
    - Added ice concentration and surface temperature tracking
- Version 2.0
    - Added ability to loop through multiple years of data
    - Added ability to save data at transition between years
- Version 2.2

    * Added ability to track parcels during freeze-up

- Version 2.4

    * Finalized logic of tracking parcel formation and extinction

    * Began preliminary work toward creating Python branch

A Python version of the code was developed in Spring 2016, and came to relative maturity in Summer 2016 [157]. This new branch of the code was created to facilitate modularizing the code, creating well commented work for research heritage, and to ensure compatibility with several Python re-gridding utilities.

The Python branch of the code was continually updated through 2016, and a release-ready set of database files were produced in Spring 2017. Further upgrades were aimed at making the code more accessible and practical to use. The version history notes from the code are included below for the reader's reference:

- Version 0_1

    * Initial Translation to Python from Matlab

- Version 0_2

    * Inclusion of IST and Concentration Products

    * New Vector and Age Files

    * Ability to Save Data to File

- Version 0_4

    * Inclusion of Freeze-Up and Melting Detection

    * Creation of Main Function Layout

- Version 0_6

    - Inclusion of PIOMAS Data

- Version 0_8

    - Inclusion of APP-X Data
    - Tracking of Float Parcel Positions in Lieu of Integers
    - New Mask Value for Pole-Hole in Concentration Data
    - Updated Initial Search to use Concentration in Lieu of Age Field
    - Began Work on Transition Between Years Code

- Version 0_9

    - Allowed Transition Between Years (Without Saving)
    - Re-Gridded Motion fields to 12.5km Grid

- Version 1_0

    - Release Version - Used to Generate 2001-2015 Files

- Version 2_0

    - Swapped NaN Fill Values with 9999
    - Added Pole-Hole Fill Values of 9998
    - Created and Included Convergence Fields
    - Added New Weekly Product Age Field

- Version 2_1

    - Completed Inter-Annual Transition Code
    - Corrected Error in PIOMAS Data
    - Updated Files to Allow Runs Through 2016
    - Second Release Version for Pangaea 2001-2016

Appendix F

Additional Aiborne Data Sources in the Beaufort Sea

Airborne observations carried out at the University of Colorado yielded data in the Beaufort Sea that were considered for use in Chapter 5. This appendix describes the Ball Experimental Sea Surface Temperature (BESST) radiometer and some preliminary results obtained by the instrument during two separate NASA campaigns in 2013 and 2016. The CU-Laser Profiling Instrument Extended (CULPIS-X) is also described, along with results from Arctic Domain Awareness flights carried out during 2016.

F.1 BESST

The Ball Experimental Sea Surface Temperature (BESST) radiometer is an IR instrument that obtains images of ocean and ice surfaces (Figure F.1). It samples in the 8-11μm range, and is calibrated throughout flight [35, 154]. Calibration and image acquisition are performed through the use of a rotating internal fold mirror that offers views of nadir, an ambient blackbody, zenith, and a hot blackbody. Zenith and blackbody measurements are obtained periodically for use in post-processing.

The BESST instrument was flown during the NASA 2013 Marginal Ice Zone Processes EXperiment (MIZOPEX) and 2016 IceBridge campaigns in order to obtain IR data for the ice-ocean surface [97]. Attempts were made to compare the resulting data with coincident satellite data for further use in the Beaufort Sea study, but they were hampered by data processing issues and missing data. Scheduling issues and instrument drift led to limited BESST data during MIZOPEX, and also

yielded unrealistically high surface temperatures as observed by the instrument. Post-processing testing failed to determine the cause of this issue, which prevented the use of these data in Chapter 5. During the IceBridge campaign the instrument was situated ahead of a pressure bulkhead in the P3 radome, which caused the control computer to fail once the temperature dropped below $0°C$ at altitude. This led to there being limited data during the campaign, with none of the valid acquisitions being located in the Beaufort Sea. The failure of BESST during the 2016 IceBridge campaign is also the cause of the missing IST measurements for 2016 in Table 5.4.

F.2 CULPIS-X

F.2.1 Instrument Package Overview

The CU Laser Profiling InStrument - eXtended (CULPIS-X) instrument package was deployed on Arctic research flights with the US Coast Guard (USCG) as part of the 2016 Seasonal Ice Zone Reconnisance Survey (SIZRS) Campaign [26, 108, 162, 171]. The instrument package contains a profiling LIDAR, an IR sensor, a multi-spectral instrument, and a visible camera. These components are mounted in an instrument box on the rear cargo door of a C-130 via a sled that is secured to the interior of a flare tube (Figure F.2).

After initial on-ground testing in 2015, the instrument package was successfully flight tested at USCG Air Station Sacramento via a transit down the California coast from Sacramento to Point Mugu Naval Air Station in Spring 2016. This test was followed by five successful Arctic deployments in June, July, August, September, and October of 2016 via USCG Air Station Kodiak. These flights yielded visible camera data that were considered for use in determining sea ice concentration along the flight tracks for comparison against coincident satellite observations (Figure F.3). Initial investigation of these data aimed to implement k-means clustering and other statistical learning techniques on the data, but these efforts were hampered by instrument timing issues and GPS inaccuracy (Figure F.4). These issues led to these data being unsuitable for the work discussed in Chapter 5.

Figure F.1: Assembled BESST instrument [35]

Figure F.2: CULPIS-X Package installed in a flare tube prior to a June 2016 Arctic Flight

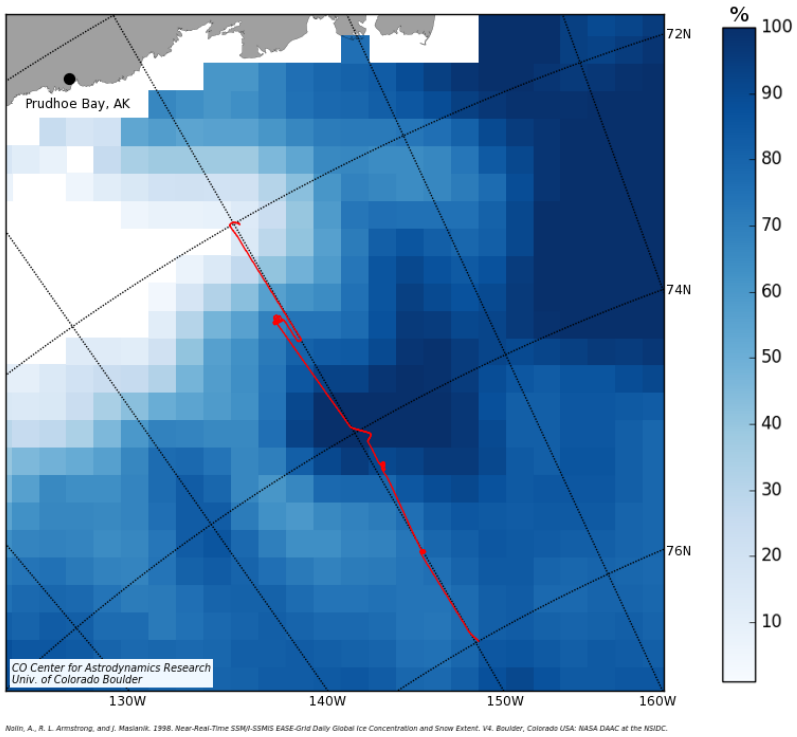

Figure F.3: CULPIS-X flight track with coincident ice concentration data from June 15^{th}, 2016

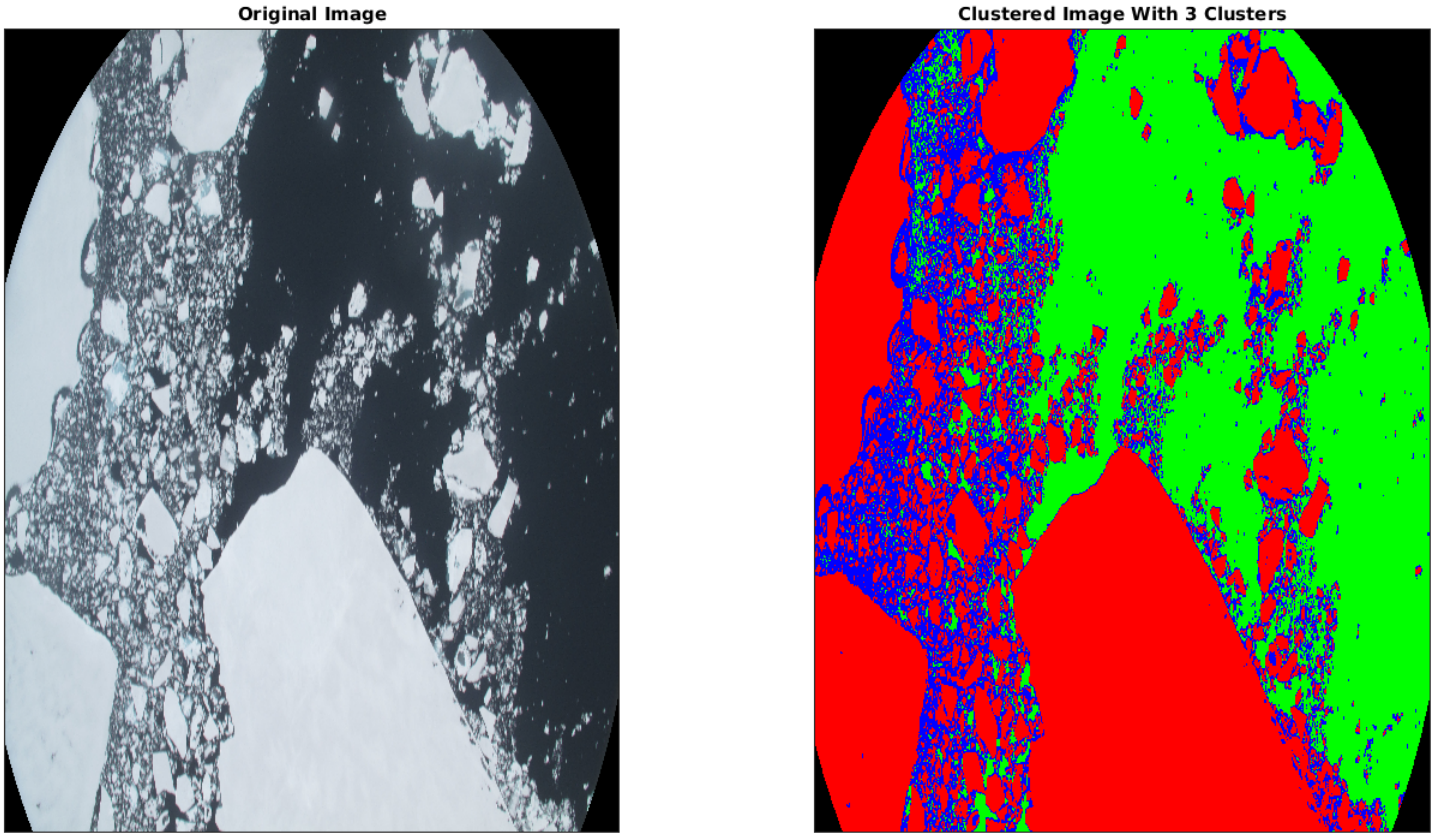

Figure F.4: Visible image of ocean surface from CULPIS-X (left), and k-means clustered image with three clusters (right)