Remote Sensing of Open Water Fraction and Melt Ponds in the Beaufort Sea Using Machine Learning Algorithms

Macarena Ortiz
University of Miami, macortiz85@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/1957

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
REMOTE SENSING OF OPEN WATER FRACTION AND MELT PONDS IN THE BEAUFORT SEA USING MACHINE LEARNING ALGORITHMS

By

Macarena D. Ortiz

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Coral Gables, Florida
August 2017
REMOTE SENSING OF OPEN WATER FRACTION AND MELT PONDS IN THE BEAUFORT SEA USING MACHINE LEARNING ALGORITHMS

Macarena D. Ortiz

Approved:

Hans C. Graber, Sc.D.
Professor of Ocean Sciences

Roland Romeiser, Dr.rer.nat.
Associate Professor of Ocean Sciences

Peter Minnett, Ph.D.
Professor of Ocean Sciences

Jeremy Wilkinson, Ph.D.
Senior Scientist
British Antarctic Survey, UK

Frank Monaldo, M.S.
Principal Staff Physicist
Johns Hopkins University
Applied Physics Laboratory

Guillermo Prado, Ph.D.
Dean of the Graduate School
The classification of open water fraction (OWF) from synthetic aperture radar (SAR) images in the marginal ice zone (MIZ) can be significantly difficult during the summer months, where melt-onset can alter the radar backscatter and melt ponds contribute to errors in OWF estimates. In this study we apply five machine learning classifiers in combination with image textures to derive OWF from TerraSAR-X stripmap images of the MIZ in the Beaufort Sea from the summer of 2014. The classifiers used include (1) Neural Networks, (2) Linear Support Vector Machines, (3) Naive Bayes, (4) K-nearest neighbor and (5) discriminant analysis. We validate our results using near-coincident high resolution optical images from a panchromatic imager with 1-m resolution as literal image derived products (LIDPs). Our results show that the different classification algorithms report overall similar results. The K-nearest neighbor algorithm achieved the highest OWF image percentage correlations from the training data against the validation dataset \((R^2 = 0.91)\), while the highest correlations from the testing dataset are reported by the Neural Network \((R^2 = 0.86)\). The Naive Bayes algorithm was the least computationally expensive method; the computation time to classify a full-sized SAR image using the Naive Bayes algorithm was roughly 7-10 minutes, while the Neural Network and Support Vector Machines approximated 25-30 minutes. Faster computation can be very practical for users on vessels
wishing to have “on-the-fly” and accurate methods to calculate ice/water in the general vicinity for navigational purposes.

In addition, much work has been published on determining changes in summer ice albedo and morphological properties of melt ponds such as depth, shape and distribution using in-situ measurements and satellite-based sensors. Although these studies have dedicated much pioneering work in this area, there still lacks sufficient spatial and temporal scales. We present a prototype algorithm using Linear Support Vector Machines designed to quantify the evolution of melt pond fraction from a recently government-declassified high-resolution panchromatic optical dataset, in an area where several in-situ instruments were deployed by the British Antarctic Survey in joint with the Marginal Ice Zone Program, from April-September, 2014. We explore both the temporal evolution of melt ponds and spatial statistics such as pond fraction, pond area, and number pond density, to name a few. We also introduce a linear regression model that can potentially be used to estimate average pond area by ingesting several melt pond statistics and shape parameters.
For Gus, the light and joy of my life.
Acknowledgements

There were so many people who helped me during the course of my Ph.D, who I would like to thank. First, I would like to thank Hans C. Graber, my advisor, who was nothing but supportive throughout these years. Hans, you let me go to the Arctic, and believed that I could succeed in −37°C weather, carrying a rifle and taking measurements in complete darkness, even though I called you in the middle of the field course and told you I was not going to make it. You also let me attend a NASA summer internship, one year before graduation, which you did not like very much, but you knew that it would benefit my career in the long-run. Thank you for including me in the many MIZ meetings we had over the years, I truly felt like I was part of this project. I also want to thank you for letting me go to so many conferences, I know that most advisors would not let their students travel as much as you let me. I truly had a once in a lifetime opportunity in working with you. Thank you for supporting me not just professionally, academically, and financially, but also emotionally. You always helped me see there was a solution to everything, and to take life one step at a time. I will be forever grateful, and will always say that I had the best Ph.D advisor in the world!

Second, I would like to thank my committee members: Roland, Frank, Jeremy and Peter. I would like to thank you all for your invaluable help, and knowledge. You all helped this project come to light and I want to take a moment to thank each of you individually.

Roland, you really deserved RSMAS’ best mentor award. You are a great teacher and you have a tremendous amount of patience. Thank you so much for sitting with me all those hours, going through my proposal, thesis edits, and codes. You always took the time to help me go through tedious problems and I cannot image having gone through this Ph.D without your expertise and kindness. Thank you also for being a good friend and being concerned about me when things got rough along the way.
Frank, I would like to thank you for agreeing to be my committee member, even though you did not know me very well. Thank you for all those times that I called you on the phone to help me with Ulaby’s fading statistics! You were also a great help in my quest to learn LaTeX. Your positive comments on my efforts to learn this program gave me a lot of confidence and a desire to keep using it. I also want to thank you for clarifying many SAR concepts that I was confused on, I certainly learned a lot from you and hope I can keep learning still.

Jeremy, thank you so much for all the help you provided me with, especially all the data. Thank you for sitting with me after all the MIZ meetings and going through what I had done and planning what to do next. You were the ice expert in this project, and without your knowledge, this would have been extremely difficult, if not, impossible to achieve. Also, thank you so much for coming to see my defense in person! Having traveled from the U.K just for my defense was truly an honor for me, and I hope that I did not disappoint.

Peter, it was an honor to have you as a committee member. You have so much knowledge about remote sensing and the Arctic, and your expertise helped me a great deal in this project. Thank you so much for your hard work and detailed comments. Being able to tie in concepts from the atmosphere and the ocean was no easy task, but you certainly made me think about the importance of these concepts and this work hopefully reflects an attempt in acknowledging this. I also want to thank you for attending all those CSTAMP seminars, that made me feel very supported.

Next, I would like to thank CSTARS. Thank you for all the help in collecting and processing the satellite data in this work, especially John Hargrove, Mike Caruso and Raymond Turner. John, I cannot thank you enough for letting me bug you all these years, calling you on the phone, in desperation! You were an amazing help to me in this project. Thank you for sitting with me for hours on end when I visited CSTARS, for listening to my ideas and for being an emotional support. Ret, you were literally the backbone of all the data. Thank you for taking the time to help me with processing all the SAR images and for constantly solving computer issues and log-in problems, you’re a life saver!
I would also like to thank my family and friends, for being so supportive in following my dreams. I want to thank my mom for telling me never to give up on my goals. I know it was hard on you when I moved to a different state, but you knew that it was all for the benefit of my career. Thank you for always telling me how proud I make you, I hope I keep making you proud. Marcela, my sister, thank you for being the rock of this family. I am so grateful to have you as my big sister. You have always supported me in this endeavor and, I hope that I can do the same for your Ph.D. You can always count on me! Hanii, thank you so much for being the best of friends to me. You always supported me and believed in me all these years, even though I doubted myself. Thank you for not letting me give up, I love you! I would also like to thank Dan. You have been an amazing pillar of strength for me during the last few months of this process. Thank you for constantly giving me words of encouragement and cheering me on. You have no idea how much that meant to me, and how special you make me feel. I hope that we can always have such admiration and respect for one another and that we continue to be a great team for many years to come.

Finally, I want to thank The Office of Naval Research, who funded this project. Your generosity helped this dissertation come to light. Thank you for supporting me and all the great scientists, engineers and students in the MIZ program. I hope that this work can aid more programs like this in the future, because the Arctic is such a special place that needs attention and many questions remain unanswered about the fate of sea ice.

MACARENA D. ORTIZ

University of Miami

August 2017
# Table of Contents

**LIST OF FIGURES**

**LIST OF TABLES**

1 **INTRODUCTION**

2 **SEA ICE**

2.1 Ice Growth

2.2 A Simple Sea Ice 1-Dimensional Growth Model

2.3 The Role of Sea Ice and Clouds in the Arctic Radiation Budget

2.4 The Marginal Ice Zone
2.5 Chapter 2 Summary .................................................. 34

3 MELT PONDS .................................................. 36

3.1 Stages of Pond Evolution .............................................. 38

3.1.1 Stage 1 .................................................. 38

3.1.2 Stage 2 .................................................. 41

3.1.3 Stage 3 .................................................. 42

3.1.4 Stage 4 .................................................. 42

3.2 Chapter 3 Summary .................................................. 42

4 REMOTE SENSING OF SEA ICE .................................. 44

4.1 SAR .......................................................... 45

4.1.1 Speckle Noise .............................................. 55

4.1.2 Sea Ice Backscatter During Winter ......................... 60

4.1.3 Sea Ice Backscatter During Summer ....................... 61

4.2 Optical ...................................................... 64

4.3 Chapter 4 Summary .................................................. 64

5 DATA USED .................................................. 66

5.1 Satellite Data ...................................................... 67

5.2 In-situ Data ....................................................... 72

5.2.1 Ice Mass-Balance Buoy ..................................... 72

5.2.2 Autonomous Weather Station ............................... 73
### List of Figures

1.1 Map of the Arctic Ocean .......................................................... 2
1.2 The Beaufort Sea ................................................................. 3
1.3 Marginal Ice Zone Dynamics .................................................. 4
1.4 Sea Ice Retreat Captured by MODIS ....................................... 5

2.1 Temperature vs. Salinity ......................................................... 10
2.2 Photograph of Grease Ice ......................................................... 11
2.3 Photograph of Pancake Ice ....................................................... 12
2.4 1-D Sea Ice Model Slab .......................................................... 12
2.5 Normalized Blackbody Radiation Curves ................................... 17
2.6 Observed Albedo Values for Sea Ice .......................................... 19
2.7 Ice Thickness for Constant Surface Temperature ....................... 26
2.8 Ice Thickness for Constant Surface Temperature with 5 W m\(^{-2}\) Oceanic
    Heat Flux .............................................................................. 27
2.9 Ice Thickness for Constant Surface Temperature with 180 W m\(^{-2}\) Oceanic
    Heat Flux ............................................................................. 28
2.10 Ice Thickness for Constant Surface Temperature and Snow ........... 29
2.11 Ice Thickness for Constant Surface Temperature and Snow with Oceanic Heat Flux .................................................. 30
2.12 Ice Thickness for Changing Surface Temperature, without Surface Melt ................................................................. 31
2.13 Ice Thickness for Changing Surface Temperature with Surface Melt ................................................................. 31

3.1 Melt Pond Photograph ......................................................... 37
3.2 Melt Pond Mechanism ......................................................... 39
3.3 Macroscopic Drainage Flows ................................................ 41

4.1 The Electromagnetic Spectrum ........................................... 45
4.2 Geometry of a SAR System ............................................. 46
4.3 Radarsat-2 Incidence Angle Dependence Example ...................... 50
4.4 Scattering Behavior for Varying Surfaces ................................ 52
4.5 Example of Ice Ridges in SAR ............................................ 53
4.6 Example of Microwave Backscatter from Different Ice Types ........... 54
4.7 SAR Speckle Statistical Distributions .................................... 59
4.8 Radarsat-2 Sea Ice Image in Freezing Conditions ...................... 61
4.9 TerraSAR-X Sea Ice Image in Melting Conditions .................... 63
4.10 Panchromatic Images Showing Sea Ice and Melt Pond Development ........................................... 65

5.1 MIZ Arrays ................................................................. 67
5.2 TSX Image and Zoom ........................................................ 68
5.3 RS2 Image and Zoom ........................................................ 69
5.4 Optical Satellite Image Examples ......................................... 71
5.5 Schematic of Ice Mass-balance Buoy ...................................... 73
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.9</td>
<td>Example of Ice Divergence in TSX</td>
<td>128</td>
</tr>
<tr>
<td>8.10</td>
<td>IMB Daily Averaged Air Temperatures</td>
<td>129</td>
</tr>
<tr>
<td>8.11</td>
<td>AWS Daily Averaged Air Temperatures</td>
<td>129</td>
</tr>
<tr>
<td>8.12</td>
<td>AWS Daily Averaged Wind Speeds</td>
<td>131</td>
</tr>
<tr>
<td>8.13</td>
<td>Radarsat 2 Open Water Fraction Evolution</td>
<td>131</td>
</tr>
<tr>
<td>8.14</td>
<td>TerraSAR-X Open Water Fraction PDF</td>
<td>132</td>
</tr>
<tr>
<td>8.15</td>
<td>Radarsat 2 Open Water Fraction PDF</td>
<td>133</td>
</tr>
<tr>
<td>8.16</td>
<td>TerraSAR-X and Radarsat 2 CDF</td>
<td>133</td>
</tr>
<tr>
<td>8.17</td>
<td>TerraSAR-X and Radarsat 2 OWF Comparison</td>
<td>134</td>
</tr>
<tr>
<td>9.1</td>
<td>IMB No. 2 Air Temperature Time Series</td>
<td>139</td>
</tr>
<tr>
<td>9.2</td>
<td>Optical LIDP Minima Example</td>
<td>140</td>
</tr>
<tr>
<td>9.3</td>
<td>LIDP Textures</td>
<td>141</td>
</tr>
<tr>
<td>9.4</td>
<td>LIDP SVM Melt Pond Corrections</td>
<td>142</td>
</tr>
<tr>
<td>10.1</td>
<td>Melt Pond Statistics</td>
<td>147</td>
</tr>
<tr>
<td>10.2</td>
<td>Wave Buoy Camera Melt Ponds</td>
<td>148</td>
</tr>
<tr>
<td>10.3</td>
<td>Main Effects of Average Pond Area Predictors</td>
<td>151</td>
</tr>
<tr>
<td>10.4</td>
<td>TSX and LIDP Co-registration</td>
<td>152</td>
</tr>
<tr>
<td>10.5</td>
<td>Initial TSX and LIDP Melt Pond Classification</td>
<td>152</td>
</tr>
<tr>
<td>10.6</td>
<td>TSX Shape Parameter Progression</td>
<td>154</td>
</tr>
<tr>
<td>10.7</td>
<td>Initial TSX and LIDP Melt Pond Classification</td>
<td>155</td>
</tr>
</tbody>
</table>
List of Tables

2.1  Surface heat balance variables ........................................ 16

4.1  Types of SAR bands used and their respective wavelength and frequencies 49

7.1  Confusion matrix of best Neural Network ‘ks’ ....................... 109
7.2  Confusion matrix of best SVM ‘mks’ ................................. 110
7.3  Confusion matrix of best Naive Bayes ‘ms’ ......................... 110
7.4  Confusion matrix of best K-NN ‘s’ .................................... 110
7.5  Confusion matrix of best Discriminant Analysis ‘s’ ................. 110
7.6  Testing Dataset Performance: Against Optical Images Only ...... 111
7.7  Training Dataset Performance: Against Optical and Model Metrics 112
7.8  GLCMs from SAR .......................................................... 115

8.1  ANOVA Test Results for Average Image Intensity .................. 120

10.1 Average Shape Parameters and Statistics from Melt Ponds ........ 146
10.2 Coefficients of linear melt pond model to predict average pond area 150
CHAPTER 1

Introduction

The polar Arctic region, situated above 66°34’N, is part of Earth’s cryosphere, and comprises of large land and ocean masses of snow and ice. Its total area is 1.45 × 10^7 km², almost equivalent in size to Antarctica, and is shown in Figure 1.1. The continental snow cover and sea ice change seasonally, leading to large variations in intra-annual and sometimes, inter-annual energy budgets of both the continental regions and upper mixed layer of the ocean. In addition to seasonal variations, large changes may occur over much longer periods of time in this region. At high latitudes, snow and ice act as insulators for the underlying land and waters, preventing them from losing heat to the atmosphere, due to the high reflectivity of snow and ice in the visible part of the electromagnetic spectrum, and the low thermal diffusivity of sea ice (Peixoto and Oort, 1992). This makes the frozen Arctic act as the Northern Hemisphere’s ‘air conditioner’, playing a central role in regulating Earth’s climate. Recently, however, it was announced that for global temperatures, 2015 was the warmest year since reliable records began in 1880, with 2014 taking second place (https://www.ncdc.noaa.gov). Furthermore, the ten warmest years have occurred since 1998 (http://www.climatecentral.org).
With Arctic sea ice experiencing a dramatic decline over the recent years, numerous studies have reported the retreat of summer ice cover (Cavalieri et al., 1997, Chapman and Walsh, 1993, Comiso and Parkinson, 2004, Comiso et al., 2008, Serreze et al., 2007). Ice thinning has also been reported (Giles et al., 2008, Haas et al., 2008, Kwok et al., 2009, Rothrock et al., 2008, Rothrock and Zhang, 2005). Extended sea ice melt in response to increasing atmospheric temperatures has been theorized as a primary driver of reduced summer sea ice (Markus et al., 2009). The total amount of solar energy absorbed during the summer melt season strongly correlates with melt-onset; earlier onsets of melt allow for earlier development of open water areas, which in turn, enhance the ice-albedo feedback (Perovich et al., 2007).

The Beaufort Sea, shown in figure 1.2, along with areas in the Canada basin north of Alaska and Yukon have experienced some of the fastest decline and greatest ice loss in Arctic summer sea ice (Shimada et al., 2006). The Beaufort Sea houses the Beaufort Gyre, an
anti-cyclonic wind-driven oceanic circulation system that that historically generated some of the thickest and oldest ice in the Arctic Ocean. However, new evidence shows an accelerated decay of old ice in this region, possibly amplified by waves and swell (Asplin et al., 2012, Barber et al., 2009), causing complete ice melt instead of surviving the summer.

The contribution of basin-wide ice decrease, summer ice retreat and opening of the Beaufort Sea are creating conditions that favor the formation of a marginal ice zone (MIZ) (Lee et al., 2012). The MIZ is the zone at the edge of the ice pack whose width is the lateral distance over which the penetration of waves can fracture sea ice, effectively changing the morphology of the floes (Weeks, 2010). Figure 1.3 portrays the fundamental air-ice-upper-ocean processes in the MIZ. The seasonal evolution of the MIZ combines interactions between sea ice, oceanographic and atmospheric processes.
Winds, waves and storms passing through the MIZ shape the size of floes through mechanical flexing and fracture (Wadhams et al., 1988). This process enhances thermodynamic melt through increased solar absorption in newly formed open water, melt of broken ice and brash, and dynamic forcing (Steer et al., 2008, Toyota et al., 2006). Figure 1.4 shows how fast the MIZ develops; during the second week of June, 2012, MODIS, a spectroradiometer, captures pack-like sea ice within close range of the Arctic coastlines in the U.S and Canada. A month later, during June 16, 2012, the MIZ can be seen, where ice retreat is evident and rounded-floes have developed.

Increasing open water allows for direct momentum transfer from the atmosphere to the ocean producing wind-driven waves. Smaller floes of fragmented ice are significantly more mobile, which in turn, can absorb more atmospheric surface stress through deformation and transfer it to the ocean surface; this may induce intense secondary circulations that drive
rapid vertical exchange, entraining heat stored below the mixed layer, increasing basal melting of the ice (Lee et al., 2012).

The regions of open water within and south of the MIZ allow for radiative upper-ocean warming, and through lateral advection, heat can be distributed through the MIZ and further accelerate ice melt, thus creating a positive feedback loop. As summer advances the atmosphere becomes warmer, melting the snow cover on the sea ice. Thus, allowing short wave radiation to more easily penetrate through the remaining snow and ice surface (as well as open water regions between the floes), and warming the water under the floes, fueling the ice-albedo feedback. Therefore, investigating the characteristics of the MIZ evolution is a key component in understanding the heat budget and geophysical dynamics in the Arctic.

Because the Arctic MIZ is such a sensitive area, experiencing extreme changes during the last century (Lee et al., 2012), remote sensing this region is critical. The difficulty in deploying in-situ instruments make passive and active remote sensing measurements attractive for measuring open water fraction, melt and ship navigation. Remote sensing images aid in the monitoring of mechanical and thermodynamical processes in the MIZ. Use of remotely sensing this region decreases field-work costs and reduces the possible
dangers in this remote and hazardous place. However, ground-truth instruments are still
needed to validate remote sensing algorithms, especially since in-situ instruments provide
more temporal coverage in comparison to satellites.

During the last three decades, numerous studies have explored the use of synthetic
aperture radar (SAR) images, which is an active sensor, to monitor the state of sea ice (Kim
Although these studies have made significant contributions to the field, it is widely known
that classifying SAR images is a difficult process; the SAR signals are different when in
freezing vs. melting regimes, and high-wind conditions can also alter open water signals
significantly.

In 2009, thousands of optical literal image derived products (LIDPs) from a government-
classified sensor were released by the United States government, giving researchers a
chance to explore extremely high resolution panchromatic images in the Beaufort Sea.
This has opened an opportunity to study small scale features requiring high resolution im-
ages, such as melt ponds, cracks and leads. The LIDPs also have their own challenges to
overcome; numerous image scenes are contaminated with cloud cover, radiometric incon-
sistency exists between images, and nearly-identical signals between deep melt ponds and
open water pose great difficulty in separating the two.

The overwhelming amount of satellite images available has led to an era of “Big Data”.
Image-by-image thresholding by human operators require customized tuning for each use
case and specialized knowledge to produce results with acceptable accuracy, leading to
inefficiency (Sopharak et al., 2010). Machine learning classification methods can aid in the
Big Data problem, since these algorithms have the capacity to learn patterns and cluster
data, even with highly variable features (Duda et al., 2012). In this work, we make use
of machine learning algorithms to classify satellite images from SAR and LIDPs, taken in
the Arctic MIZ, specifically in the Beaufort Sea region (see Figure 1.2). The classification
of images will include categories of open water and ice from SAR, and open water, ice
and melt ponds from LIDPs. The highlight of this research centers on fast and accurate
detection of open water and ice, catered to methods for ship navigation, applications on
large image datasets, and near real-time ice forecast modeling. The information derived
from these classification results are investigated against in-situ data from ice mass-balance
buoys, autonomous weather stations and buoy cameras which were deployed by the MIZ
Program (Lee et al., 2012), during Spring of 2014. The pairing from classified remote
sensing products along with in-situ data will aid in understanding linkages between open
water fraction and melt in the MIZ seasonal evolution.

The breakdown of the dissertation is as follows: Chapters 2 and 3 discuss the basic
physical properties of sea ice and melt ponds. Here, we elaborate on how sea ice forms
and ultimately, how it melts. We also touch on different factors that affect sea ice growth
and melt, linking how these play a role in the Arctic climate system overall. We include a
1-dimensional sea ice growth model and discuss the outputs when specific parameters are
added and changed.

We then move to a more technical aspect, and discuss the basics of remote sensing,
with applications to sea ice in chapter 4. Here, we explain the differences between active
and passive remote sensing methods by presenting basic concepts. First, we introduce the
use of SAR and how valuable it is for measuring sea ice parameters in the Arctic. Second,
we explain the use of optical images and how we can exploit this type of sensor, although
it is not commonly used for sea ice monitoring due to extensive cloud coverage. The data
used in this study are mentioned in chapter 5. Here, we describe the details of various
satellite sensors used in our study and what technical specifics were utilized. We also introduce three in-situ sensors used in-tandem with remote sensing, including ice mass-balance buoys, autonomous weather stations and web cameras mounted on wave buoys.

We introduce basic concepts of machine learning classification algorithms in chapter 6 including three parametric classifiers: Neural Networks, Support Vector Machines and Naive Bayes. Two non-parametric classifiers, K-nearest neighbor and discriminant analysis schemes are also discussed. We include basic mathematical derivations as well as visual concepts to understand the process behind these supervised learning methods.

In chapter 7, we use the classification algorithms discussed in chapter 6 to create binary images from SAR. The classified images are separated into two classes (1) ice and (2) open water. We validate our work using high resolution LIDP optical images, and compare the best methods by investigating a combination of machine learning metrics. Chapter 8 discusses the application of the best classification method to obtain open water fraction on two large SAR image datasets from TerraSAR-X and Radarsat-2. We analyze the key differences and similarities between the two datasets and discuss possible correlations with physical data from in-situ measurements from the MIZ program.

We then move on to classify the LIDP products in chapter 9, to quantify the evolution of melt ponds in several spatial arrays along the ice pack of the MIZ, using linear support vector machines. In chapter 10, we explore several statistical and spatial parameters of melt ponds such as size, perimeter, and number density, from the melt pond classification outputs and compare our results with previous research. Finally, we wrap our findings in a summary in chapter 11, and discuss the implications we discovered and suggest possible applications from our work and what remains to be done in future studies.
CHAPTER 2

Sea Ice

Sea ice is a solid layer of frozen sea water that forms when the surface of the ocean reaches its salinity derived freezing points, generally, but not exclusively, this happens in our planet’s polar regions. This frozen water serves as a boundary between the ocean and the atmosphere, both of which are also fluid bodies. The freezing of sea ice is salinity-determined, and once freezing begins, brine is rejected from the ice into the ocean, with some brine left behind as impurities, located at the boundaries of pure ice crystals. Seawater contains on average, 35‰ salt (parts per thousand by weight), however, near-surface salinities in the polar oceans are typically less than this, approximately in the 31-34‰ range (Weeks, 2010).

The salt present in seawater decreases its freezing point. In order for water to freeze, it has to lose enough kinetic energy and slow down so that its molecules are able to get close enough and aggregate into a solid. The solute (salt) will get in the "way" of this aggregation process, so the solvent (water) must slow down even further by losing more and more kinetic energy due to even lower temperatures. Figure 2.1 shows the relationship at the freezing point between salinity and temperature of seawater.
After seawater freezes, winds drive the ice crystals, which are termed frazil, into packs that can take many different forms. One example is grease ice, where it is an agglomeration of frazil ice congealed into a "soupy" layer (figure 2.2) having an area of low reflectivity, due to its matte-like appearance (Weeks and Ackley, 1986).

Frazil crystals have a tendency to clump together; if they remain on the surface long enough, they coalesce to form pans, which also grind against one another, becoming circular and ultimately forming pancake ice (Martin and Kauffman, 1981). Pancake ice is predominantly circular, with pieces ranging from 0.3 m to 3 m in diameter and up to 0.1 m in thickness with raised rims (Carsey, 1992).

After the ice has exceeds a thickness of 0.3 m, it is then known as first-year ice. First-year ice is usually not more than one winter’s growth, with sharp angular edges and a rough surface (Weeks and Ackley, 1986).
2.1 Ice Growth

We can examine a simple physical model for ice growth, developed by Maykut (1986). In this model, the general physical setup is shown by figure 2.4, with ice thickness $H$ (m), density $\rho_i$ (kg m$^{-3}$), and thermal conductivity $k_i$ (W m$^{-1}$ K$^{-1}$) resting on seawater at its freezing point $T_f$ ($^\circ$C). Here, the ice-ocean interface temperature is fixed at $T_f$, and the ice surface temperature $T_0$ is controlled by the surface heat balance (not necessarily equal to the near-surface air temperature $T_a$). If the ice is relatively thin, the assumption is that the temperature gradient within the ice is linear, known from field observations (Wakatsuchi and Ono, 1983). Therefore, the conductive heat flux through the ice is $F_c(z = 0) = F_c(z = H)$ and

$$F_c(H) = k_i \frac{T_0 - T_f}{H} \quad (2.1)$$

The amount of growth or melt at the bottom of the ice ($z = H$) is therefore, determined by the sum of the fluxes of the heat through the ice and the heat flux from the underlying
Figure 2.3: Photograph of brash pieces from broken-up young ice, credit: University of Washington, APL

Figure 2.4: Schematic of a slab of snow-free sea ice with according heat fluxes (Weeks, 2010).

water column \((F_c + F_w)\) (Maykut, 1986). Ice will melt if the total sum of the fluxes is positive, and if the sum is negative, ice growth occurs, releasing latent heat to balance the heat lost by conduction:

\[
- \rho_i L \frac{dH}{dt} = F_c(H) + F_w
\]  

Considering the simplest case, when there is zero flux coming from the ocean and the temperature of the interface of the ice and air are equal, equation 2.2 becomes:
\frac{- \rho_i L \frac{dH}{dt}}{H} = k_i \frac{T_f - T_a}{H} \quad (2.3)

If we assume that \( H(t = 0) = 0 \), the integration of equation 2.3 gives

\[ H^2 = \frac{2k_i}{\rho_i L} \int_0^t (T_f - T_a) \, dt = \frac{2k_i}{\rho_i L} \theta \quad (2.4) \]

This is the simple form of Stefan’s solution (Stefan, 1889), but problems arise when \( T_0 \) can be appreciably warmer than \( T_a \), when \( H \) is small. Maykut proposed to make an assumption that the rate of heat exchange between the ice and the atmosphere \( F_t \) is proportional to the difference between \( T_0 \) and \( T_a \)

\[ F_t = C_t (T_a - T_0) \quad (2.5) \]

where \( C_t \) can be considered to be an average surface heat transfer coefficient. From this, \( F_c = F_t \), therefore we can write

\[ k_i \frac{T_0 - T_f}{H} = C_t (T_a - T_0) \quad (2.6) \]

If we solve for \( T_0 \) we obtain:

\[ T_0 = \frac{k_i T_f + C_t H T_a}{k_i + C_t H} \quad (2.7) \]

When we substitute this relation into equation 2.1, we get

\[ F_c = \frac{k_i C_t}{k_i + C_t H} (T_f - T_a) \quad (2.8) \]
We assumed that the flux coming from the ocean was zero, and now we say that the conductive heat flux from the ice is greater than zero, and that at time $t=0$, $H=0$, the integration of equation 2.2 results in

$$H^2 + \frac{2k_i}{C_i} H = \frac{2k_i}{\rho_i L} \theta$$  (2.9)

Maykut then decided to add a snow cover of thickness $H$ to correct for this, with thermal conductivity $K_s$, and if the conductive heat flux in the snow equals that of the ice, gives

$$H^2 + \left[ \frac{2k_i}{K_s} H_s + \frac{2k_i}{C_i} \right] H = \frac{2k_i}{\rho_i L} \theta$$  (2.10)

Following all the derivations, we learn that using simple physical analysis allows the development for equations that provide a reasonable fit to field data on the growth of first-year ice. It is important to note that as ice thickens, the $H^2$ term becomes more and more important relative to $H$. In addition, when snow is added to the equations, as ice thickens, the $H$ term relative to the $H^2$ term increases with an increase in the thickness of the snow (Weeks, 2010). This highlights that snow greatly influences the growth of sea ice. This simple sea ice growth model is not representative of what really occurs in nature, however. For example, we are not considering important parameters such as shortwave and longwave radiation, or the upward heat flux from the ocean, nor are we considering that fact that the overlying snow cover changes with time. These will be examined in the next section.

### 2.1.1 Growth of First-Year Ice

From the simple ice growth model mentioned in the previous section, we can see that only including snow thickness, amount of incoming solar radiation, and ocean heat flux
as constant terms would lead to large errors in predictions of sea ice models. The afore-
mentioned parameters tend to vary from year to year and are different for varying field
locations. Therefore, an approach was developed in order to improve parameterizations of
these terms (Maykut, 1978, 1986). In this model, the method begins by considering a thin
layer of young ice with thickness $H$ having zero snow cover. We saw in the last section that
thin ice without snow cover can have an approximated linear relationship. If the thermal
conductivity $k_i$ and the temperature of the ice surface $T_0$ are known (assuming that the sea
water at the lower surface of the ice is fixed at its freezing point) the value of conductive
heat flux through the ice ($F_c$) can be calculated from equation 2.1.

For a more complex situation, when snow is considered, the conductive heat flux from
a combined snow and ice layer is

$$F_c = \frac{k_i k_s}{k_i H_s + k_s H} (T_f - T_0) = \gamma (T_f - T_0)$$  \hspace{1cm} (2.11)

where $k_s$ is the thermal conductivity of the snow cover, $H_s$ is the thickness of the snow
cover, and $\gamma$ is the combined thermal conductance of the ice and snow layers, $T_f$ is the
freezing temperature at the ice-ocean interface is salinity determined, and $T_0$ is the surface
temperature of the snow (Weeks, 2010). Maykut (1986) considered several terms to calcu-
late $T_0$, if it is assumed that this temperature is a resultant of the balance between heat lost
and heat gained at the surface. These terms are presented in table 2.1.

When calculating the surface heat balance, a flux toward the ice or snow is taken as
positive, while a flux away from the surface is negative. If the ice surface temperature is
taken to be always below freezing, since we are considering ice growth, the energy balance
at the ice surface can be expressed as
Table 2.1: Surface heat balance variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_r$</td>
<td>Incoming shortwave radiation</td>
</tr>
<tr>
<td>$\alpha F_r$</td>
<td>Reflected shortwave radiation. $\alpha = \text{ice albedo}$</td>
</tr>
<tr>
<td>$I_0$</td>
<td>Net influx of radiative energy passing into interior of ice</td>
</tr>
<tr>
<td>$F_{L\downarrow}$</td>
<td>Incoming longwave radiation</td>
</tr>
<tr>
<td>$F_{L\uparrow}$</td>
<td>Emitted longwave radiation</td>
</tr>
<tr>
<td>$F_s$</td>
<td>Sensible heat flux</td>
</tr>
<tr>
<td>$F_e$</td>
<td>Latent heat flux</td>
</tr>
<tr>
<td>$F_c$</td>
<td>Conductive heat flux</td>
</tr>
</tbody>
</table>

\[
(1 - \alpha) = F_r - I_0 + F_{L\downarrow} - F_{L\uparrow} + F_s + F_e + F_c = 0 \tag{2.12}
\]

The terms in equation 2.12 need to be evaluated in order to solve for the surface temperature of the ice ($T_0$). We can begin tackling equation 2.12 by investigating the terms on the left hand side first, and then move towards the right-hand side.

### 2.1.2 Shortwave Radiation

Before discussing shortwave radiation, we must first recall that all matter radiates energy according to the Stefan-Boltzmann equation

\[
F = \varepsilon \sigma T^4 \tag{2.13}
\]
where $F$ is the emitted radiation, $\varepsilon$ is the emissivity (generally a material constant), $\sigma$ is the Stefan-Boltzmann constant $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$, and $T$ is the radiative temperature of the material. In more detail, emissivity is the ratio of the actual emission of a material to the blackbody emission at the same temperature. When the emissivity is equal to one (absorbs all incident radiation), the radiating material is thought to be a blackbody. For our interests, the blackbody curves of the sun and earth are shown in figure 2.5. Note that, the y-axis is normalized for each curve.

We can see that at the shorter wavelength, the peak wavelength nears 0.5$\mu$m, which originates from the sun at a temperature of 5780K, commonly referred to shortwave radiation ($F_r$). Earth has a peak wavelength at 12$\mu$m, with a temperature of 255K and is associated to longwave radiation ($F_L$).

The amount of solar radiation incident at the top of the atmosphere depends on several factors including latitude, season and time of the day. The shortwave radiation reflected back to space depends on the solar zenith angle and properties of the local surface and atmosphere (Hartmann, 1993). The solar zenith angle, $\theta_s$, is the angle between the local
normal to Earth’s surface and a line between that point on Earth’s surface and the sun. The solar zenith angle depends on latitude, season and time of day.

Because it is often difficult to measure the shortwave radiation in the Arctic, due to difficulties in reaching sites of interest, several relationships have been derived to provide reasonable estimates. For clear sky conditions Shine (1984), calculated the downwelling shortwave radiation at the surface as

$$F_{clr} = \frac{S_0 \cos^2(Z)}{\cos(Z) + (1 + \cos(Z))e_a \times 10^3 + 0.046}$$

(2.14)

where $S_0$ is the solar constant, $e_a$ is the partial pressure of water vapor, and $Z$ is the solar zenith angle. Shine also calculated the downwelling shortwave radiation for cloudy conditions

$$F_{clr} = \frac{(53.5 + 1274.5 \cos(Z)) \cos^{0.5}Z}{[1 + 0.139(1 - 0.9345 \alpha) \tau]}$$

(2.15)

In this equation, $\alpha$ and $\tau$ are the albedo of the surface and the optical depth of the clouds (Shine, 1984). Shine combined equations 2.14 and 2.15 to produce a relationship for estimating $F_r$ for all conditions

$$F_{r(all)} = [(1 - C)F_{clr} = CF_{cld}]$$

(2.16)

### 2.1.3 Albedo

The albedo is the fraction of solar energy that has been reflected from earth’s surface relative to incoming solar energy. It is a unit-less quantity which ranges from 0-1. Commonly, albedo refers to the ”whiteness” of a surface where a perfect blackbody would have
a value of 0, and a perfect white reflector would have a value of 1. Sea ice and snow have a much higher albedo than other surfaces on Earth. Examples of typical albedo values for different ice types are shown in figure 2.6. The highest albedo values are obtained from freshly fallen snow (0.81-0.87), while features with lower albedo are associated with the presence of meltwater.

The exact values for $\alpha$ are dependent on the frequency of the incoming radiation, however for typical ice growth problems the wavelength-integrated (total albedo) is frequently used

$$\alpha_t = \frac{\int \alpha(\lambda) F_r(0, \lambda) d\lambda}{\int F_r(0, \lambda) d\lambda}$$ \hspace{1cm} (2.17)

The annual variability of surface albedo in the polar regions is controlled largely by snow cover, as well as the solar zenith angle in high latitudes (Hartmann, 1993). Before melting occurs, both multi-year and first-year ice remain covered by snow, with multi-year ice generally having deeper snow cover (Perovich and Polashenski, 2012), however studies have shown that first-year ice can also have snow that is optically thick (Sturm et al., 2002), that is, the energy cannot pass through the snow without absorption. Once melting takes place, first-year ice albedo differs greatly from that of multi-year ice. For example, thinner,
bare, first-year ice has a lower bubble fraction in its upper layers, leading to less effective light scattering (Perovich and Polashenski, 2012). Differences in albedo between ice types point to shortwave radiation having higher variability in summer than in winter (Weeks, 2010).

Due to the observed rapid Arctic surface temperature warming during the last few decades, sea ice melt has led to more open water, which in turn reduces in albedo of the region. This in turn causes additional heat uptake and impacting the strength of the ice-albedo feedback, which is a positive feedback mechanism (Perovich et al., 2002a). There are two types of feedback cycles: positive and negative feedbacks. Whatever the climate forcing is (an external driver), positive feedbacks will amplify the original effect, while negative feedbacks diminish these effects. The ice-albedo feedback works in a positive loop: As the albedo decreases due to snow and sea ice melt, more heat is taken up by the system, which then causes further melt, causing the system to further lower its albedo due to snow and ice loss. Consistent sea ice melt has led to a shift from a multi-year to first-year ice cover, leading to significant implications for the heat and mass budget of the ice (Perovich and Polashenski, 2012). Understanding feedbacks in the Arctic surface radiation budget is crucial to our understanding of climate change, especially climate forcing (Curry et al., 1995). This is because feedbacks are amplified in polar regions, since the poles are sensitive to climate variations.

Because sea ice is a translucent material, a portion of the incoming shortwave radiation is transmitted through the surface layer without immediately affecting the temperature and mass changes at the surface (Weeks, 2010). The actual amount of shortwave radiation entering the ice is $I_0$ as noted in equation 2.12, and can be computed as
\[ I_0 = i_0(1 - \alpha)F_r \] (2.18)

In equation 2.18, \( i_0 \) is the actual net shortwave radiation, making \( I_0 \) a small portion of what actually enters the material surface. The reader should be aware that there are certain particulates found in sea ice, which affect its spectral properties, but are beyond the scope of this work.

### 2.1.4 Longwave Radiation

The principal term governing the surface heat exchange over sea ice is the longwave balance \( F_L \equiv F_{L\downarrow} - F_{L\uparrow} \), which determines the differences of incoming vs. outgoing longwave terms. The incoming longwave radiation \( (F_{L\downarrow}) \) from the atmosphere is

\[ F_{L\downarrow} = \varepsilon_a \sigma T_a^4 \] (2.19)

where \( \varepsilon_a \) is the effective emissivity of the atmosphere (determined by the vertical structure of temperature and humidity in the troposphere), \( \sigma \) is the Stefan-Boltzman constant, and \( T_a \) is the radiative temperature of the atmosphere (at some specific height) (Weeks, 2010). The outgoing longwave radiation is similarly calculated as

\[ F_{L\uparrow} = \varepsilon_i \sigma T_i^4 \] (2.20)

Here, \( \varepsilon_i \) and \( T_i \) are the emissivity and temperature of the snow or ice surface in contact with the atmosphere. The net longwave radiation \( (F_L) \) is typically negative, due to the fact that the snow or ice surface is usually colder than the atmosphere. This is especially true
for the winter time, when the Arctic experiences a decrease in cloud cover with extremely cold temperatures and low water vapor content. During the summertime, the longwave loss decreases with an increase in cloud cover, higher content of water vapor in the air and air temperatures nearing that of the ice surface. However, totaling over a year, values for $F_{L\downarrow}$ are generally more than twice the value of incoming shortwave radiation; only during the summer melt period when shortwave radiation values are high is $F_r$ slightly greater than $F_{L\downarrow}$ (Weeks, 2010).

### 2.1.5 Turbulent Heat Exchange and Conductive Heat Flux

In this section we explore the terms $F_s$, $F_e$ and $F_c$ in equation 2.12. The first two, the sensible and latent heat fluxes $F_s$ and $F_e$, are very difficult to measure, therefore, requiring bulk parameterizations (Weeks, 2010):

$$F_s = \rho_a c_{p(a)} C_s u (T_a - T_0) \quad (2.21)$$

and

$$F_e = \rho_a L C_e u (q_a - q_0) \quad (2.22)$$

where $\rho_a$ and $c_{p(a)}$ are the density and specific heat of the air, $T_a$ and $T_0$ are the air temperature (at a reference height) and the surface temperature, $u$ is the wind speed at a reference height, $L$ is the latent heat of vaporization, $q_a$ and $q_0$ are specific humidities (at a reference height and the surface), and $C_s$ and $C_e$ are bulk transfer coefficients (where the values are taken to be 0.000175) (Weeks, 2010). The specific humidity can be expressed in terms of partial pressure ($e$) of water vapor, and the total atmospheric pressure ($P$).
A solution to the differences in specific humidities has been solved by Maykut (1978), following a fifth-order polynomial. Values for $L$ can be computed from

$$L = [2.5 \times 10^6 - 2.274 \times 10^3(T_a - 273.15)]$$

(2.23)

The differences between air-sea and air-ice temperatures determine the stability of the lower atmospheric boundary layer, which controls the efficiency of turbulent ocean-ice-atmospheric coupling. In essence, turbulent changes transfer heat from the ocean to the atmosphere, however ice reduces turbulent exchanges by putting a “lid” on the ocean.

Thermal conductivity refers to the ability of a material to conduct heat. Heat transfer occurs more slowly across materials which have low thermal conductivity than across materials of high thermal conductivity. If we return to the assumption that the temperature profile of the ice is linear, the conductive heat flux through the ice can be calculated by equations 2.1 or 2.11, taking into account the presence of snow. In these equations, the thermal conductivity of the sea ice ($k_i$) and snow ($k_s$) must be known. One equation (Untersteiner, 1961) was developed relating $k_i$ to brine volume, ice temperature and the conductivity of pure ice

$$k_{si} = k_i + \frac{\beta S_{si}}{T}$$

(2.24)

where $k_i$ is the thermal conductivity of pure ice, $S_{si}$ is the salinity of the sea ice, $T$ is the temperature in $^\circ$C, and $\beta = 0.13 \times \text{Wm}^{-1}$. When dealing with ice growth beneath snow cover, we can follow the simplest assumption that the thermal conductivity of snow is approximately five times less than that of ice (Maykut, 1986).
2.1.6 Oceanic Heat Flux

Oceanic heat flux is an important component in the Arctic heat budget. Because the albedo of open water is significantly less than that of sea ice, large percentages of open water allow for solar surface heating to occur, leading to larger oceanic heat fluxes (Bryden and Imawaki, 2001). A study demonstrated in a one-dimensional ice model that the sea ice in the Arctic basin would completely disappear if its upward oceanic heat flux surpassed 5 kcal cm\(^{-2}\) yr\(^{-1}\) (Maykut and Untersteiner, 1969).

In sea ice growth calculations, oceanic heat flux is often regarded as an unpredictable variable, since detailed in-situ observations of ablation at the underside of the ice are lacking, making estimates of \(F_w\) a challenging task (Weeks, 2010). Although there are studies (Ebert and Curry, 1993, Fleming and Semtner, 1991) which have calculated the annual cycle for heat flux at the base of the ice, there is no general agreement on an average value for \(F_w\). Ebert and Curry (1993) carried out model calculations to show that the annual cycle for heat flux at the base of the ice takes a bell-shaped curve, with an annual average of 1.8 W m\(^{-2}\). Fleming and Semter (1991) on the other hand, calculated an annual average more than twice this amount, with a value of 5 W m\(^{-2}\).

The discrepancies in annual averages suggest that there are regional and first-year differences in the Arctic, which must be taken into account in sea ice growth models if estimates are to be in reasonable agreement with one another.

2.2 A Simple Sea Ice 1-Dimensional Growth Model

In this section we will provide different results for a simple one-dimensional sea ice growth model, which we implemented based on Maykut’s simple sea ice growth calcu-
lations. We begin by assuming that we are in the month of January, somewhere in the Beaufort Sea, where the surface temperature remains at $T_s = -10^\circ$C for the next 30 consecutive days. We initialize an ice thickness of 0.1 m, and use equation 2.1 to calculate the heat flux in the ice based on the thickness of the previous day. We then calculate the thickness change that this heat flux can cause within one day, given from balancing the heat flux against the amount of heat removed needed to grow ice

$$F_c = \frac{L \rho \Delta H}{\Delta t}$$  \hspace{1cm} (2.25)

as

$$\Delta H = -\frac{\Delta t}{\rho L F_c}$$  \hspace{1cm} (2.26)

From this simple model, the ice thickness evolution can be captured in figure 2.7. We can see that within 30 days, the ice grew approximately 0.4 m, with a final thickness of 0.54 m, however the ice keeps increasing with increasing time. If we add an oceanic heat flux of lets say, 5 W m$^{-2}$ however, we have to add another term to equation 2.26.

The addition of 5 W m$^{-2}$ oceanic heat flux (figure 2.8) only slightly decreases the total ice growth, 0.02 m to be exact, and is nearly identical to the results without the addition of this flux.

However, when we change the heat flux to 180 W m$^{-2}$ we see a dramatic change in the ice growth evolution, shown in figure 2.9. The total ice growth is only 18% of what would have been if no oceanic heat flux had been added. This suggests that there exists a threshold at which point the ocean influences sea ice growth.
Now let’s assume that we had a certain amount of snow fall on the ice, sometime in December. Our initial ice thickness is 0.1 m, and snow thickness (which does not change) is 0.05 m. In this case, the heat flux in the snow and the ice are the same: \( \frac{k_s(T_s - T_i)}{H_s} = \frac{k_i(T_i - T_{bot}/H_i)} {H_i} \), allowing the removal of the interface temperature \( T_i \) from all of the equations. If we do not add any ocean heat flux, after 30 days the ice achieves a total depth of 0.34 meters, as observed in figure 2.10. A few things worth mentioning from this model are that the change in the surface temperature with sea ice thickness is strongly dependent on the season of the year (figure not shown), with surface temperatures rapidly decreasing with increasing ice thickness when the air is cold (Weeks, 2010). Additionally, the inclusion of a snow cover acts as an insulator, preventing the acceleration of ice growth and inhibiting melt. The presence of snow cover modulates sea ice growth.

Taking the same parameters, but now including the addition of 5 W m\(^{-2}\) of ocean heat flux, we obtain a similar ice thickness of 0.30 m after 30 days (figure not shown). In order
Figure 2.8: Ice thickness evolution given constant surface temperature of \(-10^\circ\text{C}\), with addition of 5 W/m\(^2\) oceanic heat flux.

to keep the ice thickness constant, an ocean heat flux of approximately 38.5 Wm\(^{-2}\) would be needed, which is shown in figure 2.11. If we were to remove the snow cover, we would need 180 Wm\(^{-2}\) of ocean heat flux to maintain a constant sea ice thickness. This again, highlights the importance of snow cover as an insulator in the Arctic.

In order to introduce more realistic results to changes in ice thickness, we must take into account the variation in surface temperature throughout the year by using a simplified version of the Semtner zero-layer model (Semtner, 1984). The addition of shortwave and longwave radiation must also be introduced along with latent and sensible heat fluxes. We start by taking data from a one-dimensional thermodynamic model of sea ice (Maykut and Untersteiner, 1971), assuming that the temperature gradient in the ice is linear. We obtain an expression for the surface temperature \(T_{\text{top}}\) of the ice from balances between incoming
Figure 2.9: Ice thickness evolution given constant surface temperature of $-10^\circ$C, with addition of 180 W m$^{-2}$ oceanic heat flux.

atmospheric fluxes and outgoing longwave flux (all upward fluxes being positive) along with heat fluxes in the ice

$$- (1 - \alpha)F_{SW} - F_{LW} + \epsilon \sigma T_{top}^4 = - k \frac{T_{top} - T_{bot}}{H}$$

(2.27)

where $\epsilon = 0.95$. We then assume that $H = 0.5$ m is the sea ice thickness, with a bottom temperature of the ice remaining constant at its freezing point ($T_{bot} = -1.8$ $^\circ$C) having salinity 34 g/kg and no snow cover. Obtaining a solution for the surface temperature from equation 2.27 can be achieved by transforming it into a 4-th order polynomial

$$aT_{top}^4 + bT_{top}^3 + cT_{top}^2 + dT_{top} + e = 0$$

(2.28)
Once the real roots are computed, we can introduce in the surface temperature into equation 2.1 adding no ocean heat flux, and then solving for the thickness change using equation 2.26. In this simplistic model, the surface temperature is set to 0°C, whenever it is calculated to be warmer than this to account for no surface melt. Figure 2.12 shows the evolution of the change in ice thickness for an entire year, given the changing surface temperature with no initial snow cover and no surface melt.

From day 0, sea ice starts to grow, until it slightly plateaus during early May. We then see a slight drop during mid-May, because at this point, the heat flux of the ice becomes positive. Going back to the simple calculation of ice thickness (equation 2.26), if the heat flux of the ice is positive, then the thickness of the ice becomes a negative value. If we include surface melt, we can attribute a loss of ice growth to a physical parameter. We can calculate the energy available for melting in cases where the surface temperature went
above the melting point and add this value to our new ice growth equation. However, we must also keep the albedo constant at $\alpha = 0.8$, otherwise all the ice melts. The results of this model are shown in figure 2.13.
Figure 2.12: Ice thickness evolution without surface melting and changing surface temperature.

Figure 2.13: Ice thickness evolution with surface melt and changing surface temperature.
Sea ice thickness peaks near late spring, and then starts to decrease during the early summer months. The minimum ice thickness occurs in September, after the combination of energy released from the ice melt along with a surface temperature of 0°C. During late fall, the sea ice thickness picks up again, this time with a slightly higher thickness than the year before.

### 2.3 The Role of Sea Ice and Clouds in the Arctic Radiation Budget

The radiative properties of clouds are similar to those of ice, since they both have a high albedo value and high longwave emissivity. Clouds are more complex than ice in terms of their primary effect on the surface radiation budget, as clouds have a dual effect. Clouds can reflect more solar radiation back to space, typically 100 Wm$^{-2}$, causing a decrease to incoming shortwave radiation at the surface, thus promoting a cooling effect. However, this is very variable, and depends strongly on solar zenith angle. At the same time, sea ice retreat increases the absorbed solar radiative flux. Additionally, clouds can trap longwave radiation emitted by the Earth, emitting their own energy back to the surface, thus causing an increase to the incoming longwave radiation. The increased incoming longwave radiation (approximately 40 Wm$^{-2}$) would increase the surface temperature, until a balance is reached between outgoing longwave radiation to space and incoming absorbed shortwave radiation.

The process between enhancing or inhibiting the initial surface warming signal is known as the cloud-radiation feedback. The decrease in surface albedo due to sea ice retreat is partly compensated, however, not canceled by stronger shortwave cloud cooling, which depends on solar zenith angle, cloud amount, and cloud type. Albedo, surface tempera-
ture, and clouds have a complex relationship in affecting the Arctic and its ocean surface radiative flux.

During the winter months, the atmosphere is stable due to the absence of solar radiation and a larger sea ice extent, enabling formation of stratus clouds (Herman and Goody, 1976). In an unstable environment, convective clouds are formed due to upward fluxes of moisture and heat. The cold air is advected over a warmer surface, lifting moist and warm air through turbulence and forming low clouds. Research has shown that low clouds respond to variations in Arctic sea ice extent (Barton and Veron, 2012, Kay et al., 2011, Kay and Gettelman, 2009, Palm et al., 2010, Peixoto and Oort, 1992). In situations where strong turbulent surface-ocean-atmospheric coupling can occur, reduced sea ice causes more frequent low clouds with higher liquid water content (Taylor et al., 2015). This phenomena could possibly lead to a positive feedback. Barber and Yackel (2010) reported that the increase in the downward longwave flux from low clouds was sufficient to increase the basal layer snow and ice interface temperatures (Barber and Yackel, 1999).

2.4 The Marginal Ice Zone

The marginal ice zone (MIZ) is part of the seasonal ice zone, varying in width of a few tens to hundreds of kilometers, and extends from the ice edge into the ice pack. The actual width of the MIZ is defined by ice that has very high variability of ice drift, smaller and thinner ice floes, and an abundant amount of open water (Kwok, 2014). The MIZ is heavily influenced by the open water, air and temperature changes. A transition zone between the open ocean and the sea ice, the MIZ is very dynamic and variant. It can consist of small ice floes or large pack ice, with changes in its extent taking place over hours or days.
Another definition of the MIZ can be taken as the zone at the edge of the ice pack, where the penetration of waves can fracture the ice, ever changing its morphology (Weeks, 2010).

Because waves can penetrate into the ice pack, they can cause break up of large floes, reducing their size, often in a matter of hours (Lee et al., 2012). The size of these smaller floes makes them susceptible to enhanced melting and dynamic forcing, leading to further ice decay and creating a positive feedback mechanism (Lee et al., 2012). The decrease in volume of sea ice causes retreat, exposing a large area of open water which then opens opportunity for conditions to generate more swell. Revisiting figure 1.3, we can see that the ice cover shields the shortwave radiation from penetrating, and also restricts the upper ocean from direct contact with surface wind.

Understanding the air-ice-ocean-wave dynamics in the MIZ is crucial to improve future predictions of ice edge location, floe size and concentration. This is especially significant in today’s climate, since Arctic sea ice has already experienced melt from other factors such as melt ponds, air temperature, and clouds to name a few (Weeks, 2010).

2.5 Chapter 2 Summary

In this chapter, we introduced important parameters that enhance and limit sea ice growth. We also demonstrated varying scenarios of sea ice growth by constructing a 1-dimensional sea ice model. We showed that the addition of ocean heat flux has significant effects in affecting ice growth when fluxes are large enough. In addition, we learned that implementing an initial snow cover inhibits both sea ice growth and ice melt. The role of sea ice and clouds in the Arctic radiation budget was also briefly discussed. In the next
chapter, we switch gears from ice growth to ice melt, focusing on an important variable which facilitates sea ice melt: Melt ponds.
CHAPTER 3

Melt Ponds

Throughout its annual cycle, Arctic sea ice goes through a number of changes which affect the ice cover properties, with one of the most recognizable changes being the formation of melt ponds (Barber and Yackel, 1999). From chapter 2, we learned that albedo values ranged between ice types (see figure 2.6). For first-year ice, melt pond albedo is typically lower than in multi-year ice, however, an examination of melt ponds on these two types demonstrated that the surface albedo is largely determined by the percentage cover of snow surface types (Barber and Yackel, 1999). For example, the aging of snow results in an albedo decrease in all wavelengths, since the increase in grain size and rounding of the grains reduce the effective volume scattering (Grenfell and Perovich, 1984). In addition, although melt ponds initially develop during the summer, studies have shown ponds to form earlier in the season due to presence of thin snow cover (Webster et al., 2015).

The albedo of sea ice cover depends on a quantity of parameters: the depth and state of its snow cover, the optical properties of the snow and ice, the properties and distribution of melt ponds, and the amount of open water (Carsey, 1985, Curry et al., 1995, Grenfell and Perovich, 2004, Zubov, 1945). Figure 3.1 shows a photograph taken during a sea ice field
campaign, where melt ponds of varying depths are shown: deeper ponds have a darker hue, while shallower ponds appear much lighter in color.

Spatial coverage of melt-ponds is a predominant control of albedo, with ponds reported to cover up to 50 to 60% of the sea ice area (Eicken et al., 2004, Fetterer and Untersteiner, 1998). The albedo of melt ponds is much lower than that of snow or ice: 0.7 for partially frozen ponds, 0.33 for mature ponds, and 0.19 for old ponds (for values in the visible spectrum) (Morassutti and LeDrew, 1996, Perovich and Polashenski, 2012). Decrease in surface albedo due to the presence of melt ponds cause a larger amount of shortwave radiation flux to penetrate the ice and water lying beneath it, leading to further melting and a further reduction in albedo; this ice albedo-feedback is thought to be one of the reasons for rapid decline of summer Arctic sea ice (Hall, 2004, Pinker et al., 2014).

In order to understand the ice-albedo feedback, physical properties that govern the state of the ice cover in response to forcing from the atmosphere and the ocean are critical; the evolution of pond fraction and lead fraction are primarily responsible for changes in

Melt pond evolution also plays a role in influencing sea ice extent; there is a strong negative correlation between spring pond fraction and September sea-ice extent (Liu et al., 2015, Schröder et al., 2014), where more spring and summer melt ponds result in less sea ice the following September. Schröder et al., (2014) reported that melt pond fraction in May had the strongest impact in sea ice state in the subsequent September, highlighting that knowledge of melt pond timing is critical. Therefore, understanding both the spatial and temporal evolution of melt ponds, and their physical characteristics, will further improve future climate model predictions.

### 3.1 Stages of Pond Evolution

Many factors influence the development of ponds, however, pond evolution can be characterized by four stages which exhibit unique pond behavior and control mechanisms (Eicken et al., 2002, Polashenski et al., 2012). In this section, we summarize these stages based on previous studies and provide illustrations and figures in order to aid in the conceptualization of melt pond behavior.

#### 3.1.1 Stage 1

The first stage begins with the onset of pond formation: during the summer melt season, atmospheric heat flux is the primary driver of sea-ice surface-melt, although basal melt does occur from ocean heat fluxes. Once the snow and near-surface air obtain values consistently above the freezing point, snow begins to melt and liquid water becomes available within the
surface of the ice pack. It is during the early part of the first stage where melt pond fractions are at their lowest (along with pond freeze-up, as will be shown). For example, Perovich et al. (2002b) reported pond fractions below 5% during May until the second week of June, however, by mid-June, pond fractions increased to 20%.

As snow melts, melt water collects on top of relatively impermeable ice, and flows to the lowest points of local sea ice topography, leading to an increase in pond coverage, well above sea level. The melt pond volume is controlled by melt rate, where a balance between inflow and outflow is achieved. The amount of meltwater flowing into a pond is determined by melt rate, precipitation and the size of the pond’s catchment basin, while outflow rates depend on hydraulic head and presence of outflow pathways (Polashenski et al., 2012). As seen from Figure 3.2, outflow can occur in two ways: water can percolate vertically down through the porous ice, or it can move horizontally across the surface of the ice, ultimately exiting through leads, bore holes, and cracks.
How meltwater is distributed on the ice surface is governed by topographic relief, whose forces are dictated by what occurred prior to the onset of melt such as deformation, snow drifting, and presence of hummocks (Polashenski et al., 2012). For example, lateral spread in pond area is limited by the presence of hummocks on multi-year ice (Eicken et al., 2004, Fetterer and Untersteiner, 1998, Perovich et al., 2002a,b), where ponds tend to be confined to deeper pools with less spatial coverage. Melt ponds on multi-year ice are also larger, irregular and usually interconnected (Carsey, 1992). This gives the multi-year ice the ability to have a well defined drainage network, an imprint from the previous year. On the other hand, first-year ice’s low surface relief allows meltwater to spread laterally, creating larger but shallower melt pond fractional areas (Eicken et al., 2004, Polashenski et al., 2012, Yackel and Barber, 2000). Melt ponds on first-year ice produce a regular pattern of extensive small puddles and ponds, with little connectivity (Johnston and Timco, 2008).

Pond coverage has been found to vary on different ice types, as much as 0% to 80% on undeformed first-year ice and 0% to 40% in multi-year ice (Nicolaus et al., 2012). Despite both ice types undergoing the same air temperature forcings, the different behaviors of their melt ponds come back to the effects of their contrasting ice topographies, where surface roughness influences the pace of melt; discrete ponds on low-elevation areas of multi-year ice facilitates a melt evolution that is steady throughout the melt season (Webster et al., 2015). The opposite is true for first-year ice; once melt begins, it quickly intensifies and exceeds the pace on multi-year ice, due to the smooth topography of first-year ice.

During the late part of stage 1, pond fraction peaks, since the first stage results in very low albedo and coincides with near peak solar input in much of the Arctic (Polashenski et al., 2012), along with horizontal spread.
3.1.2 Stage 2

During the second stage, meltwater percolates through the ice (Eicken et al., 2002), and continues to flow through horizontal transport to macroscopic flaws (Scharien and Yackel, 2005), where macroscopic flaws are driven by ice pore permeability (Freitag and Eicken, 2003). Figure 3.3 is a composite of photos showing the evolution of brine channels to macroscopic drainage holes at a study field site. Permeability transitions occur as ice temperature increases, and has been reported as a meltwater balance control mechanism, capable of controlling pond drainage during certain stages in the melt season (Eicken et al., 2002, Golden, 2001).

During this stage, melt ponds that were once well above sea level drop to very near sea level, due to increased outflow, and pond coverage decreases significantly on first-year ice, compared to multi-year ice (Polashenski et al., 2012).
3.1.3 Stage 3

In the third stage, melt ponds remain near sea level, since outflow pathways are no longer limiting and ice permeability is high (Polashenski et al., 2012). In addition, late in the summer, as shortwave radiation penetrates the ponds and into underlying ice, the brine volume in the ice becomes large, allowing for the pond to drain by vertical seepage to the ocean and thus causing “flushing” of the ice (Untersteiner, 1968). During the third stage, changes in topographic relief play a major role in determining pond coverage: new pond areas are created where the local surface height is below the freeboard (Polashenski et al., 2012).

3.1.4 Stage 4

At the fourth stage, ponds experience refreezing, however ponds may freeze whenever atmospheric forcing allows. In autumn, melt ponds will refreeze at the upper surface and form an ice lid. This lid insulates the pond from the atmosphere, trapping pond water between the sea ice (if any is left) and its frozen lid. The frozen pond’s albedo will increase, sealing the pond both thermodynamically and optically; the water trapped inside the pond is a store of latent heat, which is released during refreezing, thus inhibiting ice growth at the base of the sea ice (Flocco et al., 2015).

3.2 Chapter 3 Summary

The presence of melt ponds on the surface of sea ice greatly reduces its albedo, introducing a positive feedback mechanism, called the ice-albedo feedback. The presence of melt ponds enhance summer melt, and area correlated to a lower September sea ice extent.
Several factors influence the development of melt ponds, however they can be characterized by four overall stages. Each stage is attributed to different physical characteristics, where ice surface topography and meltwater balance play key roles. We now end our physical background of sea ice and melt ponds, and move to a more technical aspect. In the next chapter, we introduce two types of remote sensing techniques which we use in our work to study sea ice and their limitations.
CHAPTER 4

Remote Sensing of Sea Ice

Remote sensing is the act of collecting data by detecting energy reflected or emitted from earth, at some distance. For sensors on satellites, their frequency range must be chosen wisely, since there is some attenuation of the signal as it travels through the atmosphere; the choice of frequency bands are limited. The part of the electromagnetic (EM) spectrum relevant to remote sensing is shown in figure 4.1, however for our work, we focus on the visible and microwave part of the EM spectrum.

Remote sensing instruments encompass two primary types-active and passive. Active sensors provide their own source of energy in order to illuminate targets. Active sensors emit radiation in the direction of the target, and once the radiation has interacted with the surface, the sensor measures the radiation reflected or backscattered from the target (nas, 2017). Active-satellite sensors include altimeters, lidars, radars, SAR, scatterometers, and altimeters. Passive sensors do not emit their own energy, instead, they detect radiation emitted or reflected by the target/area. The most common passive sensors used in satellite remote sensing include hyperspectral and imaging radiometers, and spectroradiometers.

Because the Arctic is dark for parts of the year, and cloud cover is predominant, active and passive microwave sensors have been crucial in monitoring the changes taking
place (Onstott and Shuchman, 2004). For active satellites or sensors with longer wavelengths, clouds and precipitation have little to no effect on the EM propagation through the atmosphere. This enables the study of sea ice regardless of season, and weather-related phenomena, giving an opportunity to measure annual ice concentration and dynamics, to name a few. Studies have also taken advantage of optical sensors to study sea ice in the Arctic (Miao et al., 2015, Perovich et al., 2002b, Perovich, 1996, 2007, Webster et al., 2015), contributing to knowledge on melt ponds, lead development, and the evolution of albedo on varying sea ice types. In this work, we focus in the visible and microwave portions of the EM spectrum in order to study the sea ice in the Beaufort MIZ region.

4.1 SAR

Because radar is able to penetrate through cloud cover and does not depend on solar illumination, it is one of the most important instruments for observing sea ice on a variety of scale, from small to regional and larger areas. SAR does not require the illumination of
the sun in order to make its observations, instead, it provides its own source of illumination in the microwave range of the electromagnetic spectrum, at wavelengths between $0.0075 - 1\, \text{m}$. This enables the radar to operate independent of solar illumination, and the long waves can penetrate clouds and precipitation, which is extremely useful in polar regions, where dense cloud and darkness can limit remotely sensed observations from visible and infra-red sensors.

Four main radar systems used to observe sea ice are (1) SAR (2) side-looking real aperture radar (RAR) (3) scatterometers and (4) altimeters (which measure sea ice freeboard). In this work, we focus on the use of SAR as our imaging radar instrument, since SAR is considered to be an over-all operational choice sensor for monitoring sea ice, due to its high resolution and all-weather working capability (Scheuchl et al., 2004).

The geometry of a satellite SAR system is shown in figure 4.2, where the SAR has the ability to look left or right of the orbit, and whose ground projection is off-\textit{nadir}. Its two main directions are called \textit{azimuth} (along the satellite orbit) and \textit{range} (the look direction of the SAR, which is perpendicular to the azimuth).
The range resolution of a radar system is the minimum distance between two point targets that can be observed as separate bright spots in the image (Ulaby et al., 2014). The azimuth resolution of a RAR (which is much coarser than that of SAR) is proportional to the beam width of the radar

\[ \beta = \frac{\lambda}{L} \]  

(4.1)

where \( \beta \) is a function of signal wavelength \( \lambda \) and antenna length \( L \). The along-track resolution, \( r_a \), is approximated as

\[ r_a = R\beta \]  

(4.2)

where \( R \) is the slant range, which is the distance from the antenna to the midpoint of the swath.

The \( r_a \) is on the order of kilometers for satellite systems. Because the beam width of radar is a function of antenna length, with larger antennas producing a narrower beam, in order to obtain a small \( r_a \) for RAR either a very short range or a very large antenna must be used. SAR however, uses the along-track motion of the satellite in order to ‘synthesize’ a longer antenna by recording the time history of magnitude and phase of the signal received. In order to form an image in the azimuth direction, the complex signals measured by the SAR are summed systematically with a congruous phase shift in order to simulate what antenna would have generated if it were large enough in the azimuth direction to capture the responses simultaneously. The azimuth resolution is then:
\[ r_a = \frac{L}{2} \]  

(4.3)

Where \( L \) is the physical length of the antenna. Therefore, the azimuth resolution depends on the azimuth size of the antenna and increases with range. The slant range resolution \( r_r \) of a SAR is

\[ r_r = \frac{c t_p}{2} \]  

(4.4)

Where \( t_p \) is the time length of the radar pulse, and the factor of 1/2 is due to doubling of phase shifts (i.e., two-path way). The range resolution depends on the bandwidth or pulse duration of the transmitted signal, since \( t_p = 1/B_e \), where \( B_e \) is the bandwidth of the radar.

Generation of SAR images from the input signals is computationally intensive. At its basic coordinate system, a SAR product has equal line spacing corresponding to constant time steps from its platform motion forward, therefore, most processors convert from slant range to ground range for more practical purposes (Johannessen et al., 2006).

The received power \( P_R \) of a SAR is given by the following basic radar equation:

\[ P_R = \frac{P_t}{4\pi R^2} G \frac{\sigma}{4\pi R^2} A \]  

(4.5)

where \( P_t \) is the transmitted power, \( R \) is the distance between the surface of the earth and the antenna, \( G \) is the gain of the antenna, \( \sigma \) is the target radar cross-section, and \( A \) is the antenna area. The total received power from a resolution cell is proportional to the radar cross section (\( \sigma \)), and other terms in the radar equation are generally assumed constant.
(Liang, 2008). The radar cross-section is defined as the area (m²) of a discrete target, however for distributed targets a reflection coefficient is used, called normalized radar cross-section (σ⁰), or backscatter coefficient. This σ⁰ is the radar cross-section per unit surface area, making it a dimensionless number, which measures the trait of the material under the radar’s observation, and is a function of frequency, incidence angle, polarization and the scattering characteristics of the area the radar is illuminating. Typically, σ⁰ is expressed in dB (10 log σ⁰). Values for various sea ice types range from −4 dB to −25 dB (Johannessen et al., 2006). Table 4.1 shows the different wavelengths and frequencies operated of radar bands. The L-K band portion of the electromagnetic spectrum causes a strong scattering cross-section contrast between sea ice and water (Livingstone et al., 1987a).

Table 4.1: Types of SAR bands used and their respective wavelength and frequencies

<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength (cm)</th>
<th>Frequency (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>30-100</td>
<td>0.3-1.0</td>
</tr>
<tr>
<td>L</td>
<td>15-30</td>
<td>1.0-2.0</td>
</tr>
<tr>
<td>S</td>
<td>7.5-15</td>
<td>2.0-4.0</td>
</tr>
<tr>
<td>C</td>
<td>3.75-7.5</td>
<td>4.0-8.0</td>
</tr>
<tr>
<td>X</td>
<td>2.4-3.75</td>
<td>8.0-12.5</td>
</tr>
<tr>
<td>Ku</td>
<td>1.67-2.4</td>
<td>12.5-18.0</td>
</tr>
<tr>
<td>K</td>
<td>1.13-1.67</td>
<td>18.0-26.5</td>
</tr>
<tr>
<td>Ka</td>
<td>0.75-1.13</td>
<td>26.5-40.0</td>
</tr>
</tbody>
</table>

The incidence angle, or the angle to the vertical of the incident radiation, is an important parameter of a SAR. This angle is related to the ground range dimension of the image, namely, every range distance has an incidence angle. Incidence angle may have a large variation across an image, especially for sensors operating at wider swath modes (typically, 20-50°). Images taken at wide swath modes have the benefit of capturing large areas, however the image scene backscatter value is dependent on the incidence angle: areas
closer to nadir achieve a 'brighter' return value than areas at larger incidence angles, as shown in figure 4.3.

Polarization is the direction of the electric field vector of a propagating electromagnetic (EM) wave, where the electric vector is transverse to the direction in which the EM energy is transmitted (Van Zyl, 2011). Sunlight is randomly polarized, since the direction of the electric vector is randomly distributed. When a radar emits a pulse, it is either horizontally, vertically, circularly, and elliptically polarized, or have some combination of these. In the same way, the radar can pick up varying polarization information at the receiver. By manipulating different covariance matrices of varying polarizations, polarimetric features such as the cross-polarization ratio \( R_{VH/VV} \), or co-polarization \( R_{VV/HH} \) ratio to name a few, can be used (where \( H \) and \( V \) are horizontally and vertically polarized). For example, \( R_{VH/VV} \) is known as a measure of depolarization, and is related to multiple scattering within the sea ice volume or to surface roughness (Moen et al., 2013, Scharien et al., 2012). With
this feature, ridges or pancake ice may be discerned, as these have a very rough surface in contrast to other sea ice types.

The $\sigma^0$ of ice depends on radar parameters and sea ice properties. Sea ice has a very complex structure and composition. It is a composite polycrystalline of pure ice with brine, air bubbles and snow layers. Its $\sigma^0$ depends on many physical parameters such as (Onstott and Shuchman, 2004):

1. The surface roughness and properties of the material

2. The dielectric constant of the sea ice (both real and imaginary parts)

3. The dielectric discontinuities or discrete scatterers (e.g., bubbles of air or brine pockets in the ice)

4. The orientation of the ice facets to the radar (with respect to its azimuth angle and incidence angle)

This surface roughness controls how the microwave energy interacts with that surface, dominating the backscatter. For example, a smooth surface will cause a specular reflection of the incident energy, usually away from the sensor, and therefore, only a small amount of energy is returned to the radar. Smooth surfaces will appear as dark-toned in a SAR image. On the other hand, a rough surface will scatter the energy equally (approximately) in all directions, where a significant amount of the energy is backscattered to the radar receiver. The backscatter return signal from first-year, new ice, and open water are all dominated by the surface roughness (Onstott and Shuchman, 2004).

The Rayleigh criterion defines a surface as smooth if:
where \( \lambda \) is the radar wavelength, \( \theta \) is the incident angle of the radar beam and \( h \) is the surface roughness, defined as the root-mean-square height relative to a perfectly smooth surface. Figure 4.4 shows the varying types of reflection scenarios under specific roughness for two radar wavelengths (X and L band).

The dielectric constant is a measure of a material’s ability to reflect, absorb or dissipate energy. The ocean surface has a very high dielectric constant, therefore reflects most of the radar energy at the surface. The dielectric constant of sea ice is dependent on the amount of salt in the ice. Penetration depth describes the distance an electromagnetic wave can
travel through a medium before it is reduced by $1/e$, and determines the extent of volume scattering. The penetration depth is a function of radar frequency, angle of incidence, temperature, and conductivity of the ice or the snow (Onstott and Shuchman, 2004). For example, first year ice having salinity of 7-8 parts per thousand at a temperature of $-5^\circ$C has a penetration depth of 0.02-0.04m at X-band (Vant et al., 1974, 1978).

The strength of the radar’s return signal from sea ice also depends on scale-roughness. For example, smaller broken ice pieces generally give a higher backscatter than those with a level surface. Reflections of multiple ice pieces nearly perpendicular to each other can enhance a return signal by behaving like a corner reflector, increasing the signal by up to an order of magnitude (Johannessen et al., 2006). This is also seen on many SAR images where ridges exist, creating bright linear features, such as those shown in figure 4.5.

The effect of these physical parameters can be divided into either (1) surface scattering and (2) volume scattering. Surface scattering is attributed to scattering near the ice/snow surface, while volume scattering depends on the amount of volume-distributed scatterers
Figure 4.6: Microwave backscatter for multiyear ice, first year ice and smooth open water (Schuchman and Onstott, 2013).

within the upper layers of the ice. Figure 4.6 illustrates the different scattering mechanisms for multiyear ice, first year ice and open water.

Once sea ice begins to freeze, a large amount of brine is rejected from the ice into the ocean, however impurities are still left behind in the process, specifically at the boundary of ice crystals. As the salt migrates away from the ice, brine volume is reduced. First year ice typically contains approximately 6 to 10 per mil salt, which is enough to hinder the penetration of C-band radar waves into the ice to only a few centimeters (Shokr, 1998). Since new and first-year ice contain high amounts of salt, with their backscatter originating from the surface, the radar scattering from this type of ice is highly dominated by surface scattering.

As summer sets in, the snow pack on sea ice begins to melt. Water freely percolates throughout the interior of the ice, enabling the brine to flush downward into the ocean. This process leaves the upper ice layer with many large air bubbles from the expulsion of brine rejection. The air bubbles act as discrete scatterers, having dimensions within an order of magnitude in size of the radar wavelength, thus allowing for volume scattering. Multi-year ice return to the radar is a function of surface and volume scattering; multi-year ice can
have both a rough surface (due to the development of hummocks and ice convergence), as well as low salinity.

### 4.1.1 Speckle Noise

In radio communications, signal fading is a term used to describe the fluctuations in received signals caused by multi-path interference. For radar remote sensing, fading describes random-like intensity variations corresponding to signals which are backscattered from cells at different locations on a distributed target. Because a SAR image is formed by coherently processing returns from successive pulses, this effect causes pixel-to-pixel variation in intensity, leading the variation to manifest itself as a salt-and-pepper like pattern, called speckle. The random variations that produce a speckle pattern on the image are essentially due to phase interference effects: constructive and destructive interference of coherent waves reflected by many elementary scatterers contained within the imaged resolution cell. The salt-and-pepper type appearance is simply a visual manifestation of fading statistics. This implies that using a single-pixel intensity value as a measure of the distributed target’s reflectivity would be inaccurate.

We can characterize the fading statistics associated with a material/surface of uniform electromagnetic properties by modeling the surface as an ensemble of independent, randomly located scatterers, all of comparable scattering strengths (Ulaby et al., 1986). Additionally, we assume that the range distance is much larger than many radar wavelengths, and the surface is rough on the scale of the radar wavelength assumed to have its phase uniformly distributed in the interval of $(-\pi, \pi)$. By the Central Limit theorem, the summation of the real and imaginary components of the vector $(x, y)$ are independently and identically Gaussian, distributed with zero mean and a variance, and $j$ is the imaginary unit:
\[ \sum_{i=1}^{M} (x_i + jy_i) = \sum_{i=1}^{M} x_i + j \sum_{i=1}^{M} y_i = x + jy \]  

(4.7)

From this model, we would find that the amplitude (A) of the backscattered signal is Rayleigh-distributed (and \( \sigma^2 \) here represents the variance)

\[
p_1(A) = \frac{2A}{\sigma^2} \exp \left( -\frac{A^2}{\sigma^2} \right) \tag{4.8} \]

for a single-look image, with its mean

\[
M_1(A) = \sigma \sqrt{\pi}/2 \tag{4.9} \]

and variance

\[
Var_1(A) = (4 - \pi)\sigma^2/4 \tag{4.10} \]

where the subscript 1 indicates the speckle statistics are for a single-look SAR data. Taking the ratio of the standard deviation to the mean, we get \( \sqrt{4/\pi-1} = 0.5227 \), which is independent of \( \sigma \). This is the 'basic' characteristic of multiplicative noise, or an inherent signal-to-noise-ratio for a Rayleigh-fading signal, even in the absence of additive noise (Ulaby et al., 1986). The intensity of an image, defined as \( I = x^2 + y^2 = A^2 \), has a negative exponential distribution with mean

\[
M_1(I) = \sigma^2 \tag{4.11} \]
and variance

\[ \text{Var}_1(I) = \sigma^4 \]  \hspace{1cm} (4.12)

Here, the standard deviation to mean ratio is unity, which signifies that speckle noise would appear more pronounced in an intensity image than in an amplitude image.

Several methods have been applied in order to reduce speckle noise (Ulaby et al., 1986), however we will focus on a specific approach called multi-look processing. This method is accomplished by dividing the synthetic aperture length (azimuthal Doppler frequency spectrum) into \( N \) segments (also known as looks), processing each segment independently to form an intensity image, and subsequently summing the \( N \) images together to form an \( N \)-look SAR image (Lee and Pottier, 2009). Because we assume that the samples in the images are statistically independent, the end product is an average of \( N \) independent samples, which causes a reduction of azimuth resolution (degraded by a factor of \( N \)).

If the summation were carried out on the complex images instead of on amplitudes or intensities, speckle reduction would not be achieved. The process is identical to the vector sum of the total number of elementary scatterers from the \( N \) images, therefore the statistics would remain identical to that of a 1-look SAR image (Lee and Pottier, 2009).

For an \( N \)-look intensity image,

\[ I_N = \frac{1}{N} \sum_{i=1}^{N} I_1(i) = \frac{1}{N} \sum_{i=1}^{N} (x(i)^2 + y(i)^2) \] \hspace{1cm} (4.13)

where \( x(i) \) and \( y(i) \) are the real and imaginary parts of the \( i \)th look, or sample. Following that \( x(i) \) and \( y(i) \) are independently Gaussian distributed, \( N \) \( I_N \) will have a Chi-square distribution with \( 2N \) degrees of freedom, with the PDF of the \( N \)-look intensity as
\[
p_N(I) = \frac{N^N I^{N-1}}{(N-1)!\sigma^{2N}} \exp \left( -NI/\sigma^2 \right), \quad I \geq 0 \tag{4.14}
\]

with mean

\[
M_N(I) = \sigma^2 \tag{4.15}
\]

and variance

\[
\text{Var}_N(I) = \frac{\sigma^4}{N} \tag{4.16}
\]

The standard deviation to mean ratio is reduced by the factor \(1/\sqrt{N}\) of the single-look data. In other words, the \(N\)-look processing reduces the standard deviation of speckle noise by a factor or \(\sqrt{N}\).

In order to illustrate the various distributions mentioned, figure 4.7 shows a NASA/JPL AIRSAR intensity image (Lee and Pottier, 2009), illustrating an exponential distribution, a 1-look amplitude, a Rayleigh distribution, a 4-look amplitude and a chi distribution.
Figure 4.7: SAR speckle statistical distributions. (A) 1-Look intensity, (B) Histogram of (A), (C) 1-Look amplitude, (D) Histogram of (C), (E) 4-Look amplitude, (F) histogram of (E), Chi-square distribution (Lee and Pottier, 2009).
4.1.2 Sea Ice Backscatter During Winter

During winter, snow is exposed to consistently sub-zero temperatures, transforming snow to a frozen solid, and the colder the ice, the greater the penetration depth becomes. Theoretically speaking, if the snow has no conductivity then the snow can be regarded as "transparent," allowing for the radar waves to image the ice surface and volume. This is why multi-year ice has a greater penetration depth than first-year ice; once the ice begins to freeze, expelling brine, especially over many cycles, the ice approaches near zero salinity. Once the penetration depth of multi-year ice increases, the radar waves have a higher chance to pick up discrete scatterers like air bubbles in the ice. These bubbles have dimensions that are within an order of magnitude in size of the radar wavelengths at X- and C-band frequencies (3-6 cm) and their contribution significantly increases the backscatter signal of multi-year ice (Onstott and Shuchman, 2004). Because first-year ice has just started to freeze, however, the amount of salinity is still large. The backscatter of this young ice is dominated by the surface, since the penetration depth is small in this case.

A SAR image taken by Radarsat-2 during April 1st, 2014 near the Alaskan/Canadian border is shown in figure 4.8. Although this is not an image taken in the winter, the ice conditions seen in the scene resemble those of freezing; older ice has bright backscatter values in the northern part of the image, while the thin new/first-year ice has a much darker signature.

Over the winter, multi-year ice signatures are temporally stable, but vary within the region since it was observed that the backscatter of this ice type is correlated to physical processes such as extensive pressure ridging (Kwok and Cunningham, 1994). For first-year ice, it is important to note whether the ice is deformed or undeformed. During the winter, it
can be seen that first-year un-deformed (FY-U ice) has no major deformation features, with a mean backscatter of -14dB, while first-year deformed (FY-D) ice has extensively ridged ice with mean of -17 to -18 dB (using ERS-1, C-band) (Kwok and Cunningham, 1994).

4.1.3 Sea Ice Backscatter During Summer

During the summer melt season, the snow that was once shielding the ice begins to melt. This ’melt onset’ stage can be categorized in two regimes called pendular and funicular. The pendular regime consists of water held in the interstices of the snow pack, often indicated by a rapid rise in backscatter from first-year ice sites and a decrease in scattering on multi-year sites (Barber and Yackel, 1999). The increase in backscatter from first-year ice is due to volume scattering, where brine and water volume are both in liquid phase. However, further accumulation of free water in the snow and on the ice surface changes
the scattering cross-sections enough to make the distinction between ice types particularly difficult during this time (Livingstone et al., 1987a). In fact, moist, wet snow has been reported to limit the backscatter response to the snow layer, making it tedious to pick up signatures from the ice below it. Therefore, thin and thick ice may have identical backscatter signatures because of this moist wet snow, making it extremely problematic for radar to differentiate ice types under wet snow conditions.

Once the snow is saturated with liquid, a shift towards the funicular regime occurs. During the funicular stage, snow cover begins to drain and a slushy water mixture pools at the base of the ice pack (Barber and Yackel, 1999). The sea ice, unprotected, then reaches its melt point (0°C), where it becomes porous. Water and brine will start to “flush” through the ice’s intergranular veins, freshening the sea ice along the way. While this is happening, the permanent shortwave radiation hitting the sea ice surface during the summer begins to form puddles, called melt ponds.

We can further add that a ‘true’ melt-onset stage is not quite clearly defined. For example, melt may not immediately be seen on the surface, but at the base of the ice, causing basal melt. This melt-onset definition is then based on an ocean perspective. Basal melt occurs when the solar heat flux reaches the upper ocean through open leads, exceeding conductive heat losses through the ice. Another way to define melt-onset is by analyzing ‘surface’ melt, where we witness snow melt (decrease in snow depth) and pond formation. Figure 4.9 shows a typical SAR image of sea ice experiencing surface-melt, as melt ponds cover the ice floes throughout the scene.
Figure 4.9: SAR image of sea ice as captured by TerraSAR-X on June 26th, 2014 over the Beaufort Sea. Red arrows point to a few examples of melt ponds on the surface of the ice.
4.2 Optical

Optical remote sensing of sea ice falls into the category of passive remote sensing. There exists a variety of passive sensors that can study sea ice properties such as optical satellites, radiometers, spectroradiometers, and cameras mounted on drones/helicopters. In our work, we focus on optical satellite sensors. The meaning of optical comes from measuring reflected energy in the electromagnetic spectrum coincident with the wavelength radiation from the sun, from approximately 250-2500 nm (Perovich, 1996), which is referred to shortwave radiation. Optical images show the appearance of targets in the visible spectrum (0.4 – 0.7μm), which is the range that the human eye can discern. A panchromatic image, on the other hand, is created when the imaging sensor is sensitive to a wide range of wavelengths/frequencies within a single channel.

The work done in this thesis makes use of panchromatic images, which come as Literal Image Derived Products (LIDPs) from the U.S Geological Survey Global Fiducial Library (GFL), and shall be discussed in further detail in chapter 5. Figure 4.10 shows a series of LIDPs capturing the development of melt ponds on a drifting ice parcel during the summer of 2009, in the Arctic. From figure 4.10 we can see that panchromatic images provide good contrast between open water (darkest pixels), melt ponds (dark gray pixels) and sea ice (light gray/white pixels).

4.3 Chapter 4 Summary

In this chapter, two types of sensors were discussed for sea ice remote sensing: SAR and optical (panchromatic) satellites. SAR, an active sensor, is able to operate independently of solar illumination and weather, and is able to create high resolution images by synthetically
synthesizing its aperture. SAR measures the backscatter of sea ice, which is dependent on a variety of parameters (from the radar as well as the physical properties of the material). Panchromatic satellites, require solar illumination and are limited to cloud cover, however they can provide very high resolution images, rendering them very useful for discriminating ice types due to good pixel contrast. Next, in chapter 5, we discuss in more detail the specifications of these sensors, along with a few in-situ instruments used in our studies.
CHAPTER 5

Data Used

The data used are here include remote sensing and in-situ measurements, designed to compliment and validate one another to observe the state of the sea ice as the Beaufort Sea’s marginal ice zone evolves. In this chapter, we discuss three different satellite sensors comprising two types of SAR and an optical sensor used to capture information such as open water fraction and melt ponds at different scales and spatial resolutions.

We also introduce in-situ systems designed to take point measurements on the ice, which were deployed during the Marginal Ice Zone (MIZ) program in 2014 (Lee et al., 2012). During the time of instrument deployment, various sensors including autonomous technologies (ice-based, floats, drifters, and gliders) were strategically placed along an array across the MIZ pack, as shown in figure 5.1. The MIZ program, funded by The Office of Naval Research, encompassed a large collaboration between scientists from various university and government institutions. The goal of the MIZ program aimed to understand the air-ice-ocean-wave dynamics in the Beaufort Sea, in order to improve future predictions of ice edge location, floe size and concentration (Lee et al., 2012).

For this study, we exploit three MIZ in-situ instruments, whose data include images from cameras mounted on wave buoys, autonomous weather stations, and ice mass-balance
Figure 5.1: MIZ-deployed in-situ arrays (Lee et al., 2012).

buoys. Information from these systems complements the data from nearly-coincident satellite images, and these can be used to analyze the evolution of melt ponds and the thermodynamic state of the ice. During this work, we will often refer to instruments on particular clusters. Clusters make up a set of instruments along the MIZ sampling array (Figure 5.1), where cluster 1 is the set of instruments closest to the Alaskan Bay, and cluster 5 is the set of instruments at the northernmost part of the pack ice. The distance between clusters is 100 km.

5.1 Satellite Data

During the late spring up to late fall of 2014, several TerraSAR-X (TSX) and RADARSAT-2 (RS2) images were collected, processed and calibrated at the Center of Southeastern Tropical Advanced Remote Sensing (CSTARS) of the University of Miami, following the MIZ program instrument deployment. The TSX is a German SAR satellite used for both
scientific and commercial applications, and is the continuation of XSAR (1994) and SRTM (2000), which were space shuttle experiments. TSX is an advanced high-resolution X-band radar satellite with operations in Spotlight, Stripmap, and ScanSAR modes with various types of polarizations and product types (Werninghaus, 2004). In this work however, we strictly use HH data in Stripmap mode, which covers a scene size of 30 km (width) × 50 km (length), and a multi-look ground range detected (MGD) product type. Figure 5.2 shows a typical TSX SAR image of sea ice during mid-August of 2014, displaying a ‘quicklook’ image with a coarse resolution (full image scene), and a zoomed area (full resolution).

The RS2 is a follow-on to RS1. RS1 completed its first orbit in 1996. RS2 is a Canadian satellite, which similarly like TSX, is also used for scientific and commercial applications. The RS2 SAR operates in C-band, with available products in several different beam modes; the fundamental imaging modes include Single Beam, ScanSAR and Spotlight (Slade, 2009). Like TSX, RS2 can transmit and receive in different types of polarization.
For our study, we used RS2 in ScanSAR mode (swath of more than 500 km), with HH polarization, at 50 m resolution. Figure 5.3 shows a typical RS2 SAR image capturing a part of the Beaufort Sea’s ice edge.

During 1995, a group of scientists, were granted high-level security clearances and given access to portions of intelligence data. This group, named Medea, were scientists appointed by the vice president of the United States to review and advise on acquisitions of imagery obtained by classified intelligence satellites (Kwok and Untersteiner, 2011). The Medea scientists were determining whether the data and intelligence assets could aid in the study of global warming, ocean temperatures, vegetation, forest cover and the condition of polar ice caps, to name a few (Richelson, 1998). Indeed, the datasets proved to be of significant scientific value, and thus a collaboration between these scientists and the intelligence community was established, however the group was disbanded in 2000. The Medea scientists re-assembled in 2008, and in June 2009, the United States Geological Survey (USGS) released several classified satellite images from a panchromatic imager with 1-m resolution as literal image derived products (LIDPs) (Kwok, 2014). These images cover locations in
the Arctic Basin such as the Beaufort, Canadian Arctic, Fram Strait, Siberia, Chukchi and Barrow. The images cover an area of more than 15 km × 15 km, are geometrically rectified based on imaging geometry and then re-sampled onto a Universal Transverse Mercator (UTM) map projection (Kwok, 2014). As reported by the National Research Council, the LIDPs contain significantly valuable information to aid polar scientific research which encompasses topics such as snow depth, lateral melting and ice deformation, and to assess the performance of other satellite images (Counsel, 2009). In fact, along with having an extremely high resolution, the LIDPs also have a high repeat rate in the mentioned Arctic regions, with some steered on a daily basis.

Studies have shown that using passive microwave sensors such as radiometers, can lead to uncertainties in the detection of melt ponds; the signature of ponds are often mistaken for open water during the onset of melt (Comiso and Kwok, 1996). SAR images also present difficulty in discriminating between melt ponds and open water accurately; the work of (Kwok et al., 1992) explains that for SAR images in single polarization mode, the backscatter from open water is not often unique and varies with incidence angle and wind speed, where its signature can overlap with that of the surrounding ice. This suggests that we need an accurate dataset to validate algorithms which quantify open water using radiometers and SAR, making the LIDP dataset an attractive option. One of the drawbacks in using optical sensors in the arctic is their inability to penetrate through clouds; arctic stratus clouds are often present during the summer. Figure 5.4 shows how a large percentage of clouds in an LIDP can render it completely useless for surface analysis, however when clouds are not present in an image scene, LIDPs can provide extremely high resolution.

During the MIZ program campaign in 2014, several LIDPs were collected to spatially and temporally collocate with in-situ instruments deployed along the MIZ arrays, among
other sensors. At least 1,681 images were collected for spring-fall, less than 10% were usable due to the effect of clouds. The time stamp reported from the LIDPs are of the nearest day, however sun elevation is provided as a measure of solar illumination during time of acquisition (Kwok, 2014). Because the LIDPs come from a previously-classified dataset, the sensors and bands from their original source are unknown. Exploiting the sun elevation angle can help understand the ranges in image brightness; when the sun is at higher angles, the illumination within the image scene will give a better image quality, and vice versa. However, this depends on several factors and applications. The images are stretched to fit a range of 256 gray levels, which gives rise to inconsistent radiometric image-to-image information.
5.2 In-situ Data

5.2.1 Ice Mass-Balance Buoy

Understanding sea ice mass-balance is critical to study the seasonal evolution of sea ice thickness, especially since mass-balance processes require continuous monitoring. Past measurements of sea ice mass-balance have contributed to our understanding of ice growth and melt (Kwok et al., 1998, Perovich and Richter-Menge, 2006, Richter-Menge et al., 2006), however with a high cost and low spatial distribution trade-off. A novel design of an ice mass-balance buoy (IMB) was introduced to reduce these costs, allowing for a more extensive IMB network through the ice pack (Jackson et al., 2013).

The IMB design was inspired by a previous successful IMB from the Cold Regions Research Engineering Laboratory (CRREL), which used a thermistor chain to measure the temperature profile through the ice and integrated two acoustic sensors to monitor snow and ice (Richter-Menge et al., 2006). The novel IMB is easy to deploy through a standard 2 inch auger hole and has a custom designed low-cost controller and logger. The IMB transmits data via the satellite-based communication Iridium, which facilitates the reduction of cost and allows the user to take more control over data sampling after deployment (Jackson et al., 2013). Figure 5.5 shows the schematic for the main system components of the IMB. Thermistor chains consist of 0.5 m sections, with 2 cm sensor spacing, 1 m above the ice, and 3 m below (Jackson et al., 2013). These thermistor chains measure temperature profiles of the air-ice-upper ocean in order to analyze ice growth/decay or snow accumulation/ablation.
Because the IMB is drifting with its ice floe, it will allow for continuous measurements and give insight on the evolution of upper ocean heat throughout different seasons. A total of 25 IMBs were deployed on the ice, as indicated by red squares in Figure 5.1.

5.2.2 Autonomous Weather Station

The automated weather station (AWS) is an operational meteorological instrument that measures GPS location, wind speed and direction, humidity, air-temperature, pressure, solar radiation, and floe rotation. This instrument is field-tested, and is based around a Campbell CR1000 logger and Iridium SBD messaging (Lee et al., 2012). These specific components allow for altering the sampling rate remotely, and have the ability to transmit data in real time for web-viewing. Figure 5.6 shows a general layout of the AWS. A total of five AWSs were set-up along the MIZ track array, depicted as green circles in Figure 5.1.

Winds speed data are gathered from a wind monitor, mounted upon the main AWS mast at a height of 2.5 m, and a one-minute mean vector wind is calculated from six, 10-second speed and six direction measurements using a CR1000 Windvector command. The
reference for wind direction is given by a compass with heading output. Temperature and humidity are obtained by the naturally ventilated HMP155 by Vaisala, which includes a radiation shield, and is also bolted to the mast at a height of 2.5 m. Pressure is taken from a CS100 Setra barometric pressure, mounted inside the electronics enclosure. Solar radiation information is gathered from a net radiometer, equivalent to a Kipp and Zonen NR Lite2 Net Radiometer (personal correspondence with Jeremy Wilkinson, British Antarctic Survey).

5.2.3 Wave Buoy Cameras

While we will not be using wave buoy data, the wave buoys deployed during the MIZ program along the sampling track were equipped with cameras. The recorded every six hours at different clusters in the array. For Clusters 1 and 2, the cameras began sampling
during March 18, and ended on September 1, 2014. The cameras of clusters 3 and 4 began on March 18, and ended on September 1, 2014. Cluster 5 images were unavailable. Figure 5.7 shows images taken by wave buoy cameras on March, mid-June and September. These images can provide helpful qualitative information about the current state of the ice and weather within the clusters.

5.3 Chapter 5 Summary

In this chapter, we described the satellite sensors used in our work, which comprised of TSX, RS2 and LIDPs. We also introduced three in-situ instruments: IMBs, AWSs and wave buoy cameras, which were deployed during the MIZ program in 2014. The data from in-situ instruments can complement remote sensing measurements, and act as a validation dataset. Now that we have introduced the necessary background information on sea ice, remote sensing, and data to be used, we move on to the background of our classification methods: machine learning classification algorithms.
CHAPTER 6

Classification Algorithms

Classification algorithms encompass two types of methods, (1) supervised learning and (2) unsupervised learning. Supervised learning incorporates user defined training data in order to build models which later classify data they have never seen. The main idea behind supervised learning makes use of statistical learning theory, which learns from examples (Zhou et al., 2002). Unsupervised learning does not require training data from the user, and instead, for example, makes decisions by asking how many classes are to be found in an image, and separates it that way. The work covered here, only makes use of supervised learning methods.

Supervised learning is the machine learning task of inferring a function from labeled training data (Mohri et al., 2012). Machine learning classifiers learn a set of rules in order to distinguish between classes. In this work, we focus on both binary (ice or open water) and multi-class machine learning classification (ice, open water, and melt ponds). Figure 6.1 shows the task of supervised learning using machine learning classifiers.

The initial steps to train a machine learning classifier require training data, which include feature and label data. Features are a set of data used to discover potentially predictive relationships. For example, in image classification, feature data often consist of pixel in-
Figure 6.1: Schematic for predicting class labels.

tensity and or image textural information. Studies have demonstrated that accounting for textural information in classification of SAR images improves the accuracy of the results, since certain features of interest can be highlighted by exploiting particular textures (Bogdanov et al., 2005, Hara et al., 1994, Zakhvatkina et al., 2013). These features are stored in a vectorized format, usually denoted with the symbol $X$. There is no general consensus on how many features should be incorporated into pattern recognition classifiers. Instead, relying on domain knowledge of the general problem to be solved is key to providing the classifiers with adequate training information. For example, if the dataset has high variability, the features must be able to capture the high complexity and variation. In image classification, the labels are a fixed set of categorical outcomes, which are represented by integers (e.g., -1 for open water, +1 for ice). After the training dataset is completed, this information can be fed into any machine learning classifier. The classifier will “learn” a general rule that will map inputs to outputs. The goal for the algorithm is to generalize well for data it has never seen, based on the learned examples from the training dataset.

Although there are numerous algorithms used for image classification, five will be discussed in this work. The classification schemes include three parametric classifiers: Neural
Networks, Support Vector Machines, and Naive Bayes. Two non-parametric classifiers including: K-Nearest neighbor and discriminant analysis schemes are also investigated. Parametric classifiers are algorithms that can simplify functions to a known form. These types of classifiers are often easier to understand and their results are more manageable to interpret than their non-parametric counterparts. Another benefit of using parametric classifiers is speed, since these algorithms tend to learn very fast from the data. On the other hand, parametric classifiers are often constrained to a specified form, limited to simpler classification problems. Non-parametric means that the classifier does not assume a particular form of a density function, but will usually estimate one (Babich and Camps, 1996). Non-parametric procedures can be used with arbitrary distributions, and are free to learn any functional form from the given training data. However, these algorithms require much more data than parametric classifiers, are slower to train and have a tendency for over-fitting. A good reason to use non-parametric classifiers occurs when we don’t know the true form of the density function, or when the data do not fit a common density model.

6.1 Parametric Classifiers

6.1.1 Neural Networks

Neural networks (NNs) are processing devices originally inspired by the human brain’s ability to solve highly complex, nonlinear problems. The foundations of NNs stem from research showing they could compute any arithmetic or logical function (McCulloch and Pitts, 1943). During the late 1950’s, Frank Rosenblatt introduced the first practical application of NNs using the perceptron network and associated learning rule (Rosenblatt, 1958), leading to numerous research and applications of NNs.
In order to design a neural network, a training sample must be provided. Figure 6.2 shows the general layout of the neural network architecture used in this study. The training dataset consists of training inputs and the respective outputs (ice or open water, for example). The NN is simulated by feeding the training data in a normalized format, which passes through one hidden layer ($L_1$), and one output layer ($L_2$). Each neuron at a specific layer is connected with all the neurons of the previous layer (for figure 6.2 showing 10 neurons). For each individual connection, there exists a certain weight parameter, which is initialized randomly. In NNs, the weight represents the strength of a connection between two nodes.

The process to find the optimal combination of weights can be computationally expensive, according to the number of features considered and the network architecture. A popular method for accomplishing this task is the back-propagation algorithm. The goal of the back-propagation algorithm is to find the minimum of the error function in weight space based on the difference between the calculations of the neural network and the expected (target) values (Rojas, 1996). An error function is used to estimate the performance of the network and updates the weights accordingly. This error function depends on the
transfer function (input function and activation function) at each neuron. For this purpose, activation functions which are continuous and infinitely differentiable, are preferred (Rojas, 1996). One of the most popular types of activation functions are sigmoid functions, especially the logistic sigmoid and tan sigmoid functions. Figure 6.3 adopted from Rojas (1996) shows the calculation of the error function.

Given a training set \( \{(x_1, t_1), \ldots, (x_n, t_n)\} \) composed of \( n \)-ordered pairs of \( n \)- and \( m \)-dimensional vectors, where the \( x \) represents the input features and \( t \) is the target values, when we present input to the network, it produces an output \( O_i \) which is different from the target \( t_j \). Through the use of an error function, the goal of training is to make the values of \( O_i \) and \( t_j \) as close to equal as possible by minimizing the error between these values and propagating this error backwards through the network to change the values of the weight. This process is done iteratively until a satisfactory minimum value of the error function is achieved.

In the case of back-propagation, the first step is the selection of the error function. Although the mean squared error function is a common choice, the cross entropy error function
\[ E_i = -t_i \ln o_i - (1 - t_i) \ln(1 - t_i) \]  

(6.1)

for binary classification was used in this study. Using the cross entropy we seek to optimize the problem through minimization of the error. The most common method involves finding the steepest decrease in error function values. A problem with methods like gradient descent is that they involve a line search, i.e., a calculation to find the direction of steepest descent in a chosen direction and step size which can result in poor convergence to a minimum error (Tang and Skorin-Kapov, 2001). Conjugate gradient methods utilize the second derivatives of the error function to find the local minimum using information from the second order approximation of the error function \( E \). As described in detail in (Moller, 1993), a standard conjugate gradient method proceeds as follows:

1. An initial weight vector \( \tilde{w}_1 \) and a search direction \( \tilde{p}_1 = \tilde{r}_1 = -E' \tilde{w}_1 \) (which is part of a conjugate system \( \tilde{p}_1, ..., \tilde{p}_N \)) are chosen at iteration \( k = 1 \).

2. A parameter \( s_k \), containing second-order information calculated from the Hessian matrix \( E'' \tilde{w} \)

\[ \tilde{s}_k = E''(\tilde{w}_k)\tilde{p}_k \]  

(6.2)

\[ \delta_k = \tilde{p}_k^T \tilde{s}_k \]

3. The step size \( \alpha_k \) is calculated

\[ \mu_k = \tilde{p}_k^T \tilde{r}_k \]  

(6.3)

\[ \alpha_k = \frac{\mu_k}{\delta_k} \]
4. The weight vector is updated

\[ \tilde{w}_{k+1} = \tilde{w}_k + \alpha_k \tilde{p}_k \]

\[ \tilde{r}_{k+1} = -E''(\tilde{w}_{k+1}) \]  

(6.4)

5. If \( k \text{mod} N = 0 \) then restart algorithm \( \tilde{p}_{k+1} = \tilde{r}_{k+1} \); else a new conjugate direction is chosen

\[ \beta_k = \frac{(|r_{k+1}|)^2 - \tilde{r}_{k+1}^T \tilde{r}_k}{\mu_k} \]

\[ \tilde{p}_{k+1} = \tilde{r}_{k+1} + \beta_k \tilde{p}_k \]  

(6.5)

6. If the steepest descent direction \( \tilde{r}_k \neq 0 \), then set \( k = k + 1 \) and return to step 2; else the algorithm is terminated and the desired minimum weight vector \( \tilde{w}_{k+1} \) is returned.

Unlike gradient descent in standard back-propagation, a conjugate gradient method moves in a direction conjugate to the directions of previous steps, which lead to a faster convergence (Fischer and Staufer, 1999).

Once the optimal weights are calculated, each neuron sums the inputs given, adds a bias rule and applies an activation function to the result (Loukachine and Loeb, 2003). The activation function is a decision making function which determines the “presence” of a particular feature, and the bias value allows the user to shift the activation function. The output through the network can be mathematically summarized as:

\[ a^n = F^n(x^n) = F^n(W^n a^{n-1} + b^n) \]  

(6.6)

where \( a^n \) denotes the vector of the \( n \)-th layer output, \( F^n \) is the vector of the neuron activation function (where \( x^n \) are vectors of real numbers), \( W^n \) is the weight matrix and \( b^n \) is a bias vector of the \( n \)th layer.
6.1.2 Support Vector Machines

The foundations of SVMs have been developed by (Burges, 1998, Schölkopf and Smola, 1998, Vapnik et al., 1996). SVMs have gained popularity due to their attractive features and empirical performance. The goal of SVMs is to produce a classifier that will work on unseen examples (i.e., it generalizes well). If we assume that our training set consists of \( N \) vectors from the \( d \)-dimensional feature space \( x_i \in \mathbb{R}^d \) \((i = 1, 2, ..., N)\). A target \( y_i \in \{-1, +1\} \) is related to each vector \( x_i \). Assuming that two classes are linearly separable, there exists at least one hyperplane defined by a weight vector \( w \) (normal to the hyperplane) and a bias \( b \) that can separate these two classes with minimal error. The discriminant function associated with the hyperplane in a linear SVM is defined as

\[
f(x) = w \cdot x + b
\]  

(6.7)

In order to calculate this hyperplane, we need to calculate the weight vector and bias so that

\[
y_i(w \cdot x + b) > 0, \quad \text{with} \quad i = 1, 2, ..., N.
\]  

(6.8)

SVMs try to find the optimal hyperplane that maximizes the distance between the closest training sample and the separating hyperplane (Melgani and Bruzzone, 2004). The distance between the separating hyperplane and the closest training sample is equal to \( 1/||w|| \) by rescaling the weight vector and bias such that

\[
\min_{i=1,2,...,N} y_i(w \cdot x + b) \geq 1
\]  

(6.9)

The distance between two classes separated by the hyperplane is called the margin, and is presented as \( 2/||w|| \). The greater the margin, the better the separation will be between
Figure 6.4: Schematic of a Support Vector Machine and its separating hyperplane.

the classes. Figure 6.4 shows the general layout of the hyperplane which separates two classes.

For pattern recognition, SVMs have the following optimization problem (Burges, 1998):

$$\min \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \xi_i,$$  \hspace{1cm} (6.10)

(where $\xi_i$ is a slack variable)

$$\text{s.t.} \quad y_i((w \cdot x_i) + b) \geq 1 - \xi_i$$ \hspace{1cm} (6.11)

$$\xi_i \geq 0, \quad i = 1, \ldots, N$$ \hspace{1cm} (6.12)

in which $(\cdot, \cdot)$ indicates the inner product between the weights and the feature vectors. When minimizing the first term of equation 6.10, we are essentially trying to avoid over-fitting while the second term minimizes the empirical risk.
We can translate this linearly constrained optimization problem through a Lagrangian formulation by the following:

$$\max \ W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j), \quad (6.13)$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \quad (6.14)$$

$$\alpha_i \in [0, C], \quad i = 1, \ldots, N. \quad (6.15)$$

where $\alpha_i$’s denote Lagrange multipliers, estimated using quadratic programming (Vapnik, 1998). The discriminant function associated with the optimal hyperplane becomes an equation which depends on the Lagrange multipliers and on the training samples, i.e.,

$$f(x) = \sum_{i \in S} \alpha_i y_i (x_i \cdot x_j) + b \quad (6.16)$$

Here, $S$ is the subset of the training samples which correspond to the non-zero $\alpha_i$’s. The $\alpha_i$’s weigh each individual training sample according to their importance in determining the discriminant function (Melgani and Bruzzone, 2004). The training samples which attain non-zero weights are called support vectors and lie at a distance of $1 / \|w\|$ from the optimal hyperplane.

### 6.1.3 Naive Bayes Method

Before introducing the Naive Bayes classifier, a refresher on probability must be introduced. Let us begin by introducing the meaning of state of a variable. The meaning of state
will often be clear in probability problems. One example arises when we are discussing experiments about a coin, \( c \), where the meaning of \( p(\text{heads}) \) is clear from the context. However, when summing or performing other mathematical operations over a variable \( \sum_x f(x) \) (where the variable here is \( x \)), we interpret that all states of \( x \) are included.

For the purposes of Bayesian statistics and probability, events are expressions about random variables, such as two tails in six coin tosses. Two events are mutually exclusive if they cannot occur simultaneously at one time. We must also define that the probability of an event \( x \) occurring is represented by a value between 0 and 1, where \( p(x) = 1 \) means we are certain the event occurs, and \( p(x) = 0 \) means we are certain that the event does not occur. The summation of the probability over all states is:

\[
\sum_x p(x = x) = 1
\]  
(6.17)

Such probabilities are normalized (Barber, 2012). For convenience, we can write \( \sum_x p(x) = 1 \). The interaction between two events \( x \) and \( y \) can be written as \( p(x \text{ or } y) = p(x) + p(y) - p(x \text{ and } y) \). We will be using the shorthand \( p(x, y) \) for \( p(x \text{ and } y) \).

If we have a joint distribution \( p(x, y) \) the distribution of a single variable is given by

\[
p(x) = \sum_y p(x, y)
\]  
(6.18)

In this case, \( p(x) \) will be termed a marginal of the joint probability distribution \( p(x, y) \) (Barber, 2012).

An important concept in Bayesian theory is the conditional probability, known as Bayes’ Rule. The probability of an event \( x \) conditioned on knowing about event \( y \) (the probability of \( x \) given \( y \)), is defined as:
\[ p(x|y) \equiv \frac{p(x,y)}{p(y)} \]  \hfill (6.19)

where if \( p(y) = 0 \), then \( p(x|y) \) is not defined. If we revisit probability density functions, and we include a continuous variable \( x \), then the probability density \( p(x) \) is defined such that \( p(x) \geq 0 \) and

\[
\int_{-\infty}^{\infty} p(x) dx = 1 \hfill (6.20)
\]

Unlike probabilities, probability densities can take positive values greater than 1 (Barber, 2012). The topic of independence is also key to discussing Bayesian probability. If two events \( x \) and \( y \) are independent, knowing one event provides no additional information about the other event and can be mathematically expressed as \( p(x,y) = p(x)p(y) \). This falls under the constraint that \( p(x) \neq 0 \) and \( p(y) \neq 0 \) independent of \( x \) and \( y \) is equivalent to \( p(x|y) = p(x) \iff p(y|x) = p(y) \).

The Naive Bayes method is a statistical classifier based on Bayes’ Theorem which applies density estimation to the data. The naive assumption of class conditional independence, which presumes the effect of a feature on a given class is independent of the values of the other features, is frequently made in order to decrease the computational cost (John and Langley, 1995).

The Naive Bayes classifier assigns observations to the most probable class, or the maximum a posteriori decision rule. This is achieved in three steps. First, the algorithm estimates the densities of the predictors within each class. Second, the posterior probabilities are modeled according to Bayes rule. That is, for all \( k = 1, \ldots, K \),
\[ \hat{P}(y = k|x_1, \ldots, x_p) = \frac{\pi(y = k) \prod_{j=1}^{p} P(x_j|y = k)}{\sum_{k=1}^{K} \pi(y = k) \prod_{j=1}^{p} P(x_j|y = k)} \quad (6.21) \]

where \( Y \) is the random variable that corresponds to the class index of an observation, \( X_1, \ldots, X_p \) are the random predictors of an observation and \( \pi(Y = k) \) is the prior probability that a class index is \( k \). Finally, the algorithm classifies an observation by estimating the posterior probability for each class. It then assigns the observation to the class which yields the maximum posterior probability.

### 6.2 Non-Parametric Classifiers

#### 6.2.1 K-Nearest Neighbor Method

The K-Nearest neighbor (K-NN) is a non-parametric lazy learning algorithm. This means the algorithm makes no assumptions on the underlying data distribution, and it does not use the training data points to do any generalization. This means that there is no explicit training phase, or it is very minimal, leading to an efficient training phase. All the training data are needed during the testing phase, unlike other techniques such as SVMs, where non-support vectors could be discarded without penalty.

There are certain assumptions to be made with the K-NN. First, K-NN assumes that the data are in a feature space, where data points are in a metric 'space' (distance space). The data can represent scalars or vectors, and can have multiple dimensions. Distance may or may not be euclidean in nature, although this distance metric is widely used.

K-NN classifies a data point by assigning it to the label most frequently present in the k-Nearest neighbors. K-NN takes into account \( k \) neighbors, which is less sensitive to noise...
than one nearest-neighbor (Duda et al., 2012). When given \( k \), this number decides how many neighbors influence the classification (and the neighbors are defined based on the distance metric). The number \( k \) is usually an odd number if the number of classes is two, and if \( k = 1 \), the algorithm is called nearest-neighbor algorithm.

The \( k \)-Nearest neighbor fit for an output is given as

\[
\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i
\]  

(6.22)

where \( N_k(x) \) is the neighborhood of \( x \) defined by the \( k \)-closest points \( x_i \) in our training sample (Friedman et al., 2001). Different behaviors are observed when \( k \) is varied. For example, if \( k = 1 \), we have the simplest scenario: We let \( x \) be a point to be labeled, where the algorithm finds the point closest to \( x \). If we let the closest point to \( x \) be \( y \), the nearest-neighbor rule will assign the label of point \( y \) to \( x \).

This method may sound over-simplistic and bound to have large errors. This is rightfully so when the number of data points is small, however, for large numbers of data points the chance that the label of \( x \) and \( y \) are the same is very high; if the density of the data is very high, within any subspace of the data there will be a reasonable number of points. If we inspect a point \( x \) within a subspace, having a large number of neighbors, and another point \( y \) is the nearest-neighbor and sufficiently close, we can assume that the probability of \( x \) and \( y \) being in the same class is fair. In order to assign a classification label with the smallest classification error, the bound is

\[
P^* \leq P \leq P^* \left( 2 - \frac{c}{c - 1} P^* \right)
\]  

(6.23)
where $P^*$ is the Bayes error rate, $c$ is the number of classes and $P$ is the error rate of nearest-neighbor (Duda et al., 2012).

For the case when $k = 2$, we extend the nearest-neighbor algorithm by performing majority voting. As stated previously, $k$ is typically odd when the number of classes is 2. If we choose $k=5$, and we see three instances of C1 and two of C2, the K-NN algorithm will classify a new data point into C1, since this is the majority. Figure 6.5 shows another example, which makes use of the Euclidean distance: we have a training set that consists of two classes (class 1 and class 2), with five data points in each class. Two features ($x_1$ and $x_2$) are used to discriminate between the classes, making this a two-dimensional feature space. We are then presented with a new data point (black circle), and wish to classify this as either class 1 or class 2. If we use $k=3$, the new data point gets assigned to class 2. When we increase to $k=6$, the new data point is now classified to class 1.

### 6.2.2 Discriminant Analysis Method

The discriminant analysis method is designed to help distinguish differences between multiple sets of objects across several variables simultaneously. Each class ($Y$) generates
data \( (X) \) using a multivariate normal distribution. This means that the model assumes
the data has a Gaussian mixture distribution. Discriminant analysis-based classification
requires the computation of covariance matrices. When dealing with linear discriminant
analysis (LDA), we assume that all covariance matrices are equal (Cios et al., 1998, James
et al., 2014). Therefore, the model has the same covariance matrix for each class, with
only the means varying. In quadratic discriminant analysis (QDA), the covariance matrix
must be evaluated separately for each class, since they may differ from each other, there-
fore both means and covariances of each class vary. Under this modeling assumption, the
discriminant analysis algorithm infers the mean and covariance parameters of each class.

For LDA, the sample mean of each class must first be computed. Then, the sample
covariance is calculated by first subtracting the sample mean of each class from the obser-
vations of that class, and then taking the empirical covariance matrix of the final result. The
QDA computes the sample mean of each class, and then calculates the sample covariances
by subtracting the sample mean of each class from the observations of that class. The QDA
then takes the empirical covariance matrix of each class.

We begin by first constructing weighted classifiers by using the following scheme. Sup-
pose that we have a matrix \( M \), which is an \( N \)-by-\( K \) class membership matrix. We estimate
the class mean for unweighted data as:

\[
\hat{\mu}_k = \frac{\sum_{n=1}^{N} M_{nk} x_n}{\sum_{n=1}^{N} M_{nk}}
\]  

(6.24)

where \( n \) are the number of training points. If the weighted data has positive weights \( w_n \), the
natural generalization is
\[ \hat{\mu}_k = \frac{\sum_{n=1}^{N} M_{nk} w_n x_n}{\sum_{n=1}^{N} M_{nk} w_n} \]  

(6.25)

The estimated pooled-in covariance matrix for unweighted data is

\[ \hat{\Sigma} = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} M_{nk} (x_n - \hat{\mu}_k) (x_n - \hat{\mu}_k)^T}{N - K} \]  

(6.26)

When using quadratic discriminant analysis, in this case we can set \( K = 1 \). For weighted data, if we assume that the weights sum to 1, the estimate for the polled-in covariance matrix is

\[ \hat{\Sigma} = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} M_{nk} (x_n - \hat{\mu}_k) (x_n - \hat{\mu}_k)^T}{1 - \sum_{k=1}^{K} \frac{W^{(2)}_k}{W_k}} \]  

(6.27)

where, \( W_k = \sum_{n=1}^{N} M_{nk} w_n \) is the sum of the weights for class \( k \), and \( W^{(2)}_k = \sum_{n=1}^{N} M_{nk} w_n^{(2)} \) is the sum of squared weights per class.

Once the discriminant analysis model is trained, it is ready to classify or predict new data. In order to classify data, the algorithm minimizes the expected classification cost

\[ \hat{y} = \arg \min_{\hat{y}} \sum_{k=1}^{K} \hat{P}(k|x) C(\hat{y}|k) \]  

(6.28)

From equation 6.28, \( \hat{y} \) is the predicted class, \( K \) is the number of classes, \( \hat{P}(k|x) \) is the posterior probability of class \( k \) for observation \( x \), and \( C(\hat{y}|k) \) is the cost of mis-classifying a data point as \( \hat{y} \) when its true class is \( k \).

The posterior probability of an observation \( x \) belonging to a class \( k \) is the product of the prior probability and the multivariate normal density. The density function of the multivariate normal with mean \( \mu_k \) and covariance \( \Sigma_k \) at a point \( x \) is
\[ P(x|k) = \frac{1}{2\pi|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2} \left( x - \mu_k^T \right) \Sigma_k^{-1} (x - \mu_k) \right) \] (6.29)

where \(|\Sigma_k|\) is the determinant of the covariance (\(\Sigma_k\)), and \(\Sigma_k^{-1}\) is the inverse matrix. If we let \(P(k)\) represent the prior probability of class \(k\), the posterior probability that a point \(x\) belongs to class \(k\) is

\[ \hat{P}(k|x) = \frac{P(x|k)P(k)}{P(x)} \] (6.30)

where \(P(x)\) is a normalization constant (sum over \(k\) of \(P(x|k)P(k)\)). The prior probability can be chosen by the user, however in this work, we use a uniform prior: The prior probability of class \(K\) is 1 over the total number of classes. Observations of \(X\) values are divided into different regions for classification. If LDA is used, the regions are separated by straight lines and by conic sections (ellipses, hyperbolas, or parabolas) for QDA. Figure 6.6 shows an example of using discriminant analysis to separate three different flower types: Sertosa, Versicolor and Virginica.
Figure 6.6: Example of separating three different flower types using linear (left) and quadratic (right) discriminant analysis.
CHAPTER 7

Open Water Fraction: Methods

7.1 Purpose

A considerable amount of work has been devoted to using SAR imagery to estimate sea ice concentration and to locate areas of open water (Berg and Eriksson, 2012, Karvonen, 2012, 2014, Karvonen et al., 2005, Leigh et al., 2014, Scott et al., 2015). These studies ranged from using image backscatter in neural networks with autocorrelation statistics and ice chart comparisons (Berg and Eriksson, 2012), the use of Gaussian mixture models (Karvonen et al., 2005), and thresholding SAR images using a supervised maximum likelihood estimator assimilated with passive microwave images (Scott et al., 2015). The studies mentioned used SAR image resolutions coarser than 100 m, did not focus on the summer MIZ melt season, and were not particularly interested in developing methods for near real-time classification methods, which has been an interest to the Office of Naval Research for sea ice forecast models and ship navigation.

In this chapter, we train five machine learning classification models including (1) Neural Networks (NNs), (2) Linear Support Vector Machines (LSVMs), (3) Naive Bayes, (4) K-nearest neighbor and (5) discriminant analysis, to obtain open water fraction (OWF) from TSX Stripmap during boreal summer of 2014 in the Beaufort MIZ region. To investigate
the efficiency of the classifiers, we perform two separate validation tests. First, from the machine learning classifiers, we compute a confusion matrix and explore several output performance metrics. Second, we inspect OWF from the machine learning outputs images (both training and testing data) and compare them to near-coincident LIDP optical image-pairs.

All of the machine learning classification algorithms were trained using the pixel intensity information from the image itself, and various textures calculated from a modified first-order histogram technique using probability density of occurrences. The highlight of this research centers on fast and accurate detection of open water and ice, catered to methods for ship navigation, as inputs for near real-time ice forecast modeling, and for applications on large datasets. Our methods also seek to reduce overall melt pond contamination, since during the summer months, melt-onset can alter the radar backscatter and ponding can contaminate true OWF estimates (Cavalieri et al., 1990, Kim et al., 2013, Mäkynen et al., 2014).

### 7.2 Satellite Data

During the Summer of 2014, several TSX passes were collected, processed, and calibrated at the Center of Southeastern Tropical Advanced Remote Sensing (CSTARS) of the University of Miami during a field campaign from the MIZ program (Lee et al., 2012). The images were acquired in Stripmap mode, at HH polarization. The acquired TSX images had a resolution of 3 m, capturing features such as leads, cracks and individual ice floes.

Thirteen TSX images were collocated with eleven from the LIDP optical dataset. The collocation time frame was decided to 24 hours, where the SAR and optical image scenes
could captured similar features. Collocation times ranged from July 1, 2014 until September 26, 2014; six images from July, one from August and, six images from September. In order to develop nearly-identical image pairs between the SAR and optical, we extracted 21 small ‘sub-image-pairs’. This was achieved by choosing a center pixel near an identifiable floe on both datasets. A co-registration of images was not possible, due to the rotation/translation of ice floes. Nearly all images had an identical area of 36 km$^2$, however, two sub-image-pairs were extracted to an area of 2 km$^2$ due to the presence of clouds. A sub-image-pair example is shown in figure 7.1.

The inherent speckle noise found in SAR images decreases the radiometric and textural information, thus making image classification less accurate. We used an Enhanced Lee Edge Filter (ELEF) with a $25 \times 25$ window, which showed the best results over homogeneous areas such as open water, while still preserving pixel edges (Ju and Moloney, 1997, Lee, 1980). Figure 7.2 shows an original TSX image (a) next to its optical-pair (b), and the progression of TSX using various filtering windows (c-e). The original TSX clearly shows speckle noise over the open ocean where grainy-like signals can be mistaken for slush.
However, the optical image pair taken only an hour before TSX, shows no such features. Using the ELEF with a $5 \times 5$ window has almost no effect in noise reduction (figure 7.2c), while increasing the window size to $25 \times 25$ has the best effect (figure 7.2e). Although the SAR image is now essentially smoothed, areas in the open ocean appear much more homogeneous, and the smaller individual ice floes are still discerned.

After reducing the speckle noise, we transformed the image from linear normalized radar cross section, $\sigma^o$, to a logarithmic scale. After this transformation, we computed the histograms of all the images. We inspected 1% of the pixels that were in the extrema of each image histogram. These extreme values were replaced with the highest value in the upper end, and the lowest values in the lower. This method ensured that outliers did not contaminate the histogram’s distribution, since extrema tend to influence texture calculations.

### 7.3 Building the Training Dataset

As discussed in chapter 6, the first steps taken to use a machine learning supervised classification algorithm require training data. We selected 10 SAR sub-images for training and 11 for testing. The training images were selected to reflect the variability of OWF spanning from images having little open water, to mostly open water. Four images were from July, one from August, and five were from September, 2014. Heavy ponding was apparent on the July images, while re-freezing conditions appeared during the images from September. The image from August had the highest amount of open water compared to the others. We did not use every pixel of the images for training. Instead, we selected 20,000 training pixels: 50% ice and 50% open water. The optical images collocated with every
Figure 7.2: Progression of ELEF on SAR compared to its optical pair: TSX image taken on August, 26, 2014 03:36:56 UTC (a), LIDP optical image taken on August, 26, 2014 02:13:09 UTC (b), TSX using a $5 \times 5$ ELEF window (c), TSX using a $15 \times 15$ window, and TSX using a $25 \times 25$ window. All areas shown are approximately $36 \text{ km}^2$. 
SAR image, aiding the selection of our class labels. An example of how training pixels were selected is shown in figure 7.3.

### 7.3.1 Image Textures

Textures are complex visual patterns composed of individual or subpatterns, which have characteristic traits such as brightness, color, slope, size, etc (Materka et al., 1998). Studies have demonstrated that accounting for textural information in classification of SAR images improves the accuracy of the results, since certain features of interest can be highlighted by exploiting particular textures (Bogdanov et al., 2005, Hara et al., 1994, Zakhvatkina et al., 2013).

Approaches to texture analysis can be categorized into: structural, statistical, model-based and transform. In this work, we concentrate in the statistical approach, specifically, first-order textures, which represent the texture indirectly by the non-deterministic properties that govern the distributions and relationships between the gray levels of an image.
Second-order statistical textures use statistics given by pairs of pixels, and will not be discussed here, but they shall be mentioned in the discussion section.

First-order textures were preferred in this study, due to their relative fast computation time and high discrimination ability. First, we assumed the image was a function \( f(x, y) \) of two space variables \( x \) and \( y \), \( x = 0, 1, \ldots, N - 1 \) and \( y = 0, 1, \ldots, M - 1 \), where the function \( f(x, y) \) takes discrete values \( i = 0, 1, \ldots, G - 1 \), where \( G \) is the total number of intensity levels in the image (Materka et al., 1998). The histogram of an intensity-level image can be computed as

\[
h(i) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(f(x,y), i),
\]

(7.1)

Where \( \delta(j, i) \) is the Kronecker delta function.

Dividing the values \( h(i) \) by the total number of pixels in the image, we can approximate a probability density of occurrence of the gray levels

\[
p(i) = \frac{h(i)}{NM}, \quad i = 0, 1, \ldots, G
\]

(7.2)

Image features, most often called central moments, can be computed from the histogram to describe the first-order statistical properties of the image. A few examples include the mean

\[
\mu = \sum_{i=0}^{G} x_i p(i)
\]

(7.3)

and the variance

\[
\sigma^2 = \sum_{i=0}^{G} (x_i - \mu)^2 p(i)
\]
\[ \sigma^2 = \sum_{i=0}^{G} (x_i - \mu)^2 p(x). \] (7.4)

A full description of first-order statistical properties can be found in (Materka et al., 1998).

Lowitz (1983) proposed to compute first-order textures from local image histograms, where a pixel \((x, y)\) was centered in a window containing \(N\) pixels. These traditional first-order texture methods did not add significant information, because these conventional local first-order statistical measures are not interesting enough for a smoothed image using the ELEF, especially if they are computed for a window size much smaller than the one used for speckle filtering. Since our goal was to keep computation time to a minimum, it was critical that a small window size be used. Therefore, we used a combination of the methods described by Materka et al. (1998) and the work proposed by Lowitz (1983) to compute the mean, variance, skewness, and smoothness, however, with a few modifications.

We quantified the unique pixel values of an entire TSX intensity image, and computed the global histogram of those unique values. This global histogram is used for the computations of our small windows (instead of the local histogram for each window). Next, we computed the probabilities of getting those unique values (the probability density of occurrence). A window of \(5 \times 5\) is then centered to a pixel of an image array, and through a block processing system, an additional order of 2 columns and 2 lines on each side is added to the core window where possible (except when the window gets too close to the borders of the image array). This block processes in steps of the given window size, not like a sliding window in steps of 1 column and 1 line. This means, the window is essentially \(9 \times 9\) pixels, which is moved over the image in steps of 5 columns and 5 lines.
For example, the window of $9 \times 9$ pixels is moved over the image in steps of 5 columns and 5 lines, leading to an initial texture array of size $400 \times 400$, for an image that is $2000 \times 2000$ pixels. We then resize the texture array back to it’s original size ($2000 \times 2000$) using a bilinear interpolation; the output pixel value is a weighted average of pixels in the nearest $2 \times 2$ neighborhood. This block-wise method speeds-up the processing time very significantly, an important issue when working with many images. The reduced effective resolution of the texture property arrays obtained with this method is not very significant, since the image array itself has a reduced effective resolution after speckle filtering with a $25 \times 25$ window. Figure 7.4 shows the TSX image after speckle filtering (a), and the corresponding textural outputs, which are termed “modified” (b-e).
Figure 7.4: SAR intensity image (a) and its corresponding modified first-order textures: mean (b) variance (c) kurtosis (d) and smoothness (e).
Since we modified the first-order textures, how can we interpret them? For example, for the mean (figure 7.4b), if we would have computed this texture using the local histogram of the chosen window, it would compute the conventional mean gray level of that window (i.e., a conventional running mean). This is true for all other first-order texture quantities previously mentioned. For our method, defining a small window, but using the global histogram of the image gives rise to more interesting results.

For example, exploring the modified-mean texture, the SAR image has a few particularly bright spots because the global probability density of these large brightness values is small. Slightly less brighter areas do not get so much darker, because the global probability density of their gray level is large. This way, more robust classification criteria can be derived from different combinations of “dark” and “bright” in the intensity and mean image.

This non-standard way of computing a "modified" mean and other "modified" texture properties is quite nonlinear in the sense that it can make some features darker and other features brighter in a complicated way, depending on the overall statistics of the image. This may lead to strange results in some cases, but for our purpose of discriminating between open water and ice, with special attention to classifying ponds as ice, the proposed computation method has been found to produce results that lead to a clearly more successful classification than the use of conventionally computed texture properties.

Using machine learning methods for image classification is often domain specific. Certain features used in training classification models will not necessarily work across different image datasets. Therefore, we used a combination of the image intensity itself, plus the addition of first-order textures as training features in order to achieve different feature training model combinations. We used the binomial method for the combinable outcomes
of choosing the four textures as our feature inputs. The order of texture combination did not matter in the implementation of our data training structure, therefore, we chose only unique combination outputs. For example, one way to use the training data included using the intensity image, the variance and the mean. In total, there were 16 different ways to combine the features for training a model. Taking the 5 different classifiers into account, our methods totaled 80 classification models.

7.3.2 Machine Learning Classifiers Used

For the NN, we used a simple network containing one hidden layer and 10 neurons. A scaled conjugate gradient with back-propagation algorithm was implemented in order efficiently reduce errors. The support vector machine method incorporated the use of linear functions, which is essentially taking the inner dot product between the weights and features vectors. For Naive Bayes, each predictor was modeled using a Gaussian distribution, in which the Naive Bayes classifier estimated a separate normal distribution for each class by computing the mean and standard deviation of the training data in that class. For the K-NN method, we chose \( k = 3 \) neighbors, as this number has been used for K-NN classification of high resolution TSX images (Xing et al., 2013, Zhao et al., 2013), with a Euclidean distance metric, as this is the most popular distance measure used. The discriminant analysis technique usually takes advantage of linear functions, because they are less computationally expensive. As mentioned in chapter 6, in linear discriminant analysis we assume that the classes have a common covariance matrix \( \Sigma_k \). However, for our purposes we approximate a quadratic formula of
\[ \delta_k(x) = -\frac{1}{2} \log|\Sigma_k| - \frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log(\pi_k) , \]  

(7.5)

by substituting a pseudo-quadratic discriminant function in order to approximate a quadratic function. This method uses fewer computations to achieve a similar result to a true quadratic by using pseudoinverses of the quadratic covariance matrices, if necessary (Sewall et al., 2004). This approach ensured that each model trained would not result in failure, because this formula assumes that the covariance matrix can be different for each class, and is therefore estimated separately.

### 7.4 Optical OWF Data

To validate the classification outputs from the machine learning models, a manual threshold was conducted on the optical image dataset to obtain ‘true’ OWF. Open water signatures changed with onset-melt and freeze-up, hence, converting the optical images into binary (ice and open water), consisted of going through each individual image and finding the maximum intensity that could be recognized as water. For our study, we considered new ice to be classified as open water. This was done for two reasons. First, very thin new ice will pose limited threat to ship navigation, and therefore can be considered as open water. Second, thin new ice and open water have similar albedo values (Massom and Comiso, 1994). Therefore, these assumptions can satisfy users wishing to use these algorithms either for navigational purposes, or for limited ice/water discrimination if taking albedo into account.

For deep melt ponds present in the optical images having identical intensity values to open water, an erosion filter was introduced to close small features in order to reduce the
Figure 7.5 shows an example of the optical image before and after binary thresholding, where melt ponds are not counted.

### 7.5 Results

Due to the constant motion of sea ice and differences in time acquisition between the classified SAR images and their thresholded optical-pair, a pixel-by-pixel comparison was not possible. Therefore, to assess the performance of the classification algorithms, we computed total OWF estimates (percentage of open water in an image) between the classified images and their “ground-truth” pairs. We break down the results by separating statistics between the testing and training datasets. The testing models report statistics between the outputs of OWF from SAR images they have never seen against their ground-truth thresholded optical images. The training models also report the same statistics as the testing models, with the addition of model performance metrics. For performance metrics, we calculated the accuracy, false negatives (FN) and the Matthews correlation coefficient (MCC). The accuracy is calculated as the sum of true positives plus true negatives, divided by the
Table 7.1: Confusion matrix of best Neural Network ‘ks’

<table>
<thead>
<tr>
<th>n=20,00</th>
<th>Predicted: NO</th>
<th>Predicted: YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: NO</td>
<td>TN = 9,969</td>
<td>FP =31</td>
</tr>
<tr>
<td>Actual: YES</td>
<td>FN = 19</td>
<td>TP = 9,981</td>
</tr>
<tr>
<td></td>
<td>9,988</td>
<td>10,012</td>
</tr>
</tbody>
</table>

calculated positive and negative classes. We labeled the positive class as ice and open water as the negative class. FN is calculated by the number of times the model reported a pixel was ice, when it was actually open water. Finally, the MCC is a correlation coefficient between the observed and predicted binary classifications, where a value of +1 represents a perfect prediction.

Although a total of 80 machine learning classification models were created using the combinable method, we only show the best three of each algorithm, ranked by their RMSE. The RMSE was calculated as the average difference between the OWF from the optical dataset and the OWF from the classified SAR images. The classification models of interest ranged from implementing one texture, to including all four. The model code names work as follows: m=mean, v=variance, k=kurtosis and s=smoothness. For example, a model with a combination using “mvk” implemented the image intensity itself, the mean, variance and smoothness as feature inputs.

Inspecting each algorithm, the NN achieved the lowest RMSE (9.00%) and the lowest bias for the testing data (data never before seen), using the mean, kurtosis and smoothness. However, this same model reports a larger RMSE (11.55%) for the training data (images the model was trained on). Overall, the training accuracy of all the model combinations
Table 7.2: Confusion matrix of best SVM ‘mks’

<table>
<thead>
<tr>
<th></th>
<th>Predicted: NO</th>
<th>Predicted: YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: NO</td>
<td>TN = 9,906</td>
<td>FP = 94</td>
</tr>
<tr>
<td>Actual: YES</td>
<td>FN = 58</td>
<td>TP = 9,942</td>
</tr>
<tr>
<td></td>
<td>9,964</td>
<td>10,036</td>
</tr>
</tbody>
</table>

Table 7.3: Confusion matrix of best Naive Bayes ‘ms’

<table>
<thead>
<tr>
<th></th>
<th>Predicted: NO</th>
<th>Predicted: YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: NO</td>
<td>TN = 9,478</td>
<td>FP = 522</td>
</tr>
<tr>
<td>Actual: YES</td>
<td>FN = 253</td>
<td>TP = 9,747</td>
</tr>
<tr>
<td></td>
<td>9,731</td>
<td>10,269</td>
</tr>
</tbody>
</table>

Table 7.4: Confusion matrix of best K-NN ‘s’

<table>
<thead>
<tr>
<th></th>
<th>Predicted: NO</th>
<th>Predicted: YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: NO</td>
<td>TN = 9,979</td>
<td>FP = 21</td>
</tr>
<tr>
<td>Actual: YES</td>
<td>FN = 9</td>
<td>TP = 9,991</td>
</tr>
<tr>
<td></td>
<td>9,998</td>
<td>10,012</td>
</tr>
</tbody>
</table>

Table 7.5: Confusion matrix of best Discriminant Analysis ‘s’

<table>
<thead>
<tr>
<th></th>
<th>Predicted: NO</th>
<th>Predicted: YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual: NO</td>
<td>TN = 9,865</td>
<td>FP = 135</td>
</tr>
<tr>
<td>Actual: YES</td>
<td>FN = 59</td>
<td>TP = 9,941</td>
</tr>
<tr>
<td></td>
<td>9,924</td>
<td>10,076</td>
</tr>
</tbody>
</table>
Table 7.6: Testing Dataset Performance: Against Optical Images Only

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Corr</th>
<th>Stand Dev</th>
<th>Bias</th>
<th>RMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN/mks</td>
<td>0.86</td>
<td>9.40</td>
<td>-0.86</td>
<td>9.00</td>
</tr>
<tr>
<td>NN/mvks</td>
<td>0.86</td>
<td>9.37</td>
<td>-1.55</td>
<td>9.07</td>
</tr>
<tr>
<td>NN/ks</td>
<td>0.85</td>
<td>9.49</td>
<td>0.80</td>
<td>9.08</td>
</tr>
<tr>
<td>SVM/s</td>
<td>0.85</td>
<td>9.66</td>
<td>-3.24</td>
<td>9.76</td>
</tr>
<tr>
<td>SVM/ms</td>
<td>0.85</td>
<td>9.68</td>
<td>-3.28</td>
<td>9.79</td>
</tr>
<tr>
<td>SVM/mks</td>
<td>0.85</td>
<td>9.75</td>
<td>-3.32</td>
<td>9.87</td>
</tr>
<tr>
<td>NB/ms</td>
<td>0.85</td>
<td>9.75</td>
<td>-3.39</td>
<td>9.89</td>
</tr>
<tr>
<td>NB/mvs</td>
<td>0.85</td>
<td>9.85</td>
<td>-3.77</td>
<td>10.12</td>
</tr>
<tr>
<td>NB/vs</td>
<td>0.84</td>
<td>9.85</td>
<td>-3.84</td>
<td>10.15</td>
</tr>
<tr>
<td>K-NN/s</td>
<td>0.86</td>
<td>9.37</td>
<td>-1.20</td>
<td>9.01</td>
</tr>
<tr>
<td>K-NN/ks</td>
<td>0.85</td>
<td>9.83</td>
<td>-3.38</td>
<td>9.96</td>
</tr>
<tr>
<td>K-NN/mks</td>
<td>0.85</td>
<td>9.85</td>
<td>-3.40</td>
<td>9.99</td>
</tr>
<tr>
<td>DA/s</td>
<td>0.85</td>
<td>9.70</td>
<td>-3.32</td>
<td>9.82</td>
</tr>
<tr>
<td>DA/vs</td>
<td>0.84</td>
<td>9.83</td>
<td>-3.84</td>
<td>10.13</td>
</tr>
<tr>
<td>DA/vks</td>
<td>0.84</td>
<td>9.91</td>
<td>-3.85</td>
<td>10.20</td>
</tr>
</tbody>
</table>
Table 7.7: Training Dataset Performance: Against Optical and Model Metrics

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Corr</th>
<th>Stand Dev</th>
<th>Bias</th>
<th>RMSE (%)</th>
<th>Acc</th>
<th>FN</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN/mks</td>
<td>0.85</td>
<td>12.12</td>
<td>1.13</td>
<td>11.55</td>
<td>1.00</td>
<td>49</td>
<td>0.99</td>
</tr>
<tr>
<td>NN/mvks</td>
<td>0.87</td>
<td>11.20</td>
<td>-0.01</td>
<td>10.62</td>
<td>1.00</td>
<td>50</td>
<td>0.99</td>
</tr>
<tr>
<td>NN/ks</td>
<td>0.83</td>
<td>13.13</td>
<td>3.17</td>
<td>12.85</td>
<td>1.00</td>
<td>19</td>
<td>1.00</td>
</tr>
<tr>
<td>SVM/s</td>
<td>0.90</td>
<td>9.47</td>
<td>-2.78</td>
<td>9.41</td>
<td>0.99</td>
<td>76</td>
<td>0.98</td>
</tr>
<tr>
<td>SVM/ms</td>
<td>0.90</td>
<td>9.29</td>
<td>-2.98</td>
<td>9.30</td>
<td>0.99</td>
<td>77</td>
<td>0.98</td>
</tr>
<tr>
<td>SVM/mks</td>
<td>0.90</td>
<td>9.25</td>
<td>-3.11</td>
<td>9.31</td>
<td>0.99</td>
<td>58</td>
<td>0.98</td>
</tr>
<tr>
<td>NB/ms</td>
<td>0.84</td>
<td>12.23</td>
<td>-5.49</td>
<td>12.84</td>
<td>0.96</td>
<td>253</td>
<td>0.92</td>
</tr>
<tr>
<td>NB/mvs</td>
<td>0.90</td>
<td>9.23</td>
<td>-3.29</td>
<td>9.35</td>
<td>0.97</td>
<td>407</td>
<td>0.95</td>
</tr>
<tr>
<td>NB/vs</td>
<td>0.90</td>
<td>9.29</td>
<td>-3.28</td>
<td>9.40</td>
<td>0.98</td>
<td>280</td>
<td>0.96</td>
</tr>
<tr>
<td>K-NN/s</td>
<td>0.85</td>
<td>12.15</td>
<td>0.79</td>
<td>11.55</td>
<td>1.00</td>
<td>9</td>
<td>1.00</td>
</tr>
<tr>
<td>K-NN/ks</td>
<td>0.91</td>
<td>8.89</td>
<td>-3.76</td>
<td>9.23</td>
<td>1.00</td>
<td>11</td>
<td>0.99</td>
</tr>
<tr>
<td>K-NN/mks</td>
<td>0.91</td>
<td>8.91</td>
<td>-3.80</td>
<td>9.27</td>
<td>1.00</td>
<td>11</td>
<td>0.99</td>
</tr>
<tr>
<td>DA/s</td>
<td>0.89</td>
<td>9.83</td>
<td>-2.55</td>
<td>9.67</td>
<td>0.99</td>
<td>59</td>
<td>0.98</td>
</tr>
<tr>
<td>DA/vs</td>
<td>0.90</td>
<td>9.36</td>
<td>-3.20</td>
<td>9.44</td>
<td>0.98</td>
<td>303</td>
<td>0.96</td>
</tr>
<tr>
<td>DA/vks</td>
<td>0.90</td>
<td>9.21</td>
<td>-3.39</td>
<td>9.37</td>
<td>0.98</td>
<td>339</td>
<td>0.97</td>
</tr>
</tbody>
</table>
using the NN achieved an approximate value of one, with the overall highest MCC values. The NN algorithm also had the second lowest set of false negatives compared to the other algorithms. The second lowest RMSE was obtained by the KNN algorithm using the smoothness texture. Overall, the KNN had equal accuracy to that of the NN, except it only required one texture as an input. The KNN algorithm had the lowest number of false negatives compared to the other algorithms, and achieved the highest correlation values for its training datasets. The support vector machine using the smoothness achieved the 3rd lowest RMSE (9.76%), with a nearly identical RMSE for its training dataset.

The algorithms with the largest RMSEs belonged to the Naive Bayes and Discriminant analysis. These algorithms also achieved a slightly lower correlation against the testing dataset. Additionally, the Naive Bayes and discriminant analysis had much larger number of false negatives compared to the other algorithms, nearly twice the average for the discriminant analysis using the variance, kurtosis and smoothness (FN=239) and three times as large as the average for the Naive Bayes using the mean, variance and smoothness (FN=407). All algorithms were within two standard deviations from the mean for RMSEs (testing and training), except for one model which included the Naive Bayes using the mean and the smoothness for the training data.

### 7.6 Discussion

Overall, the different machine learning algorithms are comparable in terms of their testing data RMSEs and correlations. It is interesting to note that all models across the board used the smoothness to achieve high correlation results, highlighting that this particular first-order texture is useful for discriminating open water from ice in SAR images.
In terms of using second-order statistical measures for textures, we calculated gray-level co-occurrence matrices (GLCMs), and extracted nine textures (see table 7.8), since these have been widely used in SAR image classification (Soh and Tsatsoulis, 1999). However, we found that the computation of these textures were too time consuming and unnecessary for the discrimination between ice and open water. These textures also seemed to report similar values for different extrema, where extremely dark values such as those reported by open water and very bright values such as ridges or recently re-frozen ice surfaces, obtained overlapping values, which would hinder the machine learning classifiers. We believe that GLCMs are beneficial for the classification of multiple ice types, and not necessary for binary classification.

Classifying open water from SAR images during the summer can be tricky, because during the melt season, snow melt exposes bare ice to shortwave radiation, leading to build-up of melt water on the ice surface. If enough melt water accumulates, the SAR backscatter
of these deep melt ponds can be identical to that of open water. Using our methods, we were able to resolve most of the melt ponds and discriminate them from open water, however some contamination occurred.

### 7.7 Chapter 7 Summary

This study focused on using several machine learning algorithms to classify open water fraction from high resolution Stripmap TerraSAR-X images in the Beaufort sea during boreal summer of 2014. This task can be difficult during the summer months, where melt-onset can alter the radar backscatter and melt ponds potentially contaminate true OWF estimates. The machine learning classifiers investigated included (1) neural networks, (2) linear support vector machines, (3) Naive Bayes, (4) K-nearest neighbor and a (5) discrimin-
inant analysis algorithm. Each algorithm was trained using an identical training dataset consisting of the image intensity itself, and the combinable addition of four “modified” first-order textures including the mean, variance, kurtosis and smoothness. We trained 80 different models and inspected model metrics from a confusion matrix, and compared classification results against nearly-coincident panchromatic optical images that were manually thresholded for open water values as ground truth.

We found that the models had almost identical results in terms of accuracy. However, the fastest model to both train and classify was the Naive Bayes algorithm. Using a combination of parallel computing and block-processing, classification of full-sized high resolution TSX images can be performed at a relatively high speed. The computation time to classify a full-sized SAR image using the Naive Bayes algorithm was roughly 7-10 minutes, while the Neural Networks and Support Vector Machines approximated 30 minutes. Faster computation can be very practical for users on vessels wishing to have “on-the-fly” and accurate method of calculating how much ice/water is in the general vicinity for navigational purposes.

For the TSX dataset, the methods discussed in this chapter were coincidentally, limited to images taken under low wind conditions in Stripmap mode. If the readers wish to follow a similar classification set-up, we suggest incorporating separate models for images whose open-water backscatter is higher than those mentioned here. If other SAR imaging modes are desired, such as ScanSAR, the reader may also wish to incorporate additional training features such as incidence angle dependency into their models. In the next chapter, we discuss how using different models can help with varying wind conditions when we classify a much larger dataset from TSX and Radarsat-2.
CHAPTER 8

Open Water Fraction: Analysis

Based on the results from chapter 7 for open OWF methods, we decided the Naive Bayes method would be used to classify the entire SAR image dataset. Although the Naive Bayes method is much simpler than Neural Networks and Support Vector Machines, this method still achieved comparable accuracies at a fraction of the computation time.

Because the TSX and RS2 images are from different bands (X and C, respectively), they had to be treated separately. In addition, since the SAR image scenes spanned different seasons in the Arctic MIZ, several Naive Bayes models had to be trained in order to capture the changing ice conditions. For the TSX Stripmap images, seven machine learning models using the Naive Bayes method were trained to recognize areas such as high wind scenes and melting/refreezing conditions, to name a few. Model assignments for image classification included a combination of inspecting image scene dates, comparing similar changes of extrema in backscatter intensity and by eye inspection. For the Radarsat-2 dataset, four models were trained. A total of 422 SAR images were classified into ice/open water, including 251 from TSX, and 152 images from RS2.
8.1 Evolution of Open Water Fraction at MIZ Clusters from TSX

Due to TSX’s high resolution, small areas of surface melt can be seen in certain SAR images during the summer. The regime shift from freezing to surface melting conditions can be investigated through the change in backscatter due to the presence of surface water. As validation aids, we use optical images from the LIDPs to obtain further information on surface melt, as well as information from ice mass-balance (IMB) buoys, wave buoy cameras, and autonomous weather stations that are spatially and temporally collocated with TSX images.

As mentioned in chapter 4, the melt process can be categorized in two regimes: Pendular and funicular. The pendular regime consists of water held in the interstices of the snow pack where a rapid rise in backscatter from first-year ice sites and a decrease in scattering on multi-year sites have been witnessed (Barber and Yackel, 1999). In the funicular regime, snow cover begins to drain and a slushy water mixture pools at the base of the ice pack. We can investigate the pre-melt regime to pendular by inspecting the change in image backscatter through time.

We inspected the intensity of the backscatter from 338 TSX images, which ranged from mid-March to early October, 2014. Images were not taken consistently per month; they were taken at different times during the day, and multiple images were sometimes taken during the same day. The PDF of the intensity of the backscatter is shown in figure 8.1. Here, we can see that higher densities lie between values of 130-190 in gray scale, which is attributed to wet ice surfaces (in the lower end) to drier and or ‘chunkier’ sea ice surface
Figure 8.1: PDF of TerraSAR-X mean image intensity values from March-October, 2014.

types. The PDF highlights that very few images obtain, on average, low backscatter values such as open water.

The change in image backscatter throughout the different months can be measured using image statistics. We computed the average mean and median intensities for each of the images, and then averaged these values by month, looking only at June-September months. The number of data points totaled 92. We then built a linear regression model, using the averaged image intensity as the dependent variable, and the averaged median and month as the independent variables as:

\[
\text{Averaged Intensity} = 1 + \text{Averaged Median Intensity} \times \text{Month} \tag{8.1}
\]

where the month is entered into the equation as June = 6, July = 7, etc. The fitted regression lines by month are shown in Figure 8.2.

With the month as the nominal variable, the model resulted in an \( R^2 = 0.93 \), and \( p << 0.05 \). This means that the variation in the averaged image intensity is reduced by 94% when
you consider the averaged median of the intensity, the month, and their interactions. We ran an Analysis of Variance (ANOVA) test to test for significance between the slopes, whose results are shown in figure 8.1, and the median of the image \((p < 0.05)\) and the interaction between the median of the image and month \((p < 0.05)\) both provided significance to the data, suggesting evidence that these two variables were not equal and therefore, changed over time, which is expected.
Figure 8.3: TerraSAR-X (top images) and LIDPs (bottom images) showing the progression of surface melt and its consequent on visual backscatter.

From the statistics tests, we then decided to look for specific drops of intensity of the images through time, specifically looking at the median intensity value. A peculiar drop in both mean intensity and median intensity occurred during June 14th, and upon closer inspection, we decided that this date delineated surface melt/snow moisture in the SAR images. For example, figure 8.3 shows the visual progression of surface melt on SAR images, and nearly-coincident LIDPs. As meltwater collects on the low levels of topography on the ice, SAR can detect the melt ponds since the presence of meltwater decreases the penetration depth of the radar signal.
From the IMB dataset however, we were able to see phase changes in the temperature of the ice, aiding in the detection of basal melt. Looking at the data from IMB 16, which is shown in figure 8.4, and other IMB data (not shown), we were able to estimate that basal melt occurred during June 6-7. This was attributed to an increase in ice/ocean interface temperatures, sustained above 0°C, at the bottom of the ice. This gave us a clue a further validation that surface melt in the SAR images did occur during the first two weeks of June.

Out of the 338 TSX images we obtained, 251 were successfully classified using the Naive Bayes method. The images that failed classification were due to a mixture of TSX improper image model assignments and extreme high wind activity, which confused the classification models. The failures were identified qualitatively, by human visualization.

We can investigate the temporal evolution of OWF at the MIZ clusters, which can give insight as to whether there is a spatial relationship between OWF in the MIZ. Therefore, we collocated the IMB buoys which where both inside the vicinity of the TSX image, and within one day from image capture. A total of 198 TSX images collocated with 23/25
Figure 8.5: Time series of open water fraction from TerraSAR-X, at different MIZ cluster arrays.

IMBs. Specifically, IMB numbers 2-23 were collocated with the TSX OWF dataset. The evolution of OWF from TSX from clusters 1-2 is shown in figure 8.5 and figures 8.6-8.8 show the positions of the clusters from IMBs, AWSs, and wave buoys.

The amount of collocations between clusters were fairly similar, with clusters reporting 30, 40, 46, 50, and 31 for clusters 1, 2, 3, 4, 5, respectively. It is important to note that, although clusters 1-4 obtained collocations beginning from early May, cluster 5 had collocations with TSX beginning in mid-August, since instruments in this cluster were deployed during this time.

For cluster 1, OWF remained below the 5% range until the first week of July, where most OWF values began to consistently stay above 10%. Later in July, near July 22, the OWF increased to 22%, and then spiked up to 86% during the middle of August. On August 28th, the OWF reached almost 100%.

Cluster 2 shows a similar pattern to that of cluster 1, however it experienced a higher OWF about a week earlier. This increase was not expected, given that cluster 2 was north
Figure 8.6: MIZ wave buoy, AWS and IMB positions at their clusters, during May 8 through June 17, 2014.
Figure 8.7: MIZ wave buoy, AWS and IMB positions at their clusters, during July 1 through August 11, 2014.
Figure 8.8: MIZ wave buoy, AWS and IMB positions at their clusters, during August 25 through October 5, 2014.
of cluster 1. Overall, cluster 2’s OWF increased until the early week of August, where OWF reduced by 10% for a few days. The OWF then continued to increase well into the end of September.

At cluster 3, the same pattern emerges, where OWF did not increase much until the second week of July, where OWF values reached 13%. OWF consistently increased after this, until the second week of August, when OWF spiked to an average of 40%. OWF values increased and decreased after this time, however with an overall increasing trend. During the first week of September, we see a dramatic decrease from 65% to about 26% OWF, and then random increases and decreases. This type of dramatic increase/decrease in OWF can be investigated through satellite images. For example, during September 13, the OWF was 26%, but two days later, on September 15, the OWF increased to 78%. This was attributed to ice divergence, which is shown in figure 8.9. We can also see that the ice type during this stage was pancake-like, which is known to be more mobile.

Cluster 4 had a similar trend of OWF increase from early May until the first week of August, where OWF values were consistently lower than those of clusters 1-3. This is interesting, given that, during the first and second week of August, cluster 4 was situated on the same latitude as clusters 1-3. In fact, cluster 4 migrated even further south than any other cluster in early October, however cluster 4 never achieved OWFs higher than 70%. We inspected the TSX images and found that the ice type in this cluster was mostly in-tact, with larger floes which hindered ice divergence.

At cluster 5, the largest increase in OWF is seen on September 1st, which is fairly close to the values of OWF from cluster 4 during this time. However, after this first week of September, OWF values kept consistently below 10%, until we see evidence of re-freezing on the first week of October. The spike in OWF between late August is attributed to differ-
Figure 8.9: TSX images from cluster 3 on September 13, 2014 with 78% OWF (left) and on September 15, 2014 with 26% OWF (right).

ing image scene sizes, which captured different variability. For example, on August 29, the OWF is approximately 3% where the area scanned was approximately 90 × 30km, but on September 2, the OWF is reported as 32% in an image with size 338 × 30km. Such large image scenes made up less that 2% of our dataset.

The major deviations of OWF between the clusters, occurred during the first week of August. Cluster 1 clearly obtained the highest OWF estimates until end of August, while clusters 2 and 3 followed one another. However, during the second week of September, cluster 3 gained higher OWF estimates. Cluster 5 had consistently lower OWF than any other cluster, with the exception in late September, when cluster 4 obtained the lowest OWF values. In order to examine these differences further, we looked to physical data from the in-situ instruments at the cluster sites. Figure 8.10 and 8.11 show time series of daily averaged air temperatures taken from an IMB and AWS, respectively, at their clusters.
Figure 8.10: Time series of daily averaged temperature from IMBs at specific clusters.

Figure 8.11: Time series of daily averaged temperature from AWSs at clusters 1,3,4,5.
The overall trend is clear, as air temperature increased, open water fraction also increased at the cluster sites. However, inspecting the air temperatures further, we see that clusters 1-4 have almost identical values from the second week of March until early July, where cluster 4’s air temperature began to consistently decrease. Clusters 1-3 had similar air temperatures between the second week of July until late July, although during this period the clusters began to experience shifts in OWF values. It is interesting to note that, during late July, both the IMB and AWS reported air temperatures above 0°C for two weeks, however we were not able to collocate TSX images with instruments at these cites during this time to investigate this phenomena further.

We also extracted average daily wind speeds from AWSs, for clusters 1, 3, 4 and 5, shown in figure 8.12. Overall, the wind speeds at the different clusters are fairly similar, however, cluster 1 and 4 seem to report slightly higher wind speeds than clusters 3 and 5. Wind speeds do not seem to correlate with OWF values in any significant way, except for cluster 5 during the end of August, which seems to have increases in OWF with increasing wind speeds.

8.2 Evolution of Open Water Fraction from RS2

In total, we were able to classify 152 images out of 262, due to the large presence of high wind activity in many of the images, especially those in the later summer, when the open water fraction was highest.

Inspecting figure 8.13, we obtain a slightly ‘noisy’ time series. This is because, the RS2 images were taken in different places near the Beaufort MIZ, including near the Chukchi Sea, De Long Strait, and near the Canadian Archipelago.
Figure 8.12: Time series of daily averaged wind speeds from AWSs at clusters 1,3,4,5.

Figure 8.13: Radarsat-2 Open Water Fraction Evolution
8.2.1 Comparison of Open Water Fraction: TSX vs. RS2

Overall, given that the TSX and RS2 images were not taken at the same place and time, their PDFs, which are shown in figures 8.14 and 8.15, have strong similarities.

The PDFs of RS2 and TSX are fairly similar. Both datasets have a peak in OWF between 0-10%, and smaller likelihoods to below these values. Generally, RS2 reports higher OWF values than TSX, which makes sense, since RS2 captures larger image scenes and during the summer, we know that the MIZ is exposed to large areas of open water.

The similarities in the PDFs of OWF hint that the Beaufort Sea acts as a system, where remarkable similarities occur between smaller and large regions. This could indicate that changes in the Arctic sea ice can be witnessed in both small and large scales. In order to further highlight this idea, we can inspect the TSX and RS2 cumulative probability function, which is shown in figure 8.16.

From the cumulative probability function, we can see that TSX has a higher probability when the open water fractions lie between 5-35%. The opposite is true when TSX
Figure 8.15: Radarsat-2 Open Water Fraction PDF

Figure 8.16: TerraSAR-X and Radarsat-2 OWF cumulative density function.
obtains OWF values between 50-85%. Because the image scenes of RS2 are significantly larger than TSX, a one to one OWF comparison between the two datasets is not possible. However, we can make qualitative comparisons between collocated RS2 sub-image scenes against TSX. For example, we collocated sixteen RS2 images that were taken within 24 hours of TSX. Figure 8.17 shows an example of a collocation. In every case of collocation, RS2 significantly reported lower OWF values.

8.3 Chapter 8 Summary

We took the best classification method, the Naive Bayes algorithm, and applied it to classify a larger TSX and RS2 dataset. We collocated the classified TSX images against IMB instruments and reported OWF for clusters 1-5. We learned that all clusters had similar OWF values from May, until mid-July, when OWF began steadily increasing. Cluster 1 obtained the highest OWF values, while Cluster 5 obtained the lowest. We saw a correlation between increasing daily averaged surface temperature and OWF. Wind speeds reported by
the AWS did not have any significant correlation against TSX OWF values. RS2 OWF values throughout time had a more sporadic trend over time, but this was not attributed to classification error, instead, it was due to the sensor taking images at different parts of the Beaufort Sea. Finally, we saw that, overall, the OWF values of TSX and RS2 followed a similar trend. In the next chapter, we move onto a 3-classification problem to extract melt ponds from the LIDP optical dataset.
CHAPTER 9

Melt Pond Detection Algorithm

Much work has been published on determining changes in summer ice albedo and morphological properties of melt ponds such as depth, shape and distribution using in-situ measurements and satellite-based sensors (Barber and Yackel, 1999, Derksen et al., 1997, Eicken et al., 2004, Fetterer and Untersteiner, 1998, Grenfell and Perovich, 2004, Morassutti and LeDrew, 1996, Yackel and Barber, 2000). Although these studies have provided much pioneering work in this area, there still lacks sufficient spatial and temporal scales. Additionally, melt ponds are opaque to contaminate passive sensors such as SSM/I and AMSR-E, and have been reported to underestimate sea ice concentration during the summer (Comiso and Kwok, 1996).

In terms of absorbed shortwave radiation, Polashenski et al. (2012) showed that the observed spatially averaged albedo ranged from 0.25-0.60, translating to a range in absorbed shortwave radiation flux of around 90 Wm$^{-2}$, where they calculated this much flux was enough to melt at least 2.6 cm of ice per day. This means moderate changes in the duration of peak pond coverage could have significant impacts on the ice mass-balance.

Climate models have large uncertainties due to melt pond behavior (Flocco et al., 2012, Holland et al., 2012, Hunke et al., 2013, Webster et al., 2015) which is due to insufficient
physics knowledge about melt pond evolution, and the inability of climate models to resolve such high-resolution features. Melt pond knowledge has been traditionally obtained from land-fast ice observations (Eicken et al., 2004, Perovich et al., 2002a,b, 2001, POLASHENSKI et al., 2012, Yackel and Barber, 2000) due to the relative ease of revisiting the same study area. This is not the case for the MIZ, where the pack ice is constantly moving and the repeatability of visiting the same sites are both costly and dangerous for scientific crew members. With the current release of the LIDP optical dataset, we now have the opportunity to explore melt pond behavior in the moving ice packs.

A study published by Webster et al. (2015) looked into the melt pond evolution for a specific site in the Arctic during the summer of 2011; 18 images were analyzed in tandem with nearby in-situ observations from the SHEBA experiment. This study made use of masking out the open ocean along with various thresholding techniques in order to separate melt ponds from the nearby sea ice, which proved to be highly successful, however, this method was also time-consuming. For our work, we propose to make use of SVMs in order to look at a much larger dataset of LIDPs, in order to decrease computation time and still retain accurate results.

9.1 Methods

First, we collocated ice mass-balance buoys and autonomous weather stations from clusters 1-5, as seen in figure 5.1 that were taken within a day from image capture. This time was chosen in order to accurate IMB positions within the optical image frames. Because we did not have ground validation of the sites, such as photographs from helicopters, we decided to limit our classification image study area to 1 km$^2$ surrounding the buoys. Any
image with visible cloud cover was ruled out. A total of 34 buoy-centered images were extracted from original full-sized data.

Because sea ice changes phase with melt and freeze onsets, the signatures of the ice in the optical images will also change. To deal with this obstacle, we decided to create seasonally-evolving SVM models. The LIDP dataset is radiometrically inconsistent, and the fact that the amount of melt ponds present in an image increases the standard deviation of the scene (Webster et al., 2015) means that further work to separate images into respective groups must be performed. These steps include: Seasonal partitioning, Minima Sub-Partitioning, training and classifying the SVMs, and finally, post-classification processing.

9.2 Seasonal Partitioning

Separating images for classification via their respective models initially begins by analyzing the air temperature measured from IMB-2 in cluster 2, shown in figure 9.1. Pre-melt conditions were visible from time of deployment in early March to early June. From June to end of August, the buoy reported consistent temperatures above 0°C, and temperatures from early September on showed consistently to be at 0°C and/or below. Any images that were taken during these times were grouped to be classified into their respective models.

9.3 Minima Sub-Partitioning

Once our images were sorted into three main groups, the images were further partitioned. We calculated the image’s first minima and maxima by computing the derivative of the global minimum and maximum of the intensity distribution. Figure 9.2 shows the
histogram of an image which belongs to a 'melt-onset' group. Images that had a similar
distribution based on both their lowest minimum followed by their immediate first max-
imum (gray-scale values) got placed in sub-group models. We calculated that an image
would belong to a sub-group if both their first extrema fell within 1.5 standard deviations,
creating a total of nine sub-groups for images to be categorized in.

9.4 Training the LSVMs and Classifying the Images

For our nine sub-groups, we chose to train linear support vector machines (LSVMs)
to separate the LIDPs into three classes: (1) open water, (2) ice, and (3) melt ponds. The
LSVM implements a "winner-takes-all" multi-class kernel function approach. This ap-
proach acts as a binary classifier, separating data according to its weights, while "holding"
the two other classes as unwanted data.

The number of training images varied per model; some included only one training
image, and some included two, depending on how many images belonged to that specific
sub-group. However, all trained models contained 9000 training pixels; 33% belonging to open water, ice, and melt ponds, each. We also included three 'texture' features which were helpful in separating melt ponds from the other classes. These included two image filters, (1) standard deviation filter, (2) range filter and one first-order statistical texture which calculated the entropy of the image (see chapter 7).

The standard deviation filter returns an array where each output pixel contains the standard deviation of a 3-by-3 neighborhood around the corresponding pixel in the input image. The range filter returns an array where each output pixel contains the range value (maximum value - minimum value) of a 3-by-3 neighborhood around the corresponding pixel in the input image. Finally, the entropy of the image is calculated from a the modified first-order histogram method.

Upon inspecting figure 9.3, the standard deviation and range filters look identical, however, they have very different values: for this example, the standard deviation filter reported a mean of 20, with a maximum of 86. The range filter reported a mean of 50, with a maximum of 57. The entropy had a mean of 5, with a maximum of 6. All textures reported a
minimum of 0. Overall, we chose these textures for two reasons. First, they were computationally inexpensive, and the outputs did a good job in delineating edges and highlighting melt ponds.

Once the training set was complete, we classified 34 buoy-centered images, including the images trained on. The classification process included the input of the images themselves, and the aforementioned textures and then fed into their respective training models.

### 9.4.1 Post-Classification Processing

The results from the LSVM classifications left erroneous melt pond pixels. Areas of ice that were directly bordering water were classified as melt ponds, which we refer to as “border errors”, and big melt ponds were often classified as open water in the center, which we refer to as “center errors”. A correction algorithm was designed to rectify this.

The algorithm was designed using a series of expansions and contractions of certain elements (water, melt pond, or ice). Figure 9.4 indicates a visual representation of the steps described.
Figure 9.4: Steps taken to correct for erroneous melt pond classification in an LIDP.
First, in step 1, we corrected for center errors. For any pond pixel that directly touched a water pixel, it was expanded so that the erroneous water pixel became a pond pixel. This was iterated over five sequences to create a maximum five-pixel encroachment into the water. Almost all of the center errors were less than 10 pixels wide, and this sufficiently eliminated most center errors without mislabeling small areas of open water as a melt ponds. However, for border errors, this actually makes the pond borders thicker than they once were. The locations of melt pond pixels were saved in a separate variable.

Next, in step 2, we expanded the ice. For each melt pond pixel in the original image that was next to an ice pixel, it was changed to an ice pixel in the correction image. This effectively closed off very thin melt ponds that ‘drained’ out into the ocean (lateral drainage). This reduced the size of true melt ponds as well. This was saved in the same variable as the result of step 1, so that the most prominent melt pond pixels left in the image were four pixel extensions around the borders. We call this the “correction image”.

In step 3, the water in the correction image was expanded into the melt ponds, essentially eliminating extra border signatures from step 1. This also eliminated any rogue melt pond pixels that may have ended up in the middle of the open water.

A new correction image was made from these parts. The locations of ice pixels were taken from step 2, the melt ponds from step 1 were overlaid on top of this. In addition, water pixels from step three were placed on top as well. We now had a image similar to the original, however with reduced center errors.

In step 4, we executed one last expansion for the original border errors. One more expansion of water into the melt, three pixels thick, covered the last border. After that, a one-pixel expansion of water into the ice cleaned up extra ice on the borders of large ice floes from the previous 1-pixel expansion of the ice in step 2.
The correction is particularly challenging in images with a great deal of broken ice, such as the one in figure 9.4. Small cracks in the ice are often mislabeled as melt ponds at a cost for the correction of center errors and border errors. Center errors and border errors are a much more common issue than the small cracks in the ice, so it is overall a good option to run the correction on all of the images in order to avoid an overestimation of melt.

9.5 Chapter 9 Summary

In this chapter we presented a prototype algorithm using Machines LSVMs designed to quantify the evolution of melt pond fraction from a recently government-declassified high-resolution panchromatic optical dataset, in an area where several in-situ instruments were deployed by the British Antarctic Survey in joint with the Marginal Ice Zone Program, from April-September, 2014. We learned that, because of radiometric inconsistency, the LSVMs had to be sub-partitioned into separate models using the standard deviations of the intensity image’s extrema. The results from the LSVM classifications had erroneous pixels at the ice edges, which we reduced through a dilation/erosion algorithm. In the next chapter, we explore the melt pond evolution and pond statistics in order to both validate our work, and investigate possible links between melt pond shape parameters.
CHAPTER 10

Melt Pond Analysis

Given the outputs of melt ponds from our scheme discussed in chapter 9, we proceeded to further process the melt pond class results in order to get the most accurate melt pond information. The reasoning behind this comes from possible misclassified melt pond areas, such as brash ice, lead smoke, fog, and thin clouds, which have very similar histogram attributes to those of melt ponds (Webster et al., 2015). We also explored melt pond statistics such as melt pond fraction (MPF), pond number density (PND), as well as shape parameters such as mean and median pond area, pond perimeter, pond major axis length, and pond eccentricity, in order to compare our results with previous research.

10.1 Methods

We extracted the information from class melt-pond from the 34 sub-LIDP images which were classified in chapter 9. From these images, we combined classes ice and open water as one class, and melt ponds as another to generate a binary image. From this binary image, we calculated a label matrix ‘L’ that contained labels for 8-connected objects in this binary image.
Table 10.1: Average Shape Parameters and Statistics from Melt Ponds

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Pond Area</td>
<td></td>
</tr>
<tr>
<td>Minimum Pond Area</td>
<td></td>
</tr>
<tr>
<td>Mode of Pond Area</td>
<td></td>
</tr>
<tr>
<td>Average Pond Area</td>
<td></td>
</tr>
<tr>
<td>Median Pond Area</td>
<td></td>
</tr>
<tr>
<td>Average Pond Perimeter</td>
<td></td>
</tr>
<tr>
<td>Average Major Axis Length</td>
<td></td>
</tr>
<tr>
<td>Average Eccentricity</td>
<td></td>
</tr>
<tr>
<td>Number of Ponds</td>
<td></td>
</tr>
<tr>
<td>Pond Number Density</td>
<td></td>
</tr>
<tr>
<td>Melt Pond Fraction</td>
<td></td>
</tr>
</tbody>
</table>

From the matrix L, we obtained a set of properties for each eight-connected component using the Matlab function ‘regionprops’. This function returned measurements for a set of properties for each 8-connected component (object) in the binary image. The properties we extracted from the melt pond class included pond area, pond perimeter, pond major axis length, and pond eccentricity.

In order to get a general sense of these shape parameters, we decided to average these parameter values from each image, instead of analyzing each component inside every image. From these averaged values we obtained the following averaged shape parameter information, as well as overall melt pond statistics, listed in table 10.1.

The pond perimeter was calculated as the distance around the boundary of the pond region. The function regionprops calculates the perimeter as the distance between each adjoining pair of pixels around the border of the region. The major axis length (in pixels) is the major axis of the ellipse that has the same normalized second central moments as the region. Using regionprops, the eccentricity returns a scalar that specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio
of the distance between the foci of the ellipse and its major axis length, where the value is between 0 and 1.

The pond number density was calculated as the number of ponds divided by the area of ice and ponds, with units km$^{-2}$, thus eliminating effects due to changes in ice concentration (Perovich et al., 2002b). MPF also eliminates changes in ice concentration and is calculated as the number of melt pond pixels divided by the area of ice pixels in the image.

10.2 Results

We compare our results against previous research from (Perovich et al., 2002b) which report similar melt-pond statistics and shape parameters on Arctic sea ice. Figure 10.1 shows a few pond statistics for all 34 images from early June, until late September.
In Perovich et al. (2002b), statistical descriptions were given for melt ponds for 10, 22, and 30 June, and 7 August, 1998 in a 100km$^2$ area. During early June, ponds began to form, with a MPF of less than 2%, which is similar to our results of 1% for early June, 2014. During June 22, Perovich reports an initial spike of 20% MPF, while we report our first initial spike of 68% on July 1, 2014.

Why such a large pond fraction? We can use wave buoy cameras to validate the progression of melt pond behavior, specifically surface ponding, maximum ponding, and drainage, respectively. Figure 10.2 shows images taken from a camera mounted on a wave buoy on cluster 3, where during June 25, 2014, maximum ponding is apparent.

In terms of pond number density, we also had similar results. Perovich reports PND of approximately 750km$^{-2}$ in early June, while we report 400km$^2$. Perovich’s PND spike up to 2,250km$^{-2}$ on June 22, while we report 421km$^{-2}$. However, on July 30, 2014 we report a PND of 2,000km$^{-2}$, and on July 31 the PND goes up to 3,144km$^{-2}$, which is very similar to Perovich’s PND values for this time, which was approximately 3,750km$^{-2}$.

Finally, for pond area, Perovich reports both mean and median area: In early June, the mean pond area was 15m$^2$ with median 6km$^2$. During this time, we report a mean pond area of 29m$^2$ and a median pond area of 11m$^2$. The largest pond areas reported by Perovich
fall on June 22 and August 7, with mean of 60 m². For these dates, we observe pond mean areas of 1000 m² (entire image is flooded) on July 1, and 200 m² during August 8th.

### 10.3 Melt Pond Regression Analysis

In order to investigate possible correlation(s) between the pond statistics and shape parameters at different MIZ clusters, we constructed and analyzed a linear regression model, using a stepwise method with interaction effects. The model uses forward and backward stepwise regression to determine a final model. At each step, the function searches for terms to add or remove from the model based on a p-value from an F-test of the change in the sum of squared error by adding or removing the term.

Through trial an error, the variables which gave the best model results consisted of MPF, date of the image, MIZ cluster, pond number density and average pond area. The final model predicts the average pond area as a linear formula with eight terms and four predictors

\[
\text{Average Pond Area} \equiv 1 + \text{Melt Pond Fraction} \times \text{Pond Number Density} + \text{date} \times \text{Cluster} + \text{Cluster} \times \text{Pond Number Density} \quad (10.1)
\]

The adjusted linear regression model suggests a relationship to calculate average pond area, highlighted by equation 10.1, whose estimated coefficients are shown in table 10.2. The model achieved an \( R^2 = 0.96 \), with a \( p \)-value of \( p << 0.05 \), which is highly significant. For example, the linear model could calculate the average pond area using the predictors, as shown in figure 10.3. In this case, the model uses the main effects for all predictor variables. The green line shows the change in the response variable as a function of the
Table 10.2: Coefficients of linear melt pond model to predict average pond area.

<table>
<thead>
<tr>
<th>Estimated Coefficients</th>
<th>Estimate</th>
<th>SE</th>
<th>T-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.59e+06</td>
<td>1.38e+06</td>
<td>6.22</td>
<td>p«0.05</td>
</tr>
<tr>
<td>MPF</td>
<td>3025</td>
<td>145.45</td>
<td>20.80</td>
<td>p«0.05</td>
</tr>
<tr>
<td>DateNum</td>
<td>-11.67</td>
<td>1.87</td>
<td>-6.22</td>
<td>p«0.05</td>
</tr>
<tr>
<td>Cluster</td>
<td>-1.69e+06</td>
<td>4.11e+05</td>
<td>-4.12</td>
<td>p=0.0003</td>
</tr>
<tr>
<td>PND</td>
<td>0.411</td>
<td>0.05</td>
<td>7.56</td>
<td>p«0.05</td>
</tr>
<tr>
<td>MPF:PND</td>
<td>-1.17</td>
<td>0.08</td>
<td>-13.14</td>
<td>p«0.05</td>
</tr>
<tr>
<td>DateNum:Cluster</td>
<td>2.30</td>
<td>0.55</td>
<td>4.12</td>
<td>p=0.0003</td>
</tr>
<tr>
<td>Cluster:PND</td>
<td>-0.07</td>
<td>0.01</td>
<td>-6.43</td>
<td>p«0.05</td>
</tr>
</tbody>
</table>

predictor variable, when all other predictor variables are held constant. The red-dashed curves in each panel show the 95% confidence bounds for the predicted response values. The horizontal blue line in each panel shows the predicted response for the specific value of the predictor variable corresponding to the vertical dashed blue line.

For example, the predicted value of the average pond area is 259.12 m, when we input an MPF of 0.34, at date number 7355812.5, in cluster 3, with pond number density 1957.46. The values in the square brackets [173.53,344.82] show the lower and upper limits of a 95% confidence interval for the estimated response.
10.4 Melt Pond Detection Using SAR

An accurate method to map melt ponds with coarse-resolution SAR images has not yet been achieved, due to the required spatial resolution for melt pond detection, and the inherent speckle noise of this particular sensor (Kim et al., 2013). However, with TSX, we are able to achieve much higher resolution, thus, opening an opportunity to explore such high resolution features. In this section, we briefly introduce a possible method to extract melt ponds from SAR.

We used image co-registration software from ENVI, a GIS and remote sensing program, and took 16 nearly-identical surface features as reference points between a TSX and an LIDP optical image. The resulting co-registration is shown in figure 10.4.

To classify the melt ponds for the TSX image, we used the same methods mentioned in chapter 7, using the image intensity itself and the modified-mean as feature inputs, however this time, we chose to train pixels from melt ponds as a third class. For the LIDP image, we followed the same methods described in chapter 9. Figure 10.5 shows the TSX and LIDP image co-registration, and the respective melt-pond classifications.
Figure 10.4: TSX/LIDP optical co-registration: TSX taken on (left) and LIDP (right). The areas of interest cover approximately 1 km², and the images were taken approximately five hours apart.

Figure 10.5: TSX (left) and LIDP (right) image co-registration.
Upon further inspection of figure 10.5, using the Naive Bayes classifier on the TSX did not yield a 3-class separation: The melt ponds on the TSX are classified in the same category as open water. However, the LIDP image using the SVM was successful. We thus proceeded to use a SVM classifier on the TSX, however, this yielded identical results. We believe that the misclassification was due to a number of reasons, however, the most important reason was that the melt ponds on the SAR image were very advanced, and the signatures of these melt ponds were identical to that of open water. Therefore, we decided to use post-classification techniques to separate the melt ponds from open water on the SAR images.

The properties we extracted included parameters mentioned in the previous sections with the addition of the Euler number, which is the number of objects in the region minus the number of holes in those objects. We found that we could separate the ponds in the SAR image by setting a threshold of Euler number less than 5, with a minor axis length of greater than 70. Figure 10.6 shows the progression of obtaining the final melt ponds from SAR.

To get a sense of how this compares to the optical dataset, we show the melt pond classification of TSX and the optical side by side. In this example, TSX reported 15% MPF, while the optical image reported 21%.

### 10.5 Chapter 10 Summary

This chapter explored the results of melt pond statistics and shape parameters. We compared MPF, MND and mean and median pond area to a study conducted by Perovich et al. (2002b), which proved to have similar results as ours, even though this study was conducted
Figure 10.6: TSX with: de-speckled image (a), Naive Bayes multiclass classification output (b), calculation of class open water centroids (c), exploiting open-water class shape parameter Euler Number and minor axis length (d), final 3-class partitioning of ice as white pixels, light blue as open water and dark blue as melt ponds (e).
Figure 10.7: TSX (left) and LIDP (right) image co-registration.
nearly twenty years ago, with different aerial images, and the numbers reported were from a much larger area. We found that our dataset experienced higher MPFs and larger over-all pond areas. We also introduced a linear regression model that can potentially be used to estimate average pond area by ingesting several melt pond statistics and shape parameters. Finally, we showed that TSX has the potential to capture melt ponds as a separate class, if post-classification shape parameters are exploited such as the Euler number, and the minor axis length.
CHAPTER 11

Conclusion

In this work, we explored several machine learning algorithms to classify satellite images from SAR and optical sensors. These included TerraSAR-X and Radarsat-2 (SAR), and high resolution panchromatic images (optical), which were taken in the Beaufort MIZ region. The classification of images included categories of open water and ice from SAR, and open water, ice and melt ponds from the optical. The highlight of this research centered on fast and accurate detection of open water and ice, catered to methods for ship navigation, applications on large image datasets, and near real-time ice forecast modeling.

For detecting OWF in SAR images, we found that using the Naive Bayes classifier alone with a modified first-order mean texture as a feature input, was enough to get reasonably accurate results, at a fraction of computational time compared to more sophisticated methods such as neural networks and support vector machines. The computation of the modified-mean was a new method to obtain interesting features on smoothed SAR data; more robust classification criteria was derived from different combinations of “dark” and “bright” in the intensity and mean image. This non-standard way of computing a "modified" mean and other "modified" texture properties is quite nonlinear in the sense that it can make some features darker and other features brighter in a complicated way, depend-
ing on the overall statistics of the image. This may lead to strange results in some cases, but for our purpose of discriminating between open water and ice, with special attention to classifying ponds as ice, the proposed computation method has been found to produce results that lead to a clearly more successful classification than the use of conventionally computed texture properties.

For our initial findings in chapter 7, the methods were coincidentally, limited to images taken under low wind conditions in Stripmap mode for TSX. In chapter 8 however, we had a much larger dataset for both TSX and RS2, therefore, we proposed to train different models that handled relatively windy conditions. Model assignments for image classification included a combination of inspecting image scene dates, comparing similar changes of extrema in backscatter intensity and by eye inspection.

The information derived from the classified SAR images were investigated against in-situ data from ice mass-balance buoys, autonomous weather stations and buoy cameras which were deployed at different cluster sites by the MIZ Program, during spring of 2014. We saw a correlation between OWF and daily averaged surface temperatures. The overall trend is clear: As air temperature increased, open water fraction also increased at the cluster sites. We also saw that the clusters sites did not remain stationary, instead, they were very dynamic, especially at the end of August, when OWF fraction values began to steadily increase, making the ice more mobile.

To detect melt ponds from LIDPs, we used LSVMs, implementing the image intensity itself and three image ‘texture’ features that included a standard deviation filter, a range filter and a modified first-order statistical texture which calculated the entropy of the image, using our modified-intensity histogram scheme. The optical dataset had radiometric inconsistencies between images, sun angle variation, and predominant cloud cover. Al-
though careful consideration was taken to mask out clouds, there was still contamination from thin clouds and fog, which have nearly identical intensity and textural signatures as melt ponds and ice. Additionally, the LSVMs had some erroneous outputs. Areas of ice that were directly bordering water were classified as melt ponds, which we referred to as “border errors”, and big melt ponds were often classified as open water in the center, which we referred to as “center errors”. A correction algorithm was designed to rectify this.

We compared MPF, MND and mean and median pond area to a study conducted by Perovich et al. (2002b), which proved to have similar results as ours, even though this study was conducted nearly twenty years ago, with different aerial images, and the numbers reported were from a much larger area. We found that our dataset experienced higher MPFs and larger over-all pond areas. We also introduced a linear regression model that can potentially be used to estimate average pond area by ingesting several melt pond statistics and shape parameters.

Moving forward, we have a few recommendations for further research based on our methods and findings. First, we suggest SAR images should be taken at cross-polarization, in order to exploit the possibility of differentiating ice types, since this was not possible in this study. This would have been a great asset in understanding ice-type dynamics at the different MIZ cluster sites and the addition of polarimetric data would have been added as an additional training feature. For example, the Pauli Decomposition of the scattering matrix is often employed to represent all the polarimetric information in a single SAR image. Finally, in the future, information from incidence angle and wind speed from SAR images should be included in the training phase, as these additional data could alleviate misclassification, by providing more feature dimension.
Second, if computation time was not an obstacle, and we had polarimetric data, we would have used more sophisticated schemes to calculate texture and included these features in a convolutional neural network with multiple hidden layers, which has proven to be an excellent method to distinguish small differences between pixels, however they are widely known to be very computationally expensive in the machine learning community. Therefore, we also suggest running such a scheme with high performance computing and or supercomputers.

Third, if we had more knowledge about what ice types were in the optical image scenes, we would have ingested this information into the linear regression models in order to see what type of relationship could be derived in terms of shape parameters, pond statistics and ice type. This is especially useful, since the MIZ has been undergoing a shift from multi-year ice to first-year ice.
Bibliography


