Mixture Regression for Sea Ice Segmentation

by

Max Manning

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The classification of sea ice in SAR imagery is complicated by statistical nonstationarity. Incidence angle effects, heterogeneous ice conditions and other confounding variables contribute to spatial and temporal variability in the appearance of sea ice. I explore a family of models called mixture regressions which address this issue by endowing mixture distributions with class-dependent trends. I introduce mixture regression as a general technique for unsupervised clustering on nonstationary datasets and propose techniques to improve its robustness in the presence of noise and outliers. I then develop region-based mixture regression models for sea ice segmentation, focusing on the modeling of SAR backscatter intensities under the influence of incidence angle effects. Experiments are conducted on various extensions to the approach including the use of robust estimation to improve model convergence, the incorporation of Markov random fields for contextual smoothing, and the combination of mixture regression with supervised classifiers. Performance is evaluated for ice-water classification on a set of dual-polarized RADARSAT-2 images taken over the Beaufort Sea. Results show that mixture regression achieves accuracy of 92.8% in the unsupervised setting and 97.5% when integrated with a supervised convolutional neural network.

This work improves on existing techniques for sea ice segmentation which enable operational ice mapping and environmental monitoring applications. The presented techniques may also be useful for the segmentation of nonstationary images obtained from other remote sensing techniques or in other domains such as medical imaging.

Acknowledgements

I extend my sincere gratitude to my supervisors, Prof. David Clausi and Prof. Linlin Xu, for their support over the past two years. The global pandemic prevented us from meeting face-to-face for the majority of my degree, but they were always happy to provide helpful advice and suggestions over video calls. I would also like to thank Prof. Paul Fieguth and Prof. Grant Gunn for their valuable feedback on this thesis. Finally, I'd like to acknowledge my peers in the VIP lab remote sensing group, with whom I enjoyed many engaging research discussions.

Dedication

To the ones who have supported me over the past two years: my family, my friends and my poodle Machara.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Vast swaths of the polar oceans are covered with sea ice, with sea ice in the Northern Hemisphere typically covering an area of 12-14 million square kilometers during the annual maximum [64]. However, anthropogenically induced climate change has driven a consistent decline in sea ice over the past four decades. The continuation of this decline will bring far-reaching consequences for the global climate, for Arctic and Antarctic ecosystems, and for human activities in polar regions. Sea ice monitoring is important both to gain understanding of these impacts and to enable polar operational activities in the face of changing and unpredictable ice conditions.

Large-scale sea ice monitoring is made possible through the use of satellite-borne remote sensing technologies. Synthetic aperture radar (SAR) is heavily used for ice monitoring due to its large-area coverage, high resolution and ability to see through clouds and fog [69]. However, deriving data products such as high resolution ice maps from SAR imagery is a complex task. Operational ice maps are typically prepared manually by experts at organisations such as the Canadian Ice Service (CIS)[68]. Manual analyses are too time consuming to make full use of the large volumes of data available from modern SAR systems, necessitating the development of automated ice mapping algorithms to supplement them [68, 50].

Automated sea ice mapping has been an active area of research for nearly three decades, evolving alongside concurrent developments in computer vision and machine learning as well as improved understanding of the interactions between microwaves and sea ice [69, 78]. Existing approaches are divided between supervised methods, which involve training a classifier using a corpus of data with corresponding ground truth labels [38, 34], and unsupervised methods which attempt to find natural groupings of the data without access

to any training labels [77]. A further categorization exists based on the type of image features that are used. Some approaches directly employ SAR backscatter intensities at various polarizations to distinguish between media [50, 75], while others employ higher-order features such as texture statistics [38].

Backscatter intensities carry significant discriminative power, but it is difficult to use them for classification because they are statistically nonstationary due to incidence angle effects [40, 11, 43]. Methods for ice mapping based on backscatter intensities must therefore incorporate models for the variability of backscatter distributions. In this thesis, I explore a solution to this problem using a family of unsupervised models called mixture regressions, which combine clustering and regression to identify subpopulations in datasets which are corrupted by class-dependent nonstationary trends. I develop models for binary ice-water classification based on dual-polarized SAR backscatter intensity, using mixture regressions to model intensity distributions which vary as a function of the SAR sensor incidence angle. Despite the simplicity of the approach it proves highly effective for this task. The proposed framework is also quite flexible, setting the stage for future work which may consider both nonstationarity of image features other than backscatter intensity (e.g., texture statistics) and trend variability across other variables (e.g., 2D spatial coordinates or physical variables such as wind speeds).

I begin in chapter 2 by providing some background on SAR observation of sea ice, the physical origins of nonstationarity in SAR imagery, and existing algorithmic approaches to sea ice segmentation. Chapter 3 introduces mixture regression as a general clustering technique for nonstationary datasets. I emphasize the issue of model convergence to undesirable local optima and propose two techniques for mitigating this issue. In chapter 4 I present a region-based adaptation of mixture regression which is suited for segmentation of large, noisy images. I apply this model for ice-water segmentation on a set of dual-polarized RADARSAT-2 images and draw conclusions on the effectiveness of model variants. Finally, in chapter 5 I provide conclusions and an outlook on possible future developments.

Chapter 2

Background

2.1 Remote Sensing of Sea Ice

Satellite-borne remote sensing platforms have long been instrumental to the study of sea ice. Persistent cloud cover and lack of sunlight during the winter season often precludes the use of optical-wavelength instruments for sea ice analysis. Conversely, the microwave bands are ideal for this application due to their ability to penetrate cloud and their sensitivity to various physical parameters of sea ice. High resolution ice mapping is conducted primarily based on synthetic aperture radar (SAR). Below I provide a brief overview of SAR and its use for sea ice monitoring.

2.1.1 Synthetic Aperture Radar

SAR is a radar imaging technique which can achieve much higher resolution than other radar modes such as scatterometry and real-aperture radar [17]. In general, the angular resolution of a radar system is limited by the size of its antenna; SAR simulates a large "synthetic" antenna by operating on a moving platform and illuminating targets from multiple positions along its track. To form an image, echoes received at multiple positions are coherently combined in a process called focusing [17].

The sensor geometry for a typical satellite-borne SAR is shown in Figure 2.1. Most SAR satellites occupy polar orbits at an altitude of 500-900km. SAR systems are side-looking radars which point nearly perpendicularly to their direction of travel. The strip of ground imaged by the SAR is called the swath. The angle between the sensor's line of sight

and the vertical direction (the nadir) is called the incidence angle, denoted in this work by θ. For the wide-swath SAR imaging modes which are preferred for sea ice mapping, the incidence angle typically varies from around 20 degrees in the near range to 50 degrees in the far range. As will be discussed below, the incidence angle has a significant impact on the radar returns observed over different media.

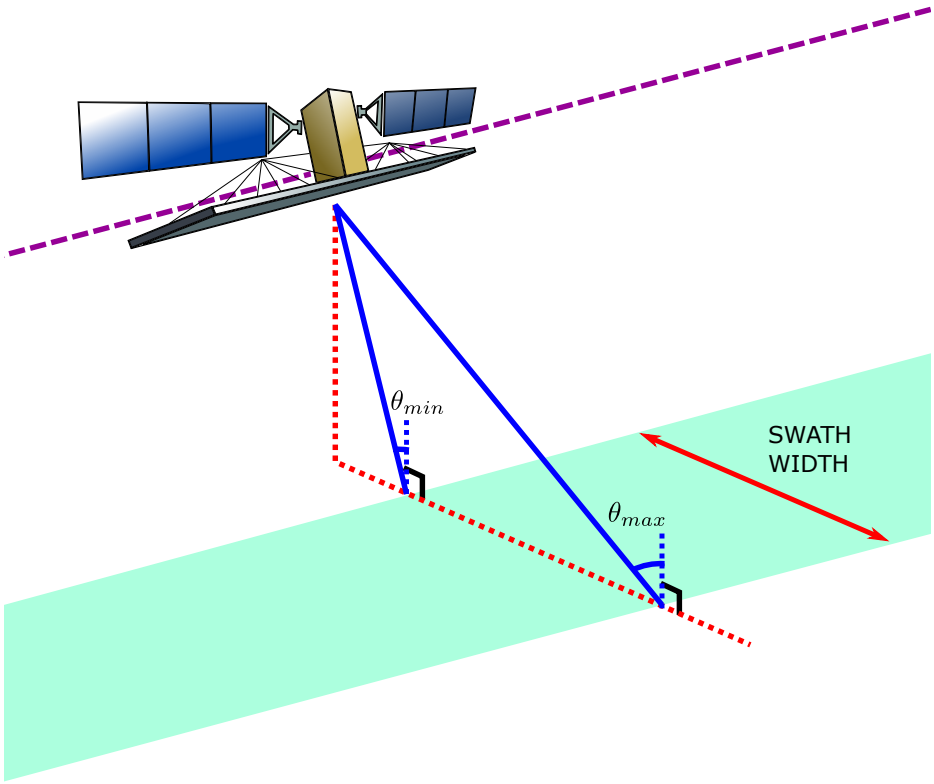

Figure 2.1: The viewing geometry of a SAR satellite. Based on a depiction of RADARSAT-2 beam modes courtesy of MacDonald, Dettweiler and Associates Ltd. [41].

SAR systems operating at different frequencies are sensitive to different target characteristics due to wavelength-dependent interactions of electromagnetic waves with target materials [14]. The primary frequencies currently used for spaceborne SAR reside in L-band ($\sim$1.2 GHz), C band ($\sim$5.4 GHz) and X-band ($\sim$9.6 GHz). All three bands have proven valuable for sea ice analysis [9, 33], but the C-band has been leveraged the most heavily for this purpose. Contemporary C-band SAR platforms such as RADARSAT-2 (the primary

data source for this thesis), RADARSAT Constellation Mission, and Sentinel-1 provide large volumes of data covering most of the globe with short revisit times.

Many SAR systems have the ability to operate under different polarimetric modes [72]. The dual-polarization SAR imagery used in this thesis comprises a co-polarized HH channel (radiation is both transmitted and received with linear horizontal polarization) and a cross-polarized HV channel (radiation is transmitted with linear horizontal polarization and received with linear vertical polarization). Contrast between these channels can help to distinguish between different media due to polarization-dependent scattering mechanisms [72].

2.1.2 Electromagnetic Interactions and Scattering Mechanisms

The interpretation of SAR sea ice imagery requires understanding the interaction of electromagnetic waves with sea ice. The interaction of a material with electromagnetic radiation is governed by its morphology and its dielectric permittivity [24]. The permittivity determines the propagation and absorption of radiation inside a material, and how it is reflected and transmitted at material interfaces [68]. High-permittivity materials tend to produce strong reflections and allow little penetration of electromagnetic waves. The interaction also depends on the frequency of the electromagnetic wave. This is because permittivity is in general a frequency-dependent property, and because wave scattering depends on the roughness scale of the material with respect to the wavelength.

Analysis of SAR backscatter is complicated by the many scattering mechanisms which contribute to radar returns [14]. Scattering which occurs after the radar signal has penetrated a medium is called volume scattering. Strong volume scatter often occurs in media which are largely transparent but which are interspersed with small scattering bodies [14, 23]. For high-permittivity media which allow little penetration of the illuminating signal, scattering occurs mostly at the surface and the return strength is determined by the surface properties. In the case of surface scattering, the surface roughness with respect to the wavelength of the radar signal strongly affects the magnitude and the direction of scattered radiation. Surfaces whose roughness scale is much smaller than the radar wavelength produce specular reflections, where the incident and reflected waves propagate at equal angles to the surface normal. Conversely, surfaces which are highly irregular at the wavelength scale produce a diffuse reflection where the reflected radiation is spread across a large range of angles. Surfaces which contain periodic structures at specific scales may induce particularly strong returns. Bragg scattering, which occurs for surfaces which contain scattering bodies arranged at multiples of half-wavelength spacing along the sensor

line of sight, produces a bright return directed back towards the illumination source [14]. Illustrations of the various scattering mechanisms described above are shown in Figure 2.2.

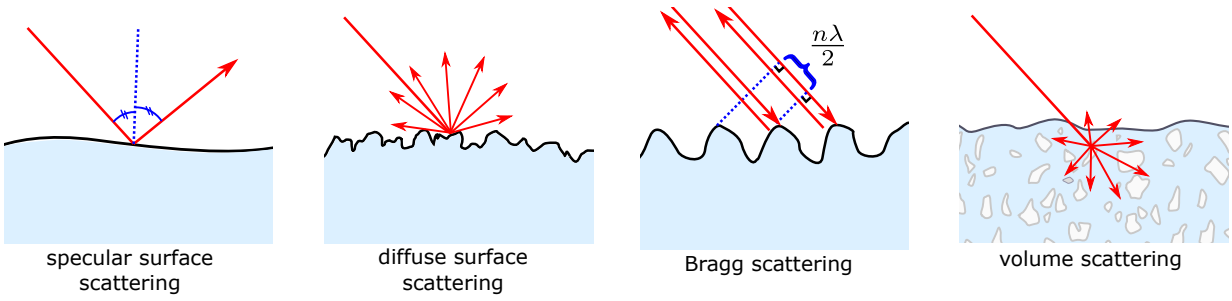

Figure 2.2: SAR scattering mechanisms

2.1.3 Characteristics of Sea Ice Types

For sea ice, the morphological and dielectric properties are modulated by many factors including the physical qualities of the ice and the nature of the surface cover on top of the ice (e.g., snow cover or melt ponds). These factors produce the contrast which allows sea ice to be identified and characterized in radar imagery. Water has a much higher permittivity than snow or ice, allowing less penetration of electromagnetic waves and producing stronger reflections from its surface [68]. The permittivity of snow depends strongly on its moisture content, with dry snow being nearly transparent at frequencies below X-band [15, 27]. The permittivity of sea ice is significantly increased by the presence of brine, giving SAR signatures of sea ice a very different appearance from freshwater ice [68].

Sea ice conditions are categorized by the World Meteorological Organization (WMO) [74], who define a standardized terminology for different sea ice types distinguished primarily by their stage of development. New ice often forms as a slushy mixture on the ocean surface which attenuates waves and appears dark in SAR imagery. Young ice passes thorough a number of stages as it continues to freeze and agglomerate, eventually becoming first-year ice (FYI) when it reaches a thickness of around 30cm [74]. FYI maintains a high salt content which reduces the penetration depth and causes relatively high backscatter intensity primarily influenced by surface effects [74]. Ice formed under calm conditions remains relatively smooth while wind and pressure buildup can cause deformation and ridging. As ice ages, brine drainage occurs and leaves air pockets in its absence. The reduced brine content allows a larger penetration depth of microwaves within the ice, and

scattering occurs at the discontinuities introduced by the air pockets [23]. This results in a much larger volume scattering contribution from multi-year ice (MYI) over young ice [74]. In winter conditions, MYI can often be distinguished from FYI by its bright cross-polarized SAR backscatter, especially at higher frequencies (C and X band).

Interpreting SAR sea ice imagery during the freeze-up and melt seasons is complex. Young ice backscatter intensities range from low values which can hardly be distinguished from calm water to anomalously high values due to frost flowers or other surface phenomena [74] (see Figure D.1). While the dry snow cover present during winter is mostly transparent, melt season brings wet snow cover and melt ponds which mask the signature of the ice underneath [9]. Understanding of the impact of surface cover on microwave backscatter has been gained through electromagnetic simulations of layered media [67, 35] and in-situ measurements obtained with scatterometers [51, 19].

The high permittivity of seawater leads to a very low microwave penetration depth, so SAR returns over open water are largely determined ocean surface conditions. Backscatter intensity over open water is therefore highly dependent on wind [49], with open water in windy conditions often exhibiting returns several dB higher than those of calm open water. Sea ice often exhibits large fissures called "leads". Water in leads typically appears darker than open water as a result of smooth surface conditions brought by wave damping or the presence of thin ice cover [69].

A consequence of the different scattering contributions for sea ice and open water is that each exhibits a different response to changes in incidence angle. A demonstration of this effect is shown in Figure 2.3. The large specular reflection contribution from open water manifests a strong dependence on incidence angle. Open water returns typically decay at a rate between 0.5-1.0 dB/degree from the near range to the far range. Ice also exhibits variability in SAR backscatter across incidence angles, but its rougher surface properties lead to a smaller effect; under winter conditions, C-band HH backscatter typically decays with incidence angle at a rate between 0.16 dB/degree (for MYI) and 0.2-0.3 dB/degree (for FYI) [43, 2]. Backscatter intensity for water in the HH polarization thus typically exceeds that of ice in the near range and falls below it in the far range. At intermediate incidence angles the backscatter intensity distributions in the HH polarization overlap, making ice-water classification difficult.

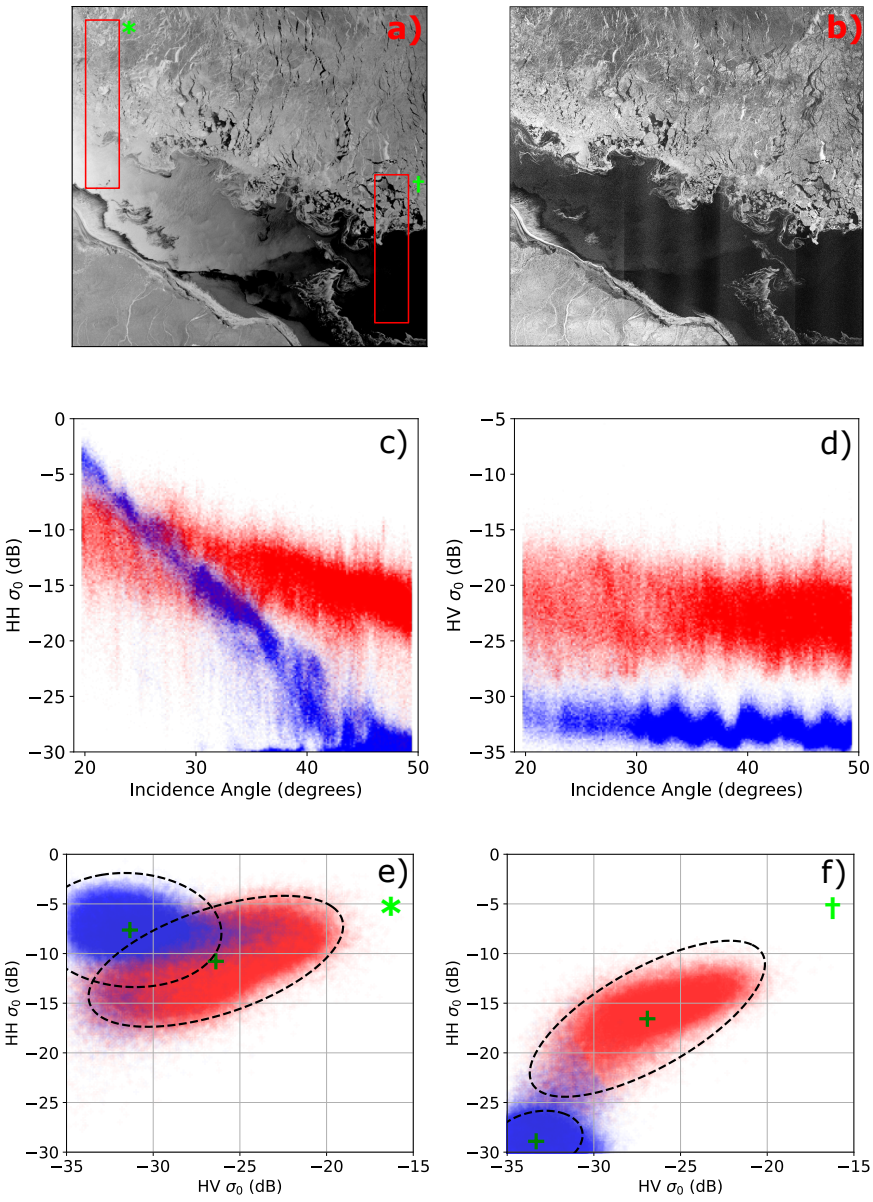

Figure 2.3: Incidence angle effects on a RADARSAT-2 scene from October 27, 2010. (a) HH polarization, (b) HV polarization. Backscatter distributions in (c-f) are shown for water in blue and for ice in red. The variation of backscatter intensities across incidence angle are shown for HH in (c) and HV in (d). Backscatter statistics are shown for the incidence angle ranges of 20-23 degrees (e, statistics extracted from region $*$) and 45-48 degrees (f, statistics extracted from region †).

2.2 Automated Sea Ice Mapping

Automatically producing sea ice maps from SAR imagery has been a long-standing challenge, and a wide variety of techniques have been proposed to address it. Many of these approaches are supervised algorithms which rely on the existence of training data. Simple examples include the use of thresholding [29] or decision trees [20, 39] to produce segmentations based on SAR backscatter intensities. Other supervised models such as support vector machines [38] have also been considered. More recently, deep learning models such as convolutional neural networks have risen to prominence and found use in sea ice segmentation applications [57, 3, 34]. While deep learning models are effective, they require large training datasets to acheive good performance and they lack interpretability due to their "black-box" nature.

On the other hand, unsupervised approaches have also demonstrated strong results for sea ice segmentation. Commonly used methods include clustering algorithms (k-means [6], fuzzy clustering [5], mixture models [6], etc.) and linear discriminant analysis [10]. Both backscatter intensities and texture statistics [10] have been used in such models. A particularly successful combination is the coupling of unsupervised clustering models with Markov random fields (MRFs)[21], which have formed the basis for several effective sea ice segmentation algorithms [77, 50]. The models developed in this thesis fall into this category, the primary deviation from existing work being the adaptation of the clustering models to account for nonstationarity. Below I briefly introduce some algorithmic techniques which provide background for the models developed in later chapters.

2.2.1 Mixture Models

Mixture models (MMs) are a ubiquitous tool for modeling data which arise from a set of unknown sub-populations [6]. MMs are well suited for the purpose of image segmentation, where they have frequently been used to perform unsupervised clustering on image features. In a multivariate MM, the probability distribution of a random vector $\mathbf{x}$ is represented as a weighted sum of K component distributions as shown in (2.1). Each mixture component has a set of parameters $\boldsymbol{\theta}_k$ and is weighted by a parameter π_k called the mixture coefficient. The mixture coefficients obey $\sum_k \pi_k = 1$ in order to ensure that $P(\mathbf{x})$ is normalized.

$$P(\mathbf{x}) = \sum_{k=1}^{K} \pi_k P_k(\mathbf{x}|\boldsymbol{\theta}_k) \tag{2.1}$$

The most common MM is the Gaussian mixture model (GMM), where each of the component distributions is a multivariate Gaussian $P_k(\mathbf{x}|\boldsymbol{\theta}_k) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. The parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ respectively denote the mean vector and the covariance matrix of the distribution for class k.

Mixture model fitting presents a chicken and egg problem. To estimate the parameters of the component distributions, it must be known which data points originate from which component; however, attributing the data points to the mixture components requires knowledge of the component parameters. EM-type algorithms provide an iterative method for overcoming this problem. The idea is to simply guess an initial set of parameters for the mixture components, and compute a set of labels based on this guess. The process is then iterated, updating the labels and component parameters at each step until convergence is reached. Variants of this procedure are used to fit a wide variety of clustering models including K-means and the GMM. A more formal development of the EM algorithm for the GMM is given in Appendix A. Since most of the algorithms developed in this work are based on the EM algorithm for GMMs, the reader is encouraged to review this procedure.

2.2.2 Markov Random Fields

Markov random fields (MRFs) provide a method for modeling the statistical dependencies between groups of unobserved variables. They have been a mainstay in computer vision since the seminal work of Geman and Geman [21], used for tasks such as denoising and segmentation. MRFs are particularly advantageous SAR image segmentation due to their ability to overcome noise, and as such they have been employed in many successful sea ice segmentation approaches [77, 50, 38]. A brief summary of MRF modeling for image segmentation is given in Appendix B along with details on the optimization approach used in this study.

2.2.3 Overcoming Nonstationarity

Random processes whose statistical properties vary with respect to some variable (for example, time or space) are called *nonstationary*. Nonstationary image data arises in several contexts, and is particularly common in remote sensing and medical imaging. For example, magnetic resonance imagery is often corrupted by a location-dependent "bias field" which leads to inhomogeneous intensities across the image [56]. Techniques have been proposed to estimate and remove the bias field [47], but this corresponds to a global correction which

does not account for class-dependent nonstationarity. The primary source of nonstationarity in SAR imagery is incidence angle effects, which existing methods have treated in a variety of ways. Global incidence angle correction schemes have been considered [2] but their effectiveness is limited to homogeneous scenes due to the class-dependent nature of the incidence angle effect. Another possibility is to use a "glocal" approach [38], in which SAR scenes are divided into regions across which are small enough that the class statistics are stationary. Segmentation is applied within each region individually, and the results are then combined in a hierarchichal gluing process. A more direct approach, and the one which is pursued in this work, is to construct a model for class-dependent nonstationary effects and incorporate it into a classifier. This approach was pioneered by Cristea et al. [11] and Lohse et al. [40]. Although they did not recognize it as such, the model proposed by Cristea et al. is a mixture regression; the models developed in this thesis are generalizations of this work.

2.3 Experimental Dataset

2.3.1 Data Description and Study Area

The main dataset used in this thesis is a set of 35 dual polarized (HH and HV) RADARSAT-2 images acquired over the Beaufort Sea between April and December for the years 2010 and 2011. Winter is excluded because near-total ice cover persists in the Beaufort sea during the winter months. Throughout the spring, summer and fall, winds and currents in the Beaufort sea drive a unique seasonal pattern of ice conditions. In particular, MYI transported into the Beaufort sea becomes trapped by a large circular current called the Beaufort Gyre. The Beaufort sea has thus historically been an important reservoir for thick ice, although recently the fraction of MYI in the region has declined precipitously [18]. The variability of ice conditions in this region make it an ideal case study for evaluating the effectiveness of automated sea ice mapping algorithms. Previous work has used a subset of the dataset for this purpose [38].

The images are acquired in the ScanSAR Wide beam mode, each covering an area of approximately 500x500km at a nominal pixel resolution of 50m. The bounding boxes for the scenes are shown in Figure 2.4. Corresponding weather data (air temperatures and wind speeds) are obtained from the Global Historical Climatology Network-Daily database (GHCN-Daily)[44, 45] for station USW00027502 located in Utquiagvik, Alaska.

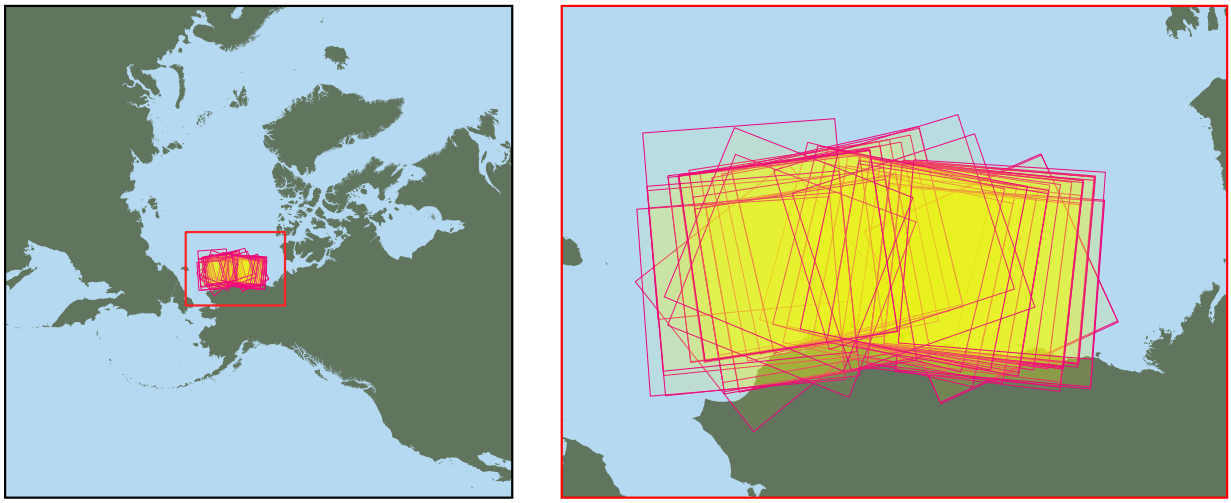

Figure 2.4: Locations of the RADARSAT-2 scenes used in this study.

2.3.2 Data Pre-processing

Radiometric calibration is applied to image each to obtain the σ_0 values for each channel following the RADARSAT-2 product description [42]. All processing is performed in the log-domain (σ_0 values in dB). For the HH channel, σ_0 is thresholded between -30dB and 0dB and linearly mapped to the range $[0, 1]$. For the HV channel, the low and high threshold levels are set to -35dB and -5dB, respectively. I apply block averaging to each image using a 4x4 pixel window, resulting in a resolution of approximately 200m and image dimensions of around 2500 by 2500 pixels. The block averaging helps to reduce speckle noise and lower the computational burden while maintaining sufficient spatial resolution for operational applications.

Chapter 3

Modeling Nonstationarity with Mixture Regressions

Mixture regression [6] combines two fundamental tasks in machine learning, clustering [6] and regression [28]. Early studies of mixture regression [54] arose from the field of econometrics, where it was known as switching regression and was used for regression analysis on data arising from multiple sub-populations or economic regimes. Efficient estimation of mixture regressions was enabled with the development of the EM algorithm [13, 12]. The first well-known application of mixture regression in nonstationary image segmentation is the work of Cristea et al. [11], who incorporated incidence angle dependencies into a clustering-based framework for SAR image segmentation. However, they did not identify their model as a mixture regression and therefore did not draw connections to previous literature on the topic.

In many applications, obtaining accurate regression parameters for each component of a mixture regression is of primary importance. Conversely, in this thesis the primary objective is clustering, with the regression estimates serving to facilitate cluster separation in the presence of nonstationarity. The following sections introduce mixture regression with an emphasis on this point of view. After establishing some terminology I introduce the mixture of linear regressions (MLR), which is a straightforward combination of the Gaussian mixture model with linear regression. I then discuss techniques for improving model convergence and accomodating various patterns of nonstationarity, and finally explore techniques for mixture regression when the dependence on the covariates is nonlinear.

3.1 Keeping Up with the Trends

When the probability distribution of a random variable changes according to one or more covariates (e.g., space or time) it is said to be nonstationary. To capture this effect in a mixture model, the mixture components must be modified to reflect their dependence on the covariates. I represent a general set of covariates with the vector $\mathbf{y}$ as shown in (3.1).

$$P(\mathbf{x}|\mathbf{y}) = \sum_{k=1}^{K} \pi_k P_k(\mathbf{x}|\boldsymbol{\theta}_k, \mathbf{y}) \tag{3.1}$$

In this thesis, the primary source of nonstationarity under consideration is the incidence angle effect, so $\mathbf{y}$ is simply the incidence angle at each image location. Other choices are however possible such as the 2D image coordinates p_x and p_y or any other quantity across which the observed data systematically varies in a class-dependent manner.

A common type of nonstationarity arises when the mean value of a distribution varies across a covariate but the distribution remains otherwise unchanged. Such distributions are commonly encountered in time series analysis where they are known as trend-stationary processes [31]. In the context of mixture distributions, the covariate may affect each of the mixture components differently resulting in a different trend for each class. An example of this type of nonstationarity in a two-component mixture distribution is shown in Figure 3.1. As can be seen, the presence of nonstationarity poses a challenge for the separation of the mixture components based on the measured values of $\mathbf{x}$. Although the two populations are well separated over most values of the covariate, marginalizing over the covariate makes the component distributions overlap significantly.

Consider the situation where the covariate values are known for each data point, that is we have a set of measurements $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$. It is then possible to account for the nonstationarity by obtaining an estimate of the trend for each mixture component. For class k, denote the trend estimate by $\mathbf{g}_k(\mathbf{y})$. The GMM is converted to the more general Gaussian mixture regression by replacing the constant mean vectors $\boldsymbol{\mu}_k$ with $\mathbf{g}_k(\mathbf{y})$. The rest of this chapter details functional forms for $\mathbf{g}_k(\mathbf{y})$ and methods for estimating them.

3.2 Mixtures of Linear Regressions

In the mixture of linear regressions (MLR) [12], the trend function is a linear function $\mathbf{g}_k(\mathbf{y}) = \mathbf{w}_k^T \boldsymbol{\phi}(\mathbf{y})$. The term $\mathbf{w}_k$ is the set of regression weights for class k and $\boldsymbol{\phi}(\mathbf{y})$ is a

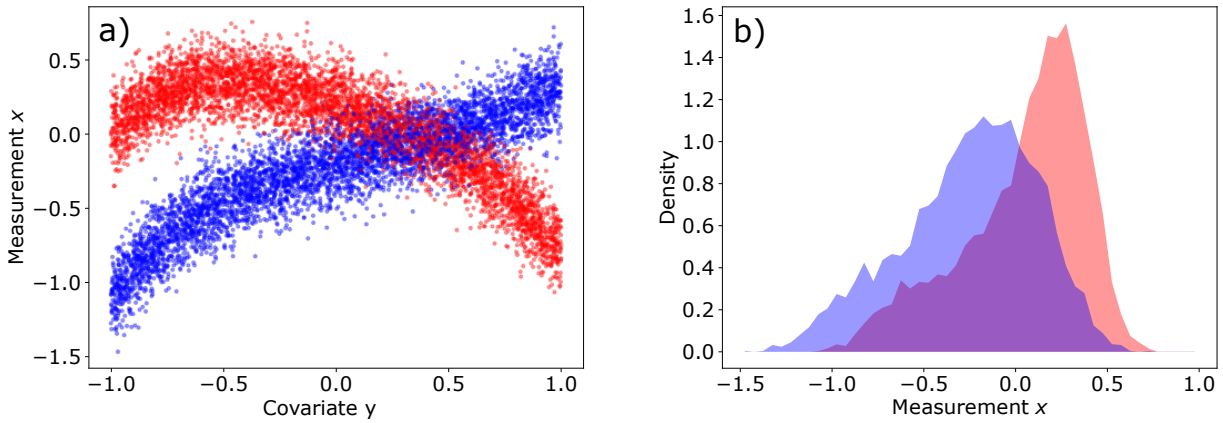

Figure 3.1: A toy dataset demonstrating a mixture of trend-stationary distributions. (a) Samples from a nonstationary mixture distribution across various values of the covariate. (b) The marginal distributions of the mixture components.

set of basis functions over $\mathbf{y}$. This model subsumes the GMM, which is recovered when $\boldsymbol{\phi}$ is 1 for all data points. The simplest non-trivial choice of basis, corresponding to a linear trend function, is to take $\boldsymbol{\phi}(y) = [1\, y_0 \ldots y_n]^T$; the values of $\mathbf{w}_k$ then consist of a mean value along with a slope for each element of $\mathbf{y}$. Alternate choices of basis are discussed in section 3.4.

Fitting a linear regression mixture is possible using a simple modification to the EM algorithm for GMMs. The E step proceeds as shown in (3.2), the only difference from A.2 being the replacement of $\boldsymbol{\mu}_k$ with $\mathbf{w}_k^T \boldsymbol{\phi}_i$. Note that I have adopted the shorthand $\boldsymbol{\phi}_i = \boldsymbol{\phi}(\mathbf{y}_i)$.

$$z_{ik} = \frac{\pi_k \mathcal{N}(\mathbf{x}_i | \mathbf{w}_k^T \boldsymbol{\phi}_i, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_i | \mathbf{w}_j^T \boldsymbol{\phi}_i, \boldsymbol{\Sigma}_j)} \tag{3.2}$$

The M step updates the mixture component parameters as shown in (3.3)-(3.5), in which $\boldsymbol{\Phi} = [\boldsymbol{\phi}_0\, \boldsymbol{\phi}_1 \ldots \boldsymbol{\phi}_N]^T$ and $\mathbf{R}_k = diag(\{z_{ik}\}_{i=1}^N)$. The update step for $\mathbf{w}_k$ is a weighted linear regression where the weighting for data point i is z_{ik}.

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} z_{ik} = \frac{N_k}{N} \tag{3.3}$$

$$\mathbf{w}_k = (\boldsymbol{\Phi}^T \mathbf{R}_k \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi}^T \mathbf{R}_k \mathbf{X} \tag{3.4}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N} z_{ik} (\mathbf{x}_i - \mathbf{w}_k^T \boldsymbol{\phi}_i)(\mathbf{x}_i - \mathbf{w}_k^T \boldsymbol{\phi}_i)^T \tag{3.5}$$

In practice, it is beneficial to add a small regularization constant to the diagonal of $\mathbf{\Phi}^T\mathbf{R}_k\mathbf{\Phi}$ to ensure that it is non-singular before carrying out the inversion [66]. Equation (3.4) is thus modified as shown in (3.6), where λ is a small positive constant and $\mathbf{I}$ is the identity matrix.

$$\mathbf{w}_k = (\mathbf{\Phi}^T\mathbf{R}_k\mathbf{\Phi} + \lambda\mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{R}_k\mathbf{X} \tag{3.6}$$

3.3 Convergence Considerations

A perennial challenge with unsupervised clustering models is convergence to undesirable local maxima. The problem stems from the non-convexity of most common clustering objectives [6] (the likelihood function in the case mixture models) and the use of EM-type algorithms which monotonically increase the objective at each step. If the iterations are initialized in the basin of attraction of a poor solution they are unable to escape. Despite some recent theoretical progress [32, 36, 37], the convergence behaviour of the EM algorithm remains poorly understood for GMMs and even more so for mixture regressions.

Mixture models are made useful in practice using a variety of techniques which can push them towards desirable solutions. A widely used solution is cross-validation [25], where the model is run many times from different starting points and some criterion is used to select the best result. This approach is time-consuming and subject to the availability of a criterion which accurately measures the quality of the solution, so it is not considered here. In this section I consider two alternative methods. The first is the use of robust estimators, which reduce the impact of outliers when fitting the cluster parameters. The second is deterministic annealing, a technique inspired by statistical mechanics which gradually decreases a "temperature" parameter over the course of model fitting.

3.3.1 Robust Estimation in Mixture Models

While the standard mixture of linear regressions model uses least squares to obtain the regression weights $\mathbf{w}_k$, (3.4) can be replaced with another suitable regression procedure without modification to the rest of the algorithm. Alternatives to the least squares objective can fulfill desiderata such as obtaining sparsity in the weights or increasing the model robustness to outliers [30]. In this work I consider the latter goal using a method called iteratively reweighted least squares (IRLS) [58].

IRLS obtains successive estimates of the regression parameters $\mathbf{w}_k^{(t)}$ in a series of steps $t = 0, 1, ..., T$. The initial estimate $\mathbf{w}_k^{(0)}$ is obtained as in (3.4). Subsequent estimates

are obtained using weighted least squares [28], where the data points are reweighted by a function of their residual with the estimated trend. Effective reweighting schemes include the family of M-estimators, the canonical example of which is the Huber function [30] $w_{ik}^{(t)} = \min\{1, \delta/r_{ik}^{(t)}\}$ where $r_{ik}^{(t)}$ is the residual magnitude shown in (3.7) and δ is a tuning parameter.

$$r_{ik}^{(t)} = |\mathbf{x}_i - \mathbf{w}_k^{(t)^T} \boldsymbol{\phi}_i| \tag{3.7}$$

The reweighted regression steps are performed in (3.8), where $\tilde{\mathbf{R}}_k^{(t)} = \text{diag}(\{z_{ik} w_{ik}^{(t)}\}_{i=1}^N)$. As in (3.6), a small regularization constant λ is added to ensure that the inversion is stable.

$$\mathbf{w}_k^{(t+1)} = (\boldsymbol{\Phi}^T \tilde{\mathbf{R}}_k^{(t)} \boldsymbol{\Phi} + \lambda \mathbf{I})^{-1} \boldsymbol{\Phi}^T \tilde{\mathbf{R}}_k^{(t)} \mathbf{X} \tag{3.8}$$

In this scheme, data points with small residuals $r_{ik}^{(t)} < \delta$ are assigned a weight of 1, while outliers with $r_{ik}^{(t)} > \delta$ are given less importance. The value of δ can thus be tuned to achieve the desired degree of outlier rejection, with $\delta \to \infty$ corresponding to ordinary least squares. When δ approaches zero the result approaches the median regression (also known as L_1 or quantile regression), which corresponds to minimizing the sum of the absolute residuals rather than their squares. Robust variants of mixture regression have been considered by Bai et al. [4], and in the context of stationary clustering it is related to the k-medians problem. A demonstration of the outlier-handling ability of a robust clustering model employing IRLS is shown in Figure 3.2.

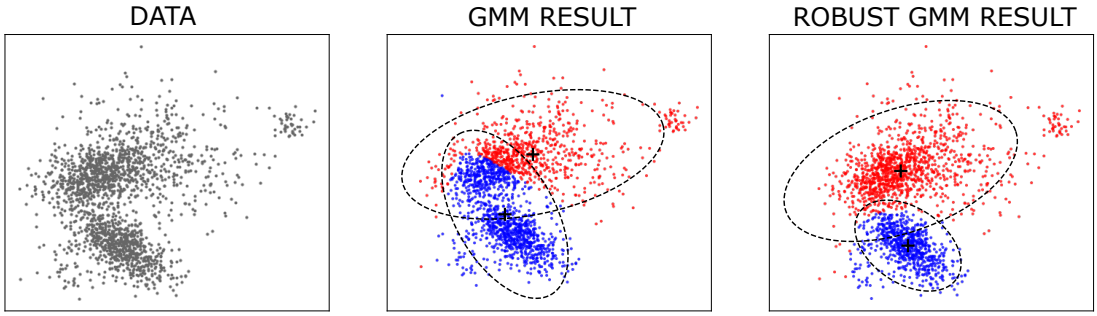

Figure 3.2: Demonstration of robust clustering on a toy dataset consisting of two main clusters along with an outlier population on the upper right. The outliers are detrimental to the GMM result but not to the robust GMM result. 95% confidence ellipses are shown for the final clusters.

Sensitivity to outliers in clustering problems can also be alleviated by changing the number of clusters, allowing outlier populations to form separate clusters. However, the

optimal number of clusters cannot usually be determined a priori and must be selected by expensive cross-validation procedures. Further, changing the number of clusters can decrease the interpretability of the results if the clusters are intended to represent specific sub-populations. This motivates the development of robust clustering procedures that are tolerant to mis-specification of the number of clusters.

3.3.2 Soft EM, Hard EM and Deterministic Annealing

The traditional EM algorithm for GMM fitting [6] obtains "soft" estimates of the class labels in the E step, that is, the responsibility for a particular data point is distributed among the classes according to a probability distribution. This contrasts with the K-means clustering algorithm [6], where a "hard" classification is used which assigns each data point entirely to the nearest cluster center. A hybrid between the two, sometimes called the Classification-EM (CEM) algorithm or elliptic K-means [16, 6], retains the M step of the GMM but uses a hard E step as shown in (3.9).

$$z_{ik} = \begin{cases} 1 & k = \arg\max_j \pi_j \mathcal{N}(\mathbf{x}_i | \mathbf{w}_j^T \boldsymbol{\phi}_i, \boldsymbol{\Sigma}_j) \\ 0 & otherwise \end{cases} \tag{3.9}$$

Comparison studies have indicated that the CEM algorithm brings improved convergence speed over regular EM for mixture regressions in certain applications [16], and several successful image segmentation models use GMMs with hard update steps [76, 53].

It is in fact possible to interpolate between the soft and hard EM algorithms by recognizing that the E step of the regular EM algorithm produces a Gibbs distribution over the class labels. Consider the log-probabilities of the mixture components for a Gaussian mixture regression with trend functions $\mathbf{g}_{ik}$ shown in (3.10).

$$u_{ik} = \log(\pi_k) - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\mathbf{x}_i - \mathbf{g}_{ik})^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \mathbf{g}_{ik}) \tag{3.10}$$

A generalized E step is $\mathbf{z}_i = \text{softmax}(\mathbf{u}_i/T)$ where T is a scalar parameter called the temperature and the softmax function is defined in (3.11).

$$\text{softmax}(\mathbf{x})_k = \frac{\exp(x_k)}{\sum_j \exp(x_j)} \tag{3.11}$$

Taking $T = 1$ corresponds to the regular soft E step, and the hard E step is recovered in the limit as $T \to 0$.

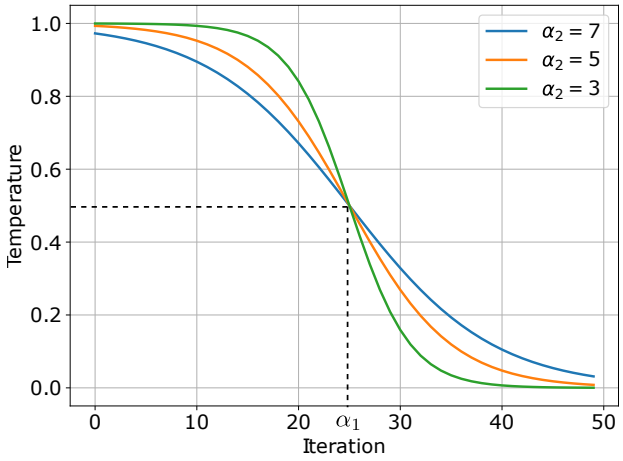

Figure 3.3: Annealing schedule for EM iterations

The temperature parameter is named in analogy to statistical mechanics, where the Gibbs distribution describes a physical system in equilibrium with a large reservoir having a particular temperature. If the reservoir has a high temperature the system can explore its high energy states, with all states becoming equiprobable as $T \to \infty$; as $T \to 0$ the probability mass concentrates entirely on the lowest energy state(s). This view of the relationship between hard and soft clustering algorithms inspired a technique called deterministic annealing [61, 60, 70] which aims to improve the convergence of clustering algorithms by systematically varying the temperature over the course of the clustering procedure. The temperature is initially set to a high value and is slowly decreased according to a function known as the annealing schedule. This strategy bears resemblance to well known stochastic sampling algorithms such as simulated annealing, but unlike these methods each update step proceeds deterministically.

In this work I use a sigmoidal annealing schedule shown in (3.12) parameterized by location and scale parameters respectively denoted α_1 and α_2. The influence of the parameters is illustrated in Figure 3.3.

$$T(\tau) = \frac{1}{1 + \exp((\tau - \alpha_1)/\alpha_2)} \tag{3.12}$$

3.4 Nonlinear Dependence on the Covariates

3.4.1 MLR with Polynomial Basis

It is common to perform nonlinear regression by performing a basis expansion over the covariates [28]. The same idea can be used to fit a nonlinear trend function in MLR. The optimal choice of basis depends on the application, and it can be used to encode prior information about the nature of the trend function. In this work I use the Legendre polynomials truncated at a maximum order B as the nonlinear basis. This is an appropriate choice for modeling trend functions which vary smoothly over a fixed interval in the covariate, since the Legendre polynomials are smooth and orthogonal under a unit weighting function on the interval $[-1, 1]$. The Legendre polynomial basis for univariate y is shown in (3.13), where $P_n(y)$ is the n^{th} Legendre polynomial.

$$\boldsymbol{\phi}(y) = [P_0(y)\, P_1(y)\, ...\, P_B(y)]^T \tag{3.13}$$

Increasing the order of the basis allows more flexible trend functions to be fit, however higher order basis functions tend to incite problems with convergence and identifiability. Consider the dataset illustrated from Figure 3.1, in which the trend functions for the two mixture components were generated by fourth-order polynomials. Examples of mixture regression trends of varying degrees fit to this dataset are shown in Figure 3.4. Overly flexible trend functions typically lead to catastrophic overfitting as shown in panel (b). Even if the order of the basis is correctly specified, nonlinear mixture regression encounters ambiguities at points where the trend functions of two components cross one another. This ambiguity is shown in panels (c) and (d) of Figure 3.4. The fit in panel (c) is close to the true result, but without prior knowledge the result in panel (d) appears equally good.

If it is known that the trend function comprises a small high-order correction on top of a larger low-order trend, an effective strategy is to use the result from a low-order mixture regression to initialize a more flexible model. The low-order trend provides a favorable starting position from which the higher-order model can "relax" onto the true trend. In the example considered above, random initializations of a mixture regression with a 4^{th} order trend converge to the solutions is panels (c) and (d) of Figure 3.4 with nearly equal probability. However, if the model is initialized with the linear trend fit from panel (a) the model converges to the result in panel (c) every time, which is an accurate fit to the true class labels shown in Figure 3.1.

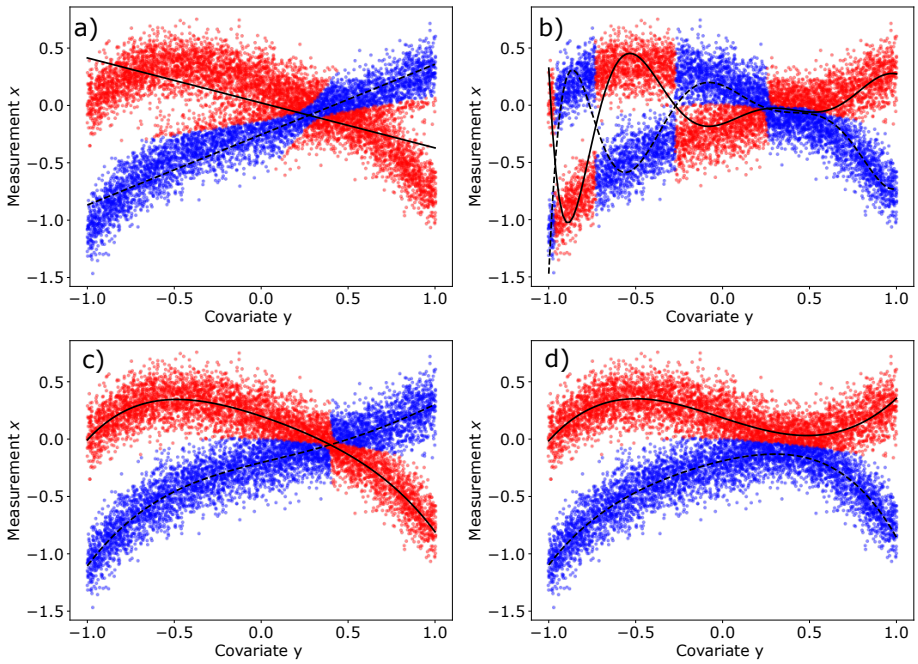

Figure 3.4: Mixture regression fits with polynomial bases on the toy dataset from Figure 3.1. (a) Linear trend, (b) Polynomial trend of degree 10, (c-d) Polynomial trends of degree 4.

3.4.2 Mixtures of Kernel Regressions

An alternative to selecting a particular basis to represent the trend function is to employ a nonparametric estimation method such as kernel regression [65]. This results in the mixture of kernel regressions (MKR). The simplest example employs the Nadaraya-Watson estimator [48, 73] shown in (3.14).

$$\tilde{\mathbf{x}}_i = \frac{\sum_{j=1}^{N} k(\mathbf{y}_i, \mathbf{y}_j)\mathbf{x}_j}{\sum_{j=1}^{N} k(\mathbf{y}_i, \mathbf{y}_j)} \tag{3.14}$$

The term $k(\cdot, \cdot)$ is the kernel function such as the Gaussian kernel $k(\mathbf{y}_i, \mathbf{y}_j) = exp(\|\mathbf{y}_i - \mathbf{y}_j\|_2^2/2\sigma^2)$, where $\|\cdot\|_2^2$ is the squared Euclidean distance. The parameter σ controls the bandwidth of the kernel, and it must be chosen to match the scale of the spatial variation in the data. Employing a non-parametric form for $\mathbf{g}_k(\mathbf{y})$ changes little about the EM model fitting procedure. The E step remains identical, and the M step is modified according to

(3.15)-(3.17).

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} z_{ik} = \frac{N_k}{N} \tag{3.15}$$

$$\mathbf{g}_k(\mathbf{y}_i) = \frac{\sum_{j=1}^{N} k(\mathbf{y}_i, \mathbf{y}_j) z_{jk} \mathbf{x}_j}{\sum_{j=1}^{N} k(\mathbf{y}_i, \mathbf{y}_j) z_{jk}} \tag{3.16}$$

$$\mathbf{\Sigma}_k = \frac{1}{N_k}\sum_{i=1}^{N} z_{ik}(\mathbf{x}_i - \mathbf{g}_k(\mathbf{y}_i))(\mathbf{x}_i - \mathbf{g}_k(\mathbf{y}_i))^T \tag{3.17}$$

A drawback of the MKR model compared to the MLR is its scalability; the computational costs of naive kernel estimation methods scale quadratically with the number of data points, although efficient approximations such as inducing point methods exist [55, 1]. Further, MKR does not bring a strong benefit over MLR for the present application because the nonstationary trends in SAR imagery are nearly linear. For these reasons MKR models are not explored further in this thesis, but they are introduced here to demonstrate the flexibility of the mixture regression framework and to anticipate possible related developments such as mixtures of Gaussian process regressions.

Chapter 4

Sea Ice Segmentation with Mixture Regression

In this chapter, I apply the mixture regression models developed in the previous chapter to the problem of sea ice segmentation in SAR imagery. I begin by developing a region-based variant of the MLR which is tailored towards the segmentation of large images corrupted by nonstationarity and speckle noise. The model can incorporate linear or nonlinear trend functions and can be combined with a MRF for contextual smoothing. I then conduct experiments to explore the effectiveness of various model configurations.

4.1 Proposed Models

4.1.1 Region-Based Mixture Regression

Following previously developed SAR image segmentation methods [77, 38], I use a region-based segmentation approach where the basic units to be classified are small homogeneous groups of pixels called regions[1]. This brings the benefit of reduced computational cost and better resistance to speckle noise than pixel-based approaches without requiring excessive downsampling of the image. The regions are obtained using the watershed algorithm applied to the vector gradient magnitude of the input image as described by Qin et al. [53].

[1] Related work in computer vision often refers to regions as "superpixels."

The MLR is modified for the region-based framework in the following way. Let N denote the number of regions in the image, which are indexed by i. Each region consists of a set of pixels Ω_i with $|\Omega_i|$ denoting the number of pixels in the region. EM updates are computed in terms of the first and second order statistics of the pixels in each region, namely the mean $\bar{\mathbf{x}}_i = \sum_{s\in\Omega_i} \mathbf{x}_s/|\Omega_i|$ and the second moment $\tilde{\mathbf{x}}_i = \sum_{s\in\Omega_i} \mathbf{x}_s\mathbf{x}_s^T/|\Omega_i|$, where $\mathbf{x}_s$ denotes the feature vector for pixel s.

I employ an E step with temperature parameter T as described in section 3.3.2. The update is given by $\mathbf{z}_i = \text{softmax}(\mathbf{u}_i/T)$ where the elements of $\mathbf{u}_i$ are shown in (4.1). The shorthand $\mathbf{g}_{ik}$ is used for $\mathbf{g}_k(\mathbf{y}_i)$.

$$u_{ik} = \log(\pi_k) - \frac{1}{2}\log|\mathbf{\Sigma}_k| - \frac{1}{2}(\bar{\mathbf{x}}_i - \mathbf{g}_{ik})^T\mathbf{\Sigma}_k^{-1}(\bar{\mathbf{x}}_i - \mathbf{g}_{ik}) \tag{4.1}$$

The region-based M step is shown in equations 4.2-4.4, where I have used the notation $\bar{\mathbf{X}} = [\bar{\mathbf{x}}_1\ \bar{\mathbf{x}}_2\ ...\ \bar{\mathbf{x}}_N]^T$ and $\mathbf{\Omega} = \text{diag}(\{|\Omega_i|\}_{i=1}^N)$.

$$\pi_k = \frac{\sum_{i=1}^N |\Omega_i| z_{ik}}{\sum_{i=1}^N |\Omega_i|} \tag{4.2}$$

$$\mathbf{w}_k = (\mathbf{\Phi}^T\mathbf{\Omega}\mathbf{R}_k\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{\Omega}\mathbf{R}_k\bar{\mathbf{X}} \tag{4.3}$$

$$\mathbf{\Sigma}_k = \frac{\sum_{i=1}^N z_{ik}\sum_{s\in\Omega_i}(\mathbf{x}_s - \mathbf{g}_{ik})(\mathbf{x}_s - \mathbf{g}_{ik})^T}{\sum_{i=1}^N |\Omega_i| z_{ik}} \tag{4.4}$$

It is possible to express 4.4 solely in terms of region-level statistics as shown in (4.5). This provides an efficiency benefit since the region statistics can be pre-computed and re-used during each iteration.

$$\mathbf{\Sigma}_k = \frac{\sum_{i=1}^N |\Omega_i| z_{ik}\left(\tilde{\mathbf{x}}_i - \mathbf{g}_{ik}\bar{\mathbf{x}}_i^T - \bar{\mathbf{x}}_i\mathbf{g}_{ik}^T + \mathbf{g}_{ik}\mathbf{g}_{ik}^T\right)}{\sum_{i=1}^N |\Omega_i| z_{ik}} \tag{4.5}$$

The regression weight update in (4.3) can be replaced with a robust variant as described in section 3.3.1. Once the EM steps have been completed, the final label for region i is computed as $\ell_i = \arg\max_k\ z_{ik}$.

4.1.2 Trend Functions

I begin by considering models in which the trend function of the backscatter log-intensity for each class is modeled as a linear function of the incidence angle as shown in (4.6)

following Cristea et al. [11] and Lohse et al [40]. I refer to this model as the linear trend model.

$$\mathbf{g}_k(\theta_i) = \mathbf{a}_k + \mathbf{b}_k\theta_i \tag{4.6}$$

In the formulation presented in chapter 3, the linear trend model corresponds to taking $\boldsymbol{\phi}(\theta_i) = [1\ \theta_i]^T$ and $\mathbf{w}_k = [\mathbf{a}_k\ \mathbf{b}_k]^T$. In the region-based approach θ_i is taken to be the incidence angle at the centroid[2] of region i.

I also consider trend functions which are nonlinear functions of the incidence angle, referring to the MLR where the basis $\boldsymbol{\phi}(\theta_i)$ is the set of Legendre polynomials up to order n as the n^{th}-order polynomial trend model. This is the model that was considered in section 3.4, which is written more explicitly in (4.7).

$$\mathbf{g}_k(\theta_i) = \mathbf{a}_k + \sum_{j=1}^{n} \mathbf{b}_{jk} P_j(\theta_i) \tag{4.7}$$

Polynomial trend model fits are obtained by first running the linear trend model to convergence and using the resulting labels as initialization. The linear trend model and the polynomial trend models are contrasted with the standard GMM which represents the case where the backscatter is stationary (no trend).

Beyond nonstationarity due to incidence angle effects, I also considered the possibility of mixture regression using a two-dimensional trend function over the spatial coordinates. The motivation for this approach is that such a model could represent trend nonstationarity arising from any (possibly unknown) spatially varying quantities provided that the response is spatially correlated. The trend could be represented for example using a MLR with a two-dimensional basis or with a two-dimensional kernel regression. Unfortunately preliminary experiments with this approach encountered problems. As the flexibility of the 2D basis was increased, the model's convergence suffered before any significant benefit was observed. Further work is required to make two-dimensional trend models viable so I do not consider them further in this work.

4.1.3 Regularizing Mixture Regressions with Markov Random Fields

Clustering models such as GMMs and mixture regressions do not make use of spatial context information leading to noisy results. Segmentation results with many small isolated

[2] regions are small enough that the within-region incidence angle variation is negligible.

regions are sometimes undesirable, for example when sea ice maps are converted into polygon shapefiles for use in GIS applications. This issue can be alleviated using MRFs. Previous work has employed a region-based MRF to regularize the results of a GMM [53]; replacing the GMM with a mixture regression is straightforward. Details on the region-based MRF and the optimization technique are given in Appendix B.

The objective function for the mixture regression-MRF is shown in (4.8).

$$\boldsymbol{\ell}^* = \arg\min_{\boldsymbol{\ell}} E(\boldsymbol{\ell}) = \arg\min_{\boldsymbol{\ell}} \left\{ \sum_{i \in \mathcal{V}} U_i(\ell_i) + \sum_{(s,t) \in \mathcal{E}} V_{ij}(\ell_i, \ell_j) \right\} \tag{4.8}$$

The unary potentials $U_i(\ell_i)$ are derived from the mixture regression parameters as shown in (4.9). The pairwise potentials $V_{ij}(\ell_i, \ell_j)$ are obtained as described in (B.3) in Appendix B.

$$U_i(k) = |\Omega_i| \big(\log(\pi_k) - \frac{1}{2} \log|\boldsymbol{\Sigma}_k| - \frac{1}{2} (\bar{\mathbf{x}}_i - \mathbf{g}_{ik})^T \boldsymbol{\Sigma}_k^{-1} (\bar{\mathbf{x}}_i - \mathbf{g}_{ik}) \big) \tag{4.9}$$

When using MRFs with mixture models as the unary potential, it is common to incorporate the MRF into the EM iterations, replacing maximum likelihood estimates for z_{ik} in the E step with maximum a posteriori estimates. This appears to negatively impact convergence in the case of mixture regressions, so I use the MRF as a post-processing step by first fitting the mixture regression without any MRF regularization and using the resulting parameters to compute the unary potential for the MRF.

4.1.4 Nonstationary MRF Weighting

When using MRFs for image segmentation it is necessary to select a weight which controls how strongly the MRF regularizes the result. If the weight is too small then the MRF has no effect, but excessive weights are also detrimental since minima of the pairwise energy correspond to all regions in the image being assigned to the same label. Yu and Clausi [76] developed a principled method for selecting the edge penalty for a pairwise MRF with a univariate GMM for the unary potential, which was extended to the multivariate case by Qin and Clausi [53]. The basic idea is that the edge penalty should be large when the unary potential provides strong class separability so as to provide sufficient regularization, but should decrease in the case of a weak unary potential. Therefore an adaptive edge weight was proposed which scales according to a measure of class separability. The class separability measure was taken to be the minimum of the two-class Fisher criterion over all pairs of classes j and k as shown in (4.10).

$$J = \min_{jk} \operatorname{trace} \left(S_{W_{jk}}^{-1} S_{B_{jk}} \right) \tag{4.10}$$

In the above, $S_{W_{jk}}$ and $S_{B_{jk}}$ respectively represent the within and between-class scatter matrices as defined in (4.11) and 4.12.

$$S_{W_{jk}} = \frac{N_j}{N_j + N_k}\boldsymbol{\Sigma}_j + \frac{N_k}{N_j + N_k}\boldsymbol{\Sigma}_k \tag{4.11}$$

$$S_{B_{jk}} = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)^T \tag{4.12}$$

The MRF edge penalty is then determined by 4.13, where C_1 and C_2 are tuning parameters. C_2 controls the sensitivity of the edge weight to changes in the Fisher criterion and C_1 scales the result.

$$\beta = C_1 \frac{J/C_2}{1 + J/C_2} \tag{4.13}$$

When the GMM is replaced with a mixture regression, it no longer makes sense to consider a global Fisher criterion since the class separability varies across the image. I propose a simple extension which replaces the between-class scatter matrix of (4.12) with the nonstationary version shown in (4.14).

$$S_{B_{ijk}} = (\mathbf{g}_{ij} - \mathbf{g}_{ik})(\mathbf{g}_{ij} - \mathbf{g}_{ik})^T \tag{4.14}$$

The elements of $S_{B_{ijk}}$ scale according to the difference between the trends for classes j and k at the location of region i, and are then weighted by the inverse of the within-class scatter matrix to obtain the modified Fisher criterion shown in (4.15). In contrast to the global Fisher criterion, the modified version provides local estimates of class separability, having high values where the classes are well separated and low values where they have a large overlap.

$$J_i = \min_{jk} \text{trace}\left(S_{W_{jk}}^{-1} S_{B_{ijk}}\right) \tag{4.15}$$

The modified Fisher criterion is then used to construct a spatially varying edge strength parameter

$$\beta_i = \beta_0 \left(\frac{J_i}{\bar{J}}\right)^{\gamma} \tag{4.16}$$

where β_0 is a scaling parameter, $\bar{J} = \sum_i J_i/N$ is the mean of the local Fisher criterion across the whole image, and γ is a positive constant which controls the sensitivity of the edge strength to the local class separability. This parameterization is chosen over the parameterization from 4.13 to allow a larger variation in edge penalty across the image; in practice I find that $\gamma = 2.0$ works well.

4.2 Results and Discussion

4.2.1 Evaluation of Least Squares MLR for Ice-Water Segmentation

This section provides a qualitative evaluation of the results of the MLR model with a linear trend function for ice-water segmentation. Fitting is performed using standard least squares and no MRF regularization is used. Initial results support the findings of Cristea et al. [11] who achieved good ice-water separation using the same model for class-dependent backscatter decay. An example segmentation from the linear-trend MLR is shown in Figure 4.1 and is contrasted with the result from a GMM. Plots of the incidence angle dependent backscatter intensities for each channel are overlaid with the obtained trend fits for comparison. It is clear that the incidence angle induced nonstationarity prevents the GMM from separating the classes, while the MLR achieves a good fit.

While the MLR performs very well on some scenes, for others the convergence is unreliable; different model initializations converge to different results, some successful and others not. For these scenes models initialized with randomly typically converge to one of a handful of basins of attraction. Certain conditions in particular prove adversarial to MLR convergence. One difficult scenario occurs during advanced melt conditions, where moist snow cover and melt ponds attenuate backscatter. This is particularly detrimental for the HV polarization where backscatter intensity during melt approaches the sensor noise floor. An example of unreliable convergence for a scene with low HV contrast is shown in Figure 4.2 for a scene taken on July 30, 2010. Another confounding factor in this scene is the presence of very dark regions of open water in the HH polarization; these may be caused by local wind conditions or by wave attenuation by sub-pixel ice floes.

A second scenario which proves challenging for the convergence of mixture regression occurs when the incidence angle ranges for which ice and water are present are nearly disjoint. Mixture regression works best when there is an ice-water boundary that separates the scene horizontally (across all incidence angles). It is worth noting that the favourable scenario is over-represented in the dataset used in this study; narrow channels of open water tend to form along the coastline of the Beaufort Sea, appearing horizontal in the swaths of polar-orbiting satellites. This results in scenes such as the one shown in Figure 4.1 where both water and ice cover the full range of incidence angles. An example of poor convergence for a scene where the ice-water boundary crosses the scene vertically is shown in Figure 4.3.

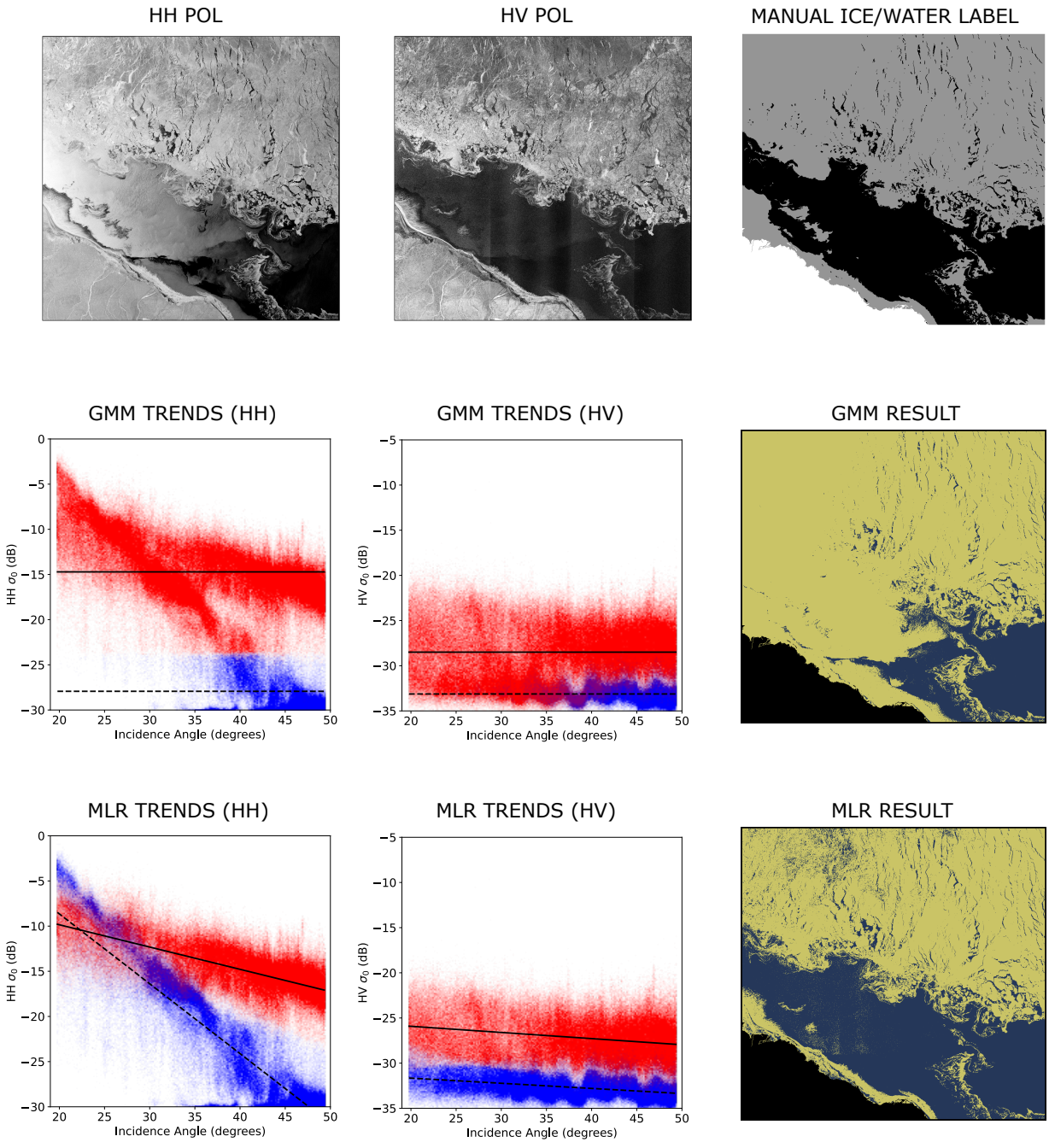

Figure 4.1: Comparison of GMM with MLR for scene 20101027

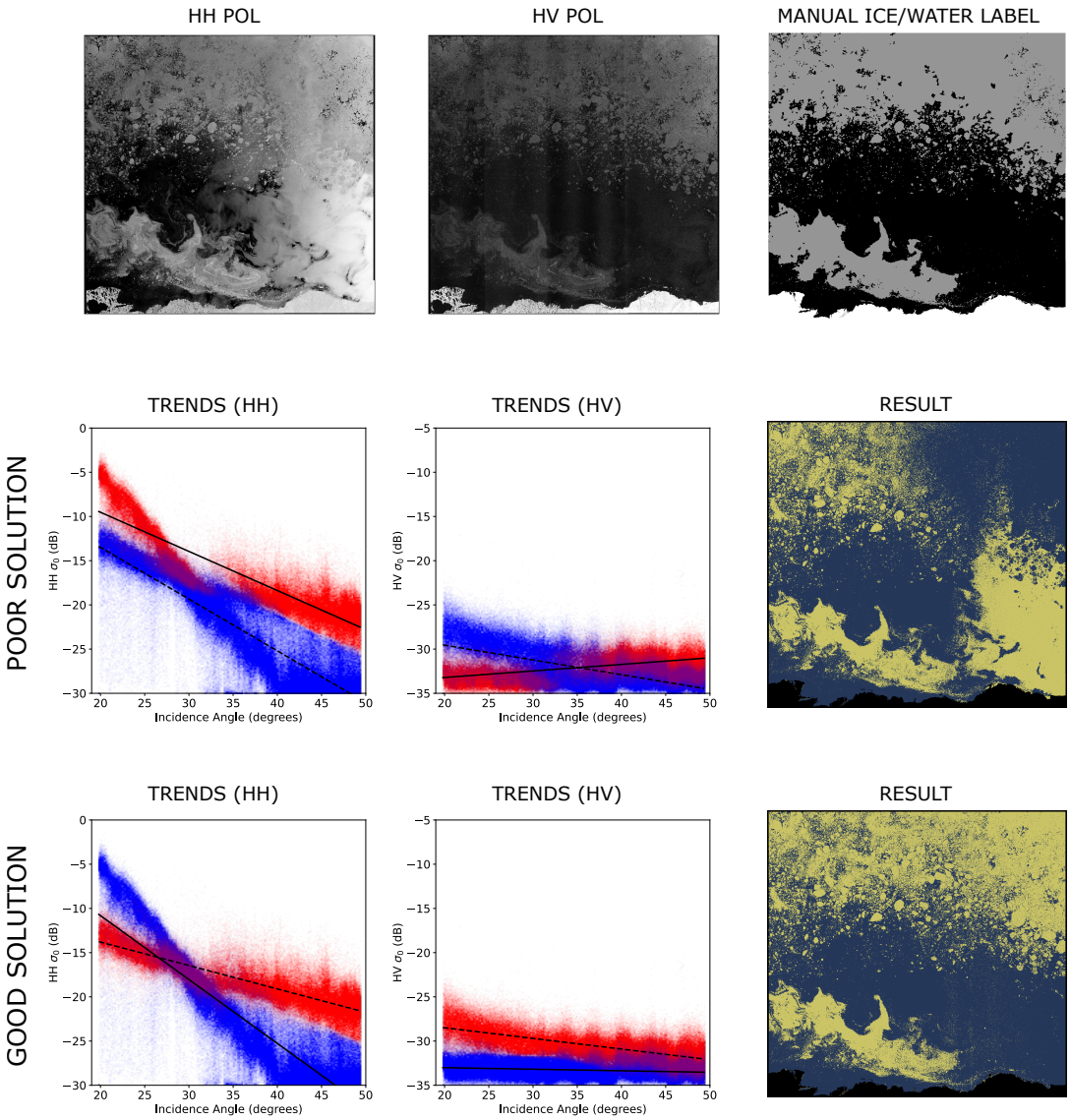

Figure 4.2: Least squares mixture regression results for scene 20100730 corresponding to two different random initializations. Middle row - a solution obtaining poor ice-water separation (pixel accuracy 57.3%). Bottom row - a solution acheiveing good ice-water separation (pixel accuracy 90.7%). A weather station in Utquiagvik, AK (approximately 140 km from the nearest point in the scene) reported a 24h-average air temperature of 8.2 degrees Celsius on the day of observation, and 24h average temperatures exceeded 6.3 degrees Celsius during the preceeding week indicating an advanced stage of melting.

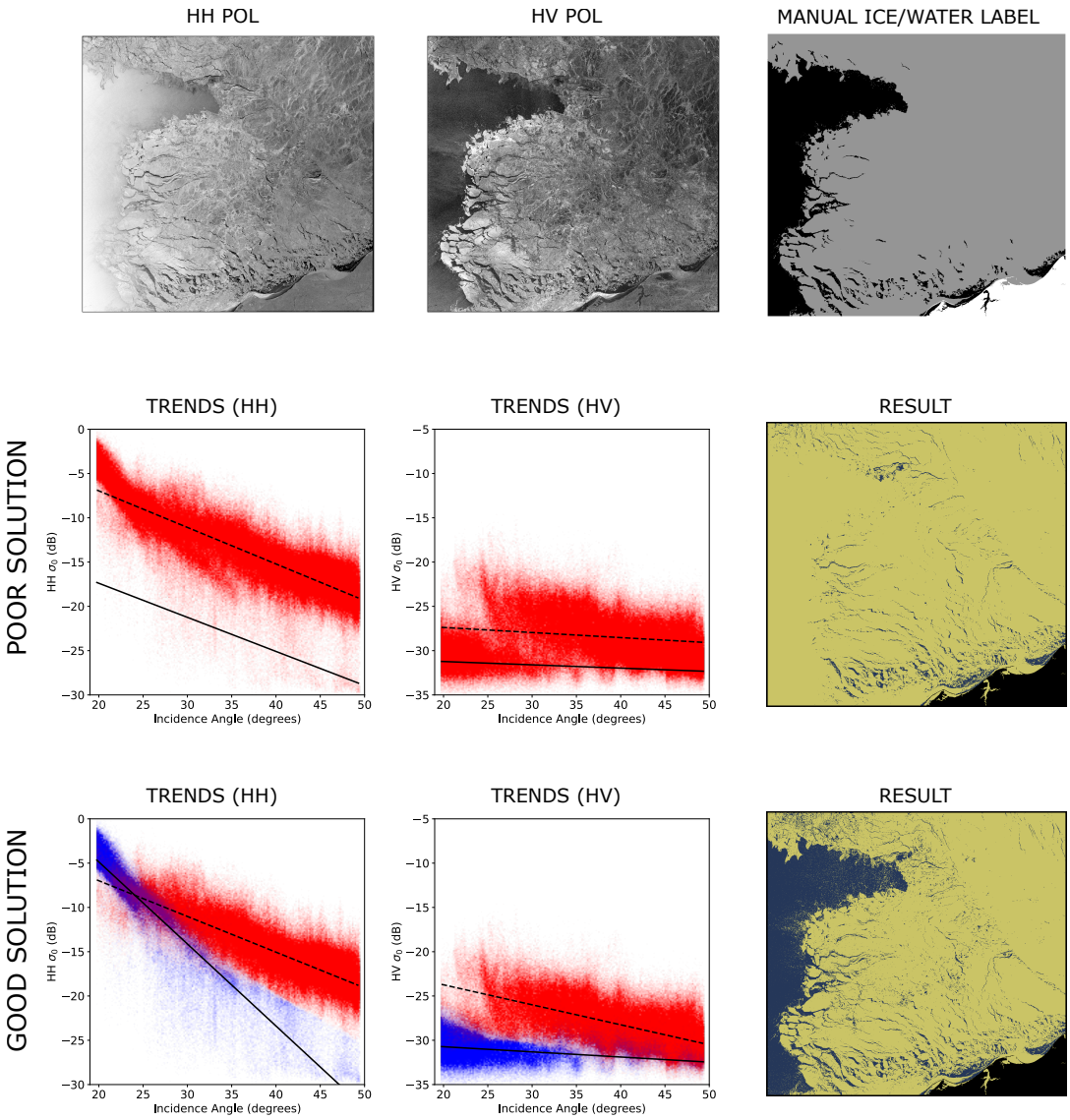

Figure 4.3: Least squares mixture regression results for scene 20101114 corresponding to two different random initializations. Middle row - a solution obtaining poor ice-water separation (pixel accuracy 80.2%). Bottom row - a solution achieveing good ice-water separation (pixel accuracy 94.7%).

4.2.2 Improving the Reliability of Mixture Regression

In this section, I explore schemes to overcome the reliablity problems encountered by the standard mixture regression in section 4.2.1 by incorporating robust regression and deterministic annealing. I conduct experminents on the two challenging scenes described above (20100730 and 20101114). In general, if convergence is unreliable using the linear trend model then it is unreliable for the polynomial trend model as well, so only the linear trend model is considered in these experiments. The experimental configurations for the regression are as follows:

- *Least Squares*: The trend coefficients $\mathbf{w}_k$ are obtained using an ordinary least squares regression as shown in (4.3).
- *Robust Regression*: The trend coefficients are obtained with a robust objective and the IRLS algorithm described in section 3.3.1. I use the Huber loss with robustness parameter δ, and consider the three settings $\delta = 0.001$, $\delta = 0.01$, and $\delta = 0.1$. Note that δ should be chosen in proportion to the range of the input data since it is used as an outlier threshold on the regression residuals; the choices used here apply to input values normalized to the range $[0, 1]$.

For each of the variants described above, I perform experiments with two types of EM fitting procedures:

- *Constant Temperature EM*: Each configuration is run for 50 EM iterations using a constant temperature parameter T. I perform several experiments for each configuration, sweeping over temperature values between 0.1 and 1.0.
- *Deterministic Annealing*: I fit the models using determinisitic annealing following the sigmoidal annealing schedule detailed in section 3.3.2 with 50 iterations. I set $\alpha_1 = 25$ and $\alpha_2 = 4$ since these settings seem to work well in practise; systematic investigation of annealing schedules is beyond the scope of this work.

For each pairing of a model configuration with a fitting procedure I run 50 trials, each starting from a random set of initial labels. A seed is used to generate the initial random labels so that all experiments use the same 50 starting points. Models are evaluated on their pixel accuracy against manual labels for each scene (the percentage of pixels for which the correct ice/water label is obtained).

Experimental results are presented in Figure 4.4. The reliablility problems with the least squares method are apparent from the left-most panes. The model converges to a

satisfactory solution for some of the runs, and the performance when this occurs is quite good (pixel accuracy 90.7% for scene 20100730 and 94.7% for 20101114). However, the least squares model reaches this solution for only 43% of the tested initializations for scene 20100730 and for 46% of the initializations for scene 20101114. Other runs are drawn to various lower-accuracy basins of attraction. Varying the temperature parameter has only a minor effect on the consistency of the solutions for least squares, and employing deterministic annealing similarly shows no benefit.

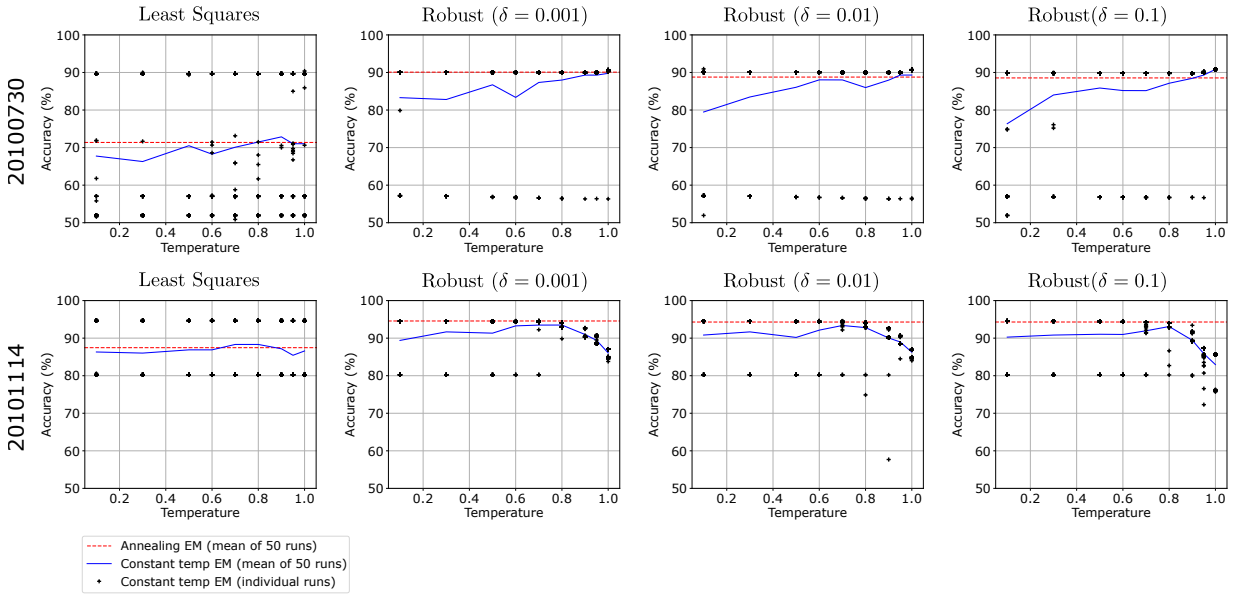

Figure 4.4: Comparison of results for least squares and robust regression, both for constant temperature EM and annealing EM. Top row shows results for scene 20100730 and middle row shows results for scene 20101114. In all cases, robust regression converges to high-accuracy solutions more frequently than least squares. Annealing EM leads to better convergence for robust regression but not for least squares.

Robust regression results for various values of δ are shown in the rightmost 3 panes of Figure 4.4. In most cases, similar basins of attraction exist for the robust regression models as for least squares. The maximum accuracy over 50 trials is thus no better for robust regression than least squares, but the high-accuracy solutions are achieved much more consistently as is shown by the increased average accuracy. Performance is similar across the three values of δ that were tested. The lack of significant improvement when

decreasing δ below 0.1 indicates that outliers with large residuals are primary contributors to the convergence difficulties faced by the least squares procedure.

Unlike the least squares case, the performace for robust regression depends on the choice of temperature. However, the optimal temperature varies between scenes. This is likely due to the fact that in the trend estimation step for robust mixture regression each data point is reweighted by both the expected class labels and an outlier rejection term (see 3.8 in section 3.3.1). The weighting terms interact differently depending on the temperature and the severity of outlier contamination in the scene. Experimental results show that deterministic annealing brings improved consistency while also avoiding the need to tune the temperature parameter on a scene-by-scene basis. For both scenes, the combination of robust regression with $\delta = 0.001$ and deterministic annealing reaches the highest-accuracy solution in all 50 trials.

4.2.3 Nonlinear Dependence on Incidence Angle

While a linear trend function provides a reasonable approximation of the incidence angle dependence of SAR backscatter for water and sea ice, a more flexible model for the trend can provide a better fit. In this section, I experiment with polynomial trend models from degree 1 (linear trend) to 5. I use the model settings determined in the previous section to give the most consistent results (robust regression with $\delta = 0.001$ and deterministic annealing), and conduct experiments on the 25 labelled scenes in the dataset where both ice and water are present. I find that the presence of outliers reduces the quality of the polynomial fits when the number of clusters is set to 2, so I use 3 clusters and evaluate the pixel accuracy using the best mapping of the clusters to the two output classes (ice and water)[3].

The trends for scene 20100730 are shown in Figure 4.5. It is apparent that the polynomial model provides a better representation of the trend, especially for the water class. The main source of nonlinearity for this scene appears to be low backscatter values in the far range which fall below the noise-equivalent sigma zero (NESZ) of the SAR sensor.

Interestingly, the improved fit for the polynomial trend model does not bring a corresponding increase in classification accuracy. In fact, increasing the trend order beyond the linear case corresponds to a slight decrease in the average accuracy over the 25 scenes. Visually, the segmentation results obtained using nonlinear trends are very similar to those obtained with the linear trend model. The lack of improvement when increasing the trend

[3] "Best" in this context means the mapping which gives the highest pixel accuracy.

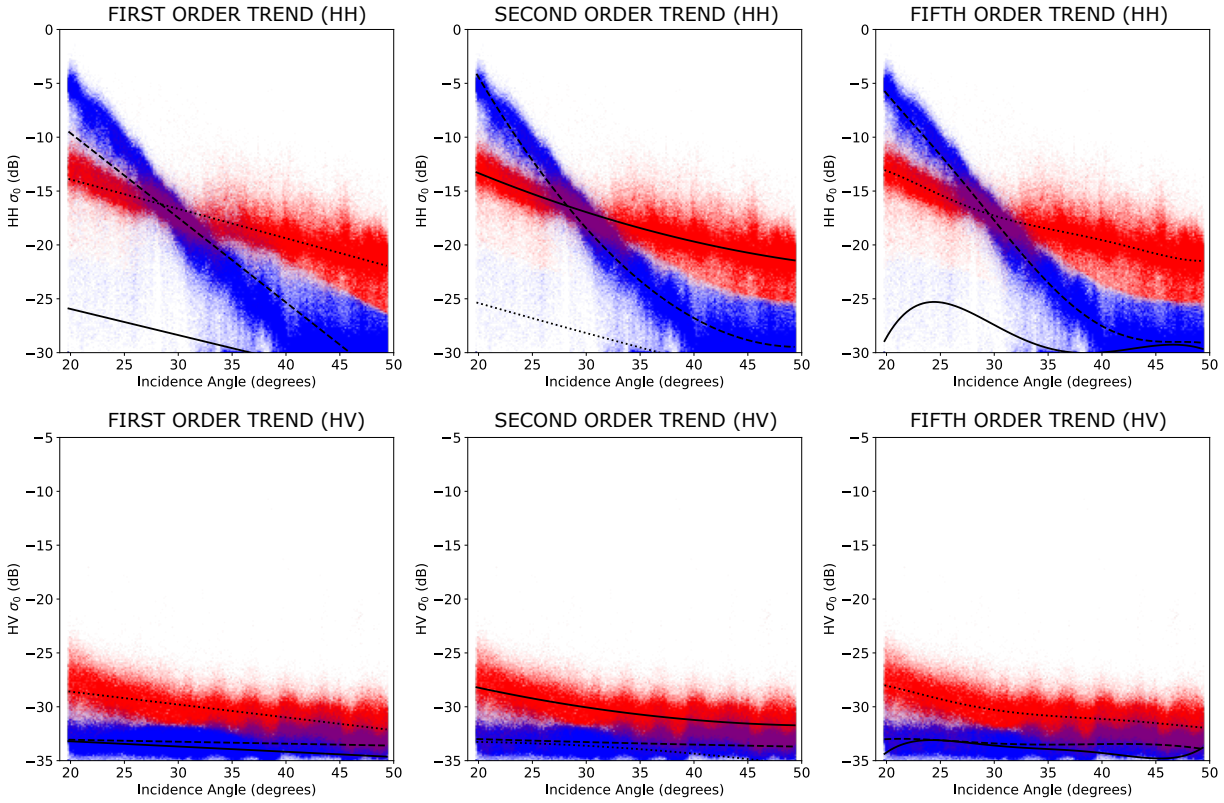

Figure 4.5: First, second and fifth order polynomial trend fits for scene 20100730. Note that

order can be attributed to the fact that nonlinear fit deviates from the linear fit mostly at incidence angles where ice and water backscatter values are well separated. Although including nonlinear incidence angle terms does not help the classification performance, doing so may still be beneficial if the trends themselves are of interest. Results for the average pixel accuracy across the 25 scenes are shown in table 4.1. Accuracies for all scenes are given in table C.1 in Appendix C.

Table 4.1: Comparison of polynomial models for the incidence angle effect.

Trend Order	**1**	**2**	**3**	**4**	**5**
Pixel Accuracy (%)	92.8	92.4	92.4	92.7	92.2

4.2.4 MRF Regularization

This section presents the results of adding the MRF regularization scheme presented in section 4.1.3 to the mixture regression models presented above. While mixture regression without MRF regularization often obtains high pixel accuracy, the results are noisy. This effect is particularly strong at incidence angles where the HH backscatter distributions for ice and water overlap, leading to reduced class separability. An example is shown in the center panel of Figure 4.6, where banding noise towards the leftmost third of the scene causes misclassified pixels. As can be seen, the MRF alleviates this issue, leading to accurate results in the regions where there is homogeneous cover of ice or water. The MRF results in some loss of detail, causing some of the smaller leads to be erased; however, the majority of the lead structure is preserved.

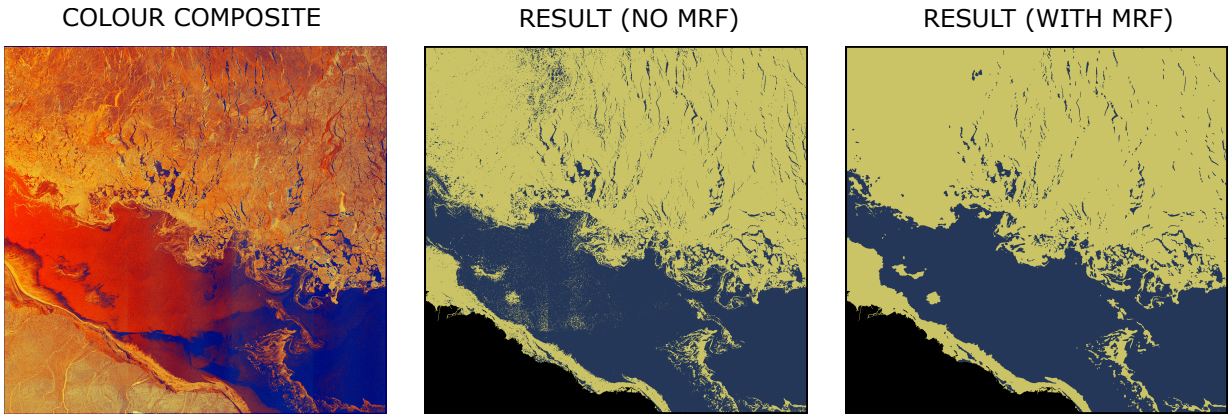

Figure 4.6: Effect of MRF regularization on the segmentaton result. The leftmost image is a colour composite of the original SAR image with R=HH, G=HV, B=HV/HH. All bands are stretched for better viewing.

The impact of employing the adaptive nonstationary edge penalty is illustrated in Figure 4.7. The degree of smoothing introduced by the MRF varies across the scene when a constant edge penalty is used as a result of the variable contrast introduced by the incidence angle effect. The inconsistent regularization allows noise to persist on the left side of the scene while over-smoothing details towards the right side. The adaptive edge penalty applies stronger regularization at incidence angles where the contrast is high and reduces the regularization strength where the contrast is low. This results in a more consistent smoothing effect across the scene. Although the effect is relatively minor and is not accompanied by an increase in measured accuracy, it constitutes a qualitative improvement.

4.2.5 Integrating Mixture Regression into a Supervised Classification Scheme

This research has considered sea ice classification using mixture regression as an unsupervised clustering technique. However, it is possible to pursue a hybrid approach which combines mixture regression with essentially any supervised classifier. The hybrid approach plays to the strengths of each component. While it is well known that both tonal and textural information are beneficial for the classification of sea ice [10, 71], the mixture regression models considered so far have left texture features unleveraged. On the other hand, many supervised classifiers can make use of texture features, but they tend to neglect the information from raw backscatter intensities due to their inability to model nonstationarity.

As a proof of concept, I explore a hybrid model which combines mixture regression with an off-the-shelf U-Net convolutional neural network (CNN) [59]. No particular effort is made to optimize the performance of the CNN on this dataset, but its performance proves sufficient to demonstrate the benefits of the hybrid approach; alternative models could presumably be substituted to similar effect. A leave-one-out approach [38] is used for CNN training where the model evaluated on each scene is trained on the remaining 34 scenes.

Denote by $f_s^{CNN}(k)$ the output of the CNN for class k at pixel s, which is used without applying any activation function. The hybrid segmentation approach is obtained by modifying the unary potential of the MRF as shown in (4.17), where $U_i^{MR}(k)$ is the mixture regression unary from 4.9 and λ is a parameter which controls the relative weights of the CNN and the mixture regression. For the results presented below I use $\lambda = 1$ to give the models equal weight; Differing values for λ tend to bias the solution heavily towards the output of only one of the models.

$$U_i(k) = U_i^{MR}(k) - \lambda \sum_{s \in \Omega_i} f_s^{CNN}(k) \tag{4.17}$$

Accuracy results for the hybrid model are presented for all the scenes in the study dataset in table C.2 in Appendix C. The average pixel accuracy across all scenes of the hybrid model is 97.5%, surpassing both the CNN (96.2%) and the unsupervised mixture regression (92.8%, evaluated on the 25 scenes with both ice and water present). On individual scenes, the hybrid model consistently outperforms the CNN, achieving a higher accuracy for 32 of the 35 scenes.

Figure 4.8 compares the results of the hybrid model with the CNN and the unsupervised mixture regression for scene 20101017. In this scene, relying on backscatter intensity alone proved insufficient to obtin a satisfactory result; the unsupervised mixture regression obtained the lowest accuracy (73.7%) on this scene of the 25 on which it was tested. This is probably due to very heterogeneous ice backscatter across the scene and an unusual pattern across the open water region which may be related to wind or thin ice formation. The CNN fares much better, but suffers from artefacts and poor segmentation alignment to natural edges in the image. The hybrid model improves on the CNN result both qualitatively and quantitatively.

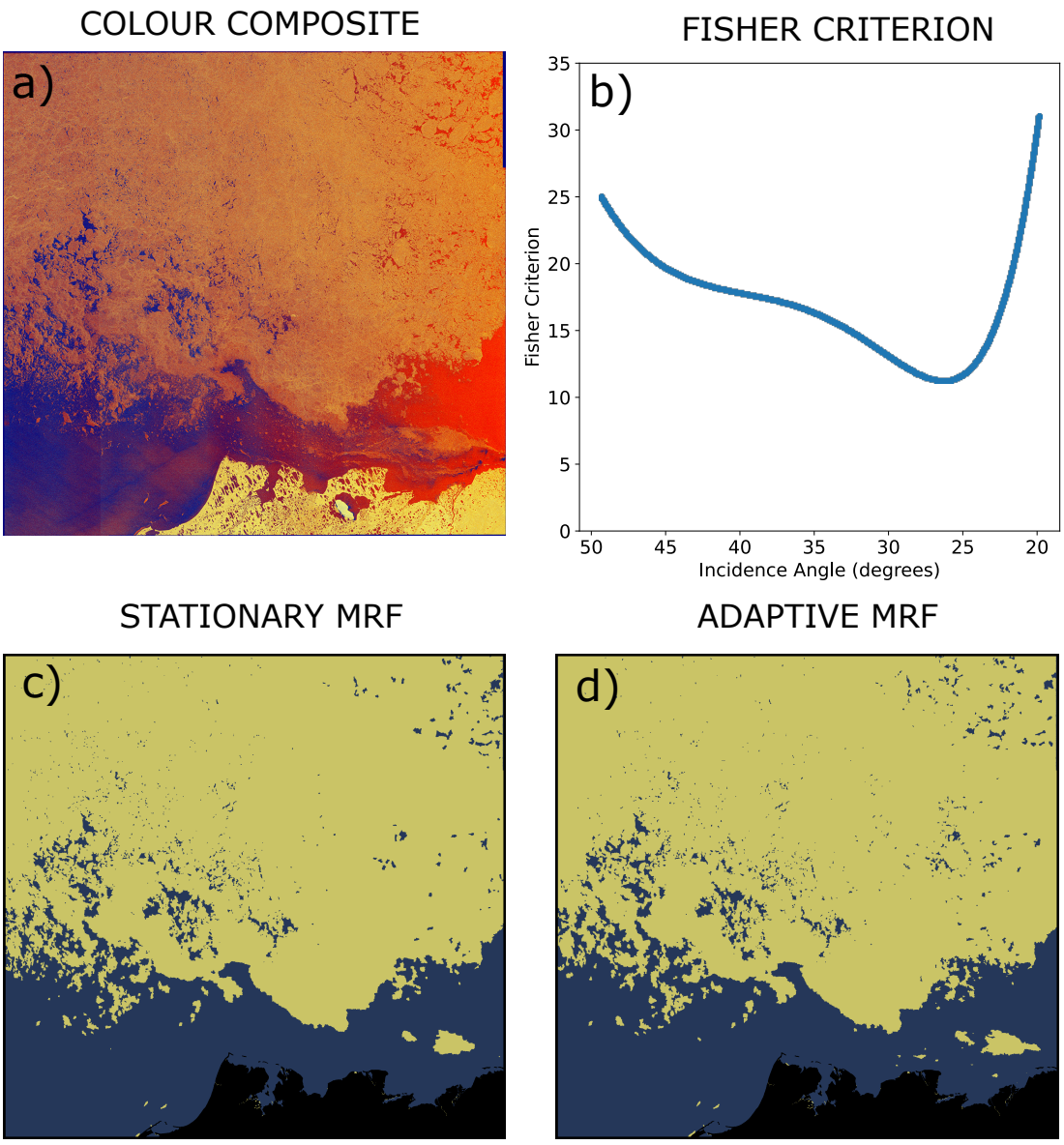

Figure 4.7: Comparison of MRF regularization with constant edge penalty ($\beta = 20$) and adaptive edge penalty ($\beta_0 = 20, \gamma = 2.0$) for scene 20100710. (a) Colour composite of the original SAR image with R=HH, G=HV, B=HV/HH. (b) Value of the Fisher criterion plotted against the incidence angle, showing low values in the center of the scene where constrast is low and higher values at the edges. (c) Mixture regression result with constant edge penalty. (d) Mixture regression result with adaptive edge penalty. Best viewed zoomed in.

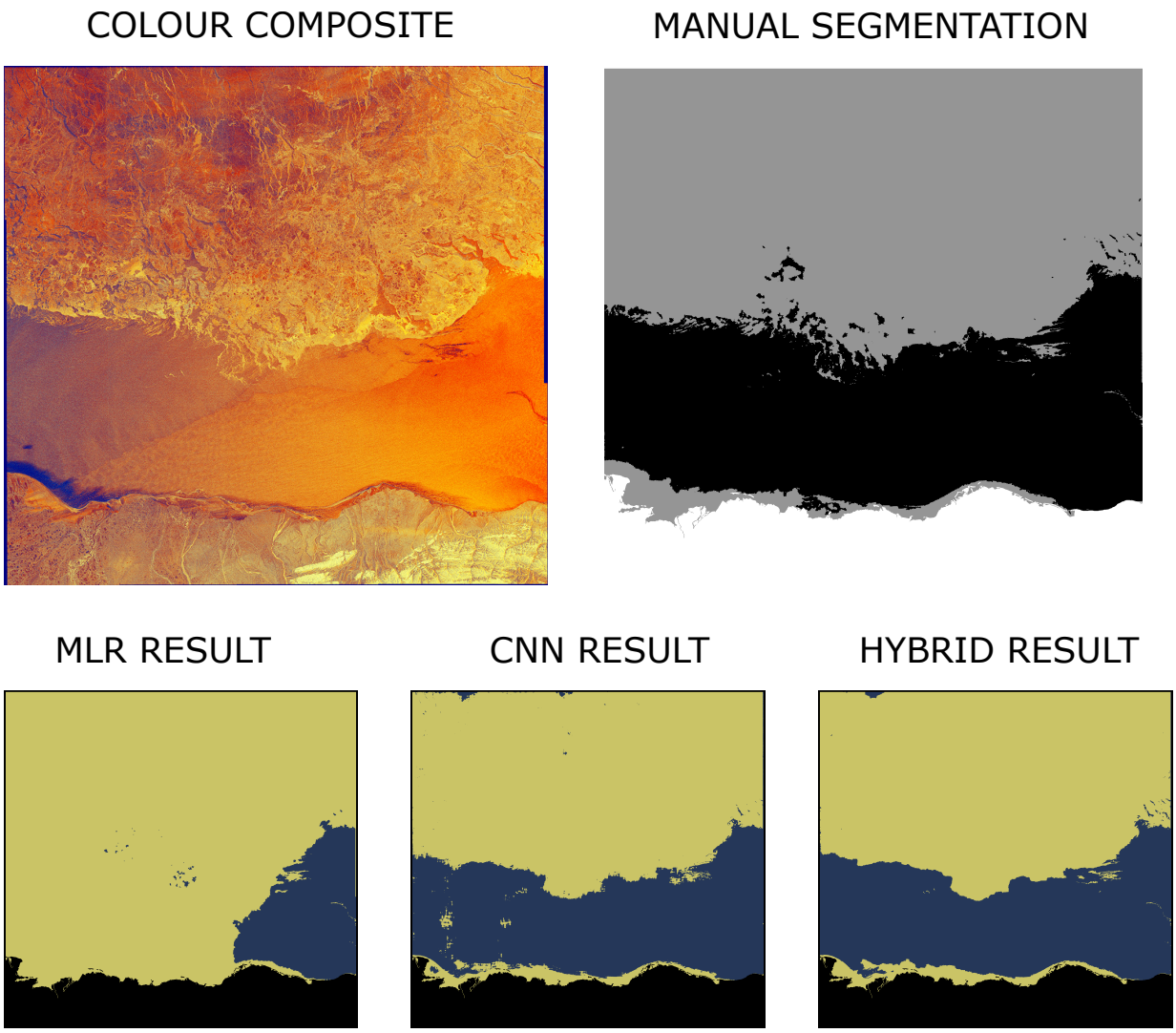

Figure 4.8: Comparison of segmentation results for different models on scene 20101017. The bottom row shows results for mixture regression (pixel accuracy 73.7%), CNN (pixel accuracy 94.8%), and the hybrid CNN-mixture regression model (pixel accuracy 95.8%).

Chapter 5

Conclusions

5.1 Summary

In this thesis, I have advocated the use of mixture regression models as a simple yet effective method for segmenting nonstationary imagery. I presented a general framework for nonstationary segmentation problems based on the estimation of class-dependent trends which describe the variability of image features across covariates. I then developed mixture regression models tailored for the segmentation of sea ice based on nonstationary SAR backscatter intensity, focusing on the variability of backscatter distributions introduced by incidence angle dependencies.

The experiments presented in chapter 4 make a strong case for the effectiveness and flexibility of mixture regression. Although mixture regression fitting using regular least squares suffered from unreliable convergence, I showed that this problem can be resolved using a combination of robust estimation and a deterministic annealing. The resulting model was able to capture the nonstationary backscatter distributions present in SAR sea ice imagery and to to obtain accurate, high-resolution ice maps based on the modeled distributions. Both linear and nonlinear incidence angle dependencies were considered. Although including nonlinear incidence angle dependence did not increase the overall classification accuracy, it led to a better representation of the class-dependent backscatter distributions.

Another strength of mixture regression models is their ease of incorporation with other models. The addition of MRF regularization removed noise while maintaining fine structures such as leads. Beyond the unsupervised setting, I showed that mixture regression can

be combined with supervised classification techniques to obtain higher performance than either model on its own. This was demonstrated by combining mixture regression with a CNN, acheiving an average pixel accuracy of 97.5% over the 35 RADARSAT-2 scenes considered in this study. Several promising directions exist to extend the work presented in this thesis, which are outlined in the following section.

5.2 Future Work

Use of mixture regression to model nonstationary texture features. Many sea ice classification methods employ texture features. The incidence angle dependence of texture features has attracted some recent studies [62, 26], but further work is needed to integrate this into effective classifiers.

Extending mixture regression models for fully polarimetric SAR. This work considered the modeling of dual-polarized SAR data with nonstationary mixtures of Gaussians. Unsupervised segentation of sea ice from Quad-Pol [75] and Compact-Pol [22] SAR has also been accomplished using classifiers based on Wishart statistics. These methods may benefit from the integration of mixture regression techniques to account for incidence angle effects and other sources of nonstationarity.

Integrating auxiliary data sources On top of incidence angle dependencies, numerous other factors affect the SAR backscatter intensity in sea ice imagery. It would be interesting to consider the use of auxiliary variables (such as surface temperatures, wind fields, and other relevant quantities) as covariates with respect to which backscatter may exhibit class-dependent trends. To make full use of this approach it is likely that a co-segmentation approach would be appropriate where several scenes are simultaneously segmented with trend fits computed for all of the scenes jointly.

Region Merging. I considered mixture regression in the same region-based framework employed by the IRGS algorithm [76, 53], which employs a region-merging strategy to grow regions over the course of segmentation. However, I did not consider region merging in this work. Incorporating a region merging step into the mixture regression models considered in this thesis may enable more effective spatial context modeling.

References

[1] Andrew Adams, Jongmin Baek, and Myers Abraham Davis. Fast high-dimensional filtering using the permutohedral lattice. *Computer Graphics Forum*, 29(2):753–762, May 2010.

[2] Wiebke Aldenhoff, Leif E. B. Eriksson, Yufang Ye, and Celine Heuze. First-year and multiyear sea ice incidence angle normalization of dual-polarized sentinel-1 SAR images in the beaufort sea. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:1540–1550, 2020.

[3] Nazanin Asadi, K. Andrea Scott, Alexander S. Komarov, Mark Buehner, and David A. Clausi. Evaluation of a neural network with uncertainty for detection of ice and water in SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13, 2020.

[4] Xiuqin Bai, Weixin Yao, and John E. Boyer. Robust fitting of mixture regression models. *Computational Statistics & Data Analysis*, 56(7):2347–2359, Jul 2012.

[5] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer, New York, 1981.

[6] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information science and statistics. Springer, New York, 2006.

[7] Reid M Bixler. Sparse matrix belief propagation. Master's thesis, Virginia Polytechnic Institute and State University, 2018.

[8] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.

[9] J. Alec Casey, Stephen E.L. Howell, Adrienne Tivy, and Christian Haas. Separability of sea ice types from wide swath C- and L-band synthetic aperture radar imagery acquired during the melt season. *Remote Sensing of Environment*, 174:314–328, Mar 2016.

[10] David A. Clausi. *Texture Segmentation of SAR Sea Ice Imagery*. PhD thesis, University of Waterloo, 1996.

[11] Anca Cristea, Jeroen van Houtte, and Anthony P. Doulgeris. Integrating incidence angle dependencies into the clustering-based segmentation of SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2925–2939, 2020.

[12] Richard D. De Veaux. Mixtures of linear regressions. *Computational Statistics & Data Analysis*, 8(3):227–245, Nov 1989.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, Sep 1977.

[14] C. Elachi. *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE Press, 1987.

[15] S Evans. Dielectric properties of ice and snow - a review. *Journal of Glaciology*, 5(42):773–792, 1965.

[16] Susana Faria and Gilda Soromenho. Fitting mixtures of linear regressions. *Journal of Statistical Computation and Simulation*, 80(2):201–225, Feb 2010.

[17] G. Franceschetti and R. Lanari. *Synthetic Aperture Radar Processing*. Electronic Engineering Systems. Taylor & Francis, 1999.

[18] Karen E. Frey, G.W.K. Moore, Lee W. Cooper, and Jacqueline M. Grebmeier. Divergent patterns of recent sea ice cover across the bering, chukchi, and beaufort seas of the pacific arctic region. *Progress in Oceanography*, 136:32–49, 2015.

[19] T. Geldsetzer, J.B. Mead, J.J. Yackel, R.K. Scharien, and S.E.L. Howell. Surface-based polarimetric C-band scatterometer for field measurements of sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3405–3416, Nov 2007.

[20] T Geldsetzer and J J Yackel. Sea ice type and open water discrimination using dual co-polarized C-band SAR. *Canadian Journal of Remote Sensing*, 35(1):12, 2009.

[21] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, Nov 1984.

[22] Mohsen Ghanbari, David Clausi, and Linlin Xu. CP-IRGS: A region-based segmentation of multilook complex compact polarimetric SAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:6559–6571, 06 2021.

[23] Kenneth M Golden, D Borup, Margaret Cheney, E Cherkaeva, Michael S Dawson, Kung-Hau Ding, Adrian K Fung, David Isaacson, SA Johnson, Arthur K Jordan, et al. Inverse electromagnetic scattering models for sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1675–1704, 1998.

[24] D.J. Griffiths. *Introduction to Electrodynamics*. Cambridge University Press, 2017.

[25] Kevin J Grimm, Gina L Mazza, and Pega Davoudzadeh. Model selection in finite mixture models: A k-fold cross-validation approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2):246–256, 2017.

[26] Wenkai Guo, Polona Itkin, Suman Singha, Anthony Paul Doulgeris, Malin Johansson, and Gunnar Spreen. *Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture*. May 2022.

[27] Martti Hallikainen, F Ulaby, and Mohamed Abdelrazik. Dielectric properties of snow in the 3 to 37 ghz range. *IEEE transactions on Antennas and Propagation*, 34(11):1329–1340, 1986.

[28] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[29] D. Haverkamp, L.K. Soh, and C. Tsatsoulis. A dynamic local thresholding technique for sea ice classification. In *Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium*, pages 638–640, Tokyo, Japan, 1993. IEEE.

[30] Peter J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:492–518, 1964.

[31] Gabriel Huerta, Wenxin Jiang, and Martin A. Tanner. Time series modeling vis hierarchical mixtures. *Statistica Sinica*, 13(4):1097–1118, 2003.

[32] Chi Jin, Yuchen Zhang, Sivaraman Balakrishnan, Martin J. Wainwright, and Michael I. Jordan. Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 4123–4131, Red Hook, NY, USA, 2016.

[33] A. Malin Johansson, Camilla Brekke, Gunnar Spreen, and Jennifer A. King. X-, C-, and L-band SAR signatures of newly formed sea ice in arctic leads during winter and spring. *Remote Sensing of Environment*, 204:162–180, Jan 2018.

[34] Juha Karvonen. Baltic sea ice concentration estimation from C-band dual-polarized SAR imagery by image segmentation and convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2022.

[35] Alexander S. Komarov, Dustin Isleifson, David G. Barber, and Lotfollah Shafai. Modeling and measurement of C-band radar backscatter from snow-covered first-year sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):4063–4078, Jul 2015.

[36] Jeongyeol Kwon and Constantine Caramanis. EM converges for a mixture of many linear regressions. (arXiv:1905.12106), Nov 2019. arXiv:1905.12106 [cs, stat].

[37] Jeongyeol Kwon, Wei Qian, Constantine Caramanis, Yudong Chen, and Damek Davis. Global convergence of EM algorithm for mixtures of two component linear regression. (arXiv:1810.05752), May 2019. arXiv:1810.05752 [cs, math, stat].

[38] Steven Leigh, Zhijie Wang, and David A. Clausi. Automated ice-water classification using dual polarization SAR satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5529–5539, Sep 2014.

[39] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. An optimal decision-tree design strategy and its application to sea ice classification from SAR imagery. *Remote Sensing*, 11(13):1574, Jul 2019.

[40] Johannes Lohse, Anthony P. Doulgeris, and Wolfgang Dierking. Mapping sea-ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle. *Annals of Glaciology*, 61(83):260–270, Dec 2020.

[41] MacDonald, Dettweiler and Associates Ltd. RADARSAT-2 beam modes. https://www.asc-csa.gc.ca/eng/multimedia/search/image/watch/4249. Accessed: 2022-07-22.

[42] MacDonald, Dettweiler and Associates Ltd. RADARSAT-2 product description, rev. 14. https://earth.esa.int/eogateway/documents/20142/0/Radarsat-2-Product-description.pdf/f2783c7b-6a22-cbe4-f4c1-6992f9926dca. Accessed: 2022-07-22.

[43] Mallik S. Mahmud, Torsten Geldsetzer, Stephen E. L. Howell, John J. Yackel, Vishnu Nandan, and Randall K. Scharien. Incidence angle dependence of HH-polarized C- and L-band wintertime backscatter over arctic sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11):6686–6698, Nov 2018.

[44] Matthew J. Menne, Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910, Jul 2012.

[45] M.J. Menne, I. Durre, B. Korzeniewski, S. McNeill, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E. Gleason, and T.G. Houston. Global historical climatology network - daily (GHCN-Daily), version 3.12, 2012. Accessed July 2022.

[46] Marc Mezard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., USA, 2009.

[47] Suvadip Mukherjee and Scott T. Acton. Region based segmentation in presence of intensity inhomogeneity using legendre polynomials. *IEEE Signal Processing Letters*, 22(3):298–302, 2015.

[48] E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, Jan 1964.

[49] S.V. Nghiem and C. Bertoia. Study of multi-polarization c-band backscatter signatures for arctic sea ice mapping with future satellite SAR. *Canadian Journal of Remote Sensing*, 27(5):387–402, Oct 2001.

[50] Shuhratchon Ochilov and David A Clausi. Operational SAR sea-ice image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):12, 2012.

[51] Robert G. Onstott. *SAR and scatterometer signatures of sea ice*, volume 68, pages 73–104. American Geophysical Union, Washington, D. C., 1992.

[52] Judea Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the Second AAAI Conference on Artificial Intelligence*, AAAI'82, pages 133–136. AAAI Press, 1982.

[53] A K Qin and David A Clausi. Multivariate image segmentation using semantic region growing with adaptive edge penalty. *IEEE Transactions on Image Processing*, 19(8):2157–2170, Aug 2010.

[54] Richard E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880, Dec 1958.

[55] Joaquin Quiñonero Candela and Carl Edward Rasmussen. A unifying view of sparse approximate gaussian process regression. *J. Mach. Learn. Res.*, 6:1939–1959, dec 2005.

[56] Jagath C. Rajapakse and Frithjof Kruggel. Segmentation of mr images with intensity inhomogeneities. *Image and Vision Computing*, 16(3):165–180, 1998.

[57] Rudolf Ressel, Anja Frost, and Susanne Lehner. A neural network-based classification for sea ice types on X-band SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7):3672–3680, Jul 2015.

[58] John R. Rice and Karl H. Usow. The lawson algorithm and extensions. *Mathematics of Computation*, 22(101):118–127, 1968.

[59] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[60] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, Nov 1998.

[61] Kenneth Rose, Eitan Gurewitz, and Geoffrey Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11(9):589–594, Sep 1990.

[62] Randall Kenneth Scharien and Sasha Nasonova. Incidence angle dependence of texture statistics from sentinel-1 HH-polarization images of winter arctic sea ice. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[63] Bernd Scheuchl, Dean Flett, Ron Caves, and Ian Cumming. Potential of RADARSAT-2 data for operational sea ice monitoring. *Canadian Journal of Remote Sensing*, 30(3):448–461, Jan 2004.

[64] J. Stroeve and W. N. Meier. Sea ice trends and climatologies from SMMR and SSM/I-SSMIS, 2018. Accessed June 10, 2022.

[65] Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366, Feb 2007.

[66] A. N. Tikhonov and Vasiliy Yakovlevich Arsenin. *Solutions of ill-posed problems*. Winston Publishing, 1977.

[67] S. Tjuatja, A.K. Fung, and J. Bredow. A scattering model for snow-covered sea ice. *IEEE Transactions on Geoscience and Remote Sensing*, 30(4):804–810, Jul 1992.

[68] Jean-Francois Lemieux Tom Carrieres, Mark Buehner and Leif Toudal Pedersen, editors. *Sea Ice Analysis and Forecasting: Towards an Increased Reliance on Automated Prediction Systems*. Cambridge University Press, 2017.

[69] Costas Tsatsoulis and Ron Kwok. *Analysis of SAR Data of the Polar Oceans: Recent Advances*. Springer, 1998.

[70] Naonori Ueda and Ryohei Nakano. Deterministic annealing EM algorithm. *Neural Networks*, 11(2):271–282, Mar 1998.

[71] Fawwaz Ulaby, F. Kouyate, B. Brisco, and T. H. Williams. Textural information in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(2):235–245, Mar 1986.

[72] J.J. van Zyl and Y. Kim. *Synthetic Aperture Radar Polarimetry*. JPL Space Science and Technology Series. Wiley, 2011.

[73] Geoffrey S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372, 1964.

[74] WMO. WMO sea ice nomenclature, volume 1 - terminology and codes. Technical report, WMO, 2014. Retreived from https://library.wmo.int/doc_num.php?explnum_id=4651.

[75] Peter Yu, A. K. Qin, and David A. Clausi. Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1302–1317, Apr 2012.

[76] Qiyao Yu and D.A. Clausi. IRGS: Image segmentation using edge penalties and region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2126–2139, Dec 2008.

[77] Qiyao Yu and David A. Clausi. SAR sea-ice image analysis based on iterative region growing using semantics. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):3919–3931, Dec 2007.

[78] Natalia Zakhvatkina, Vladimir Smirnov, and Irina Bychkova. Satellite sar data-based sea ice classification: An overview. *Geosciences*, 9(4), 2019.

APPENDICES

Appendix A

The EM Algorithm for Mixture Models

Fitting a K-component GMM to a dataset $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ involves finding a set of parameters $\boldsymbol{\Theta}^* = \{\pi_k^*, \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\}_{k=1}^K$ corresponding to a maximum of the log-likelihood function of the observed data shown in equation A.1.

$$\log P(\mathbf{X}|\boldsymbol{\Theta}) = \sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{A.1}$$

Although it is simple to analytically obtain maximum-likelihood parameter estimates for a single Gaussian distribution, directly maximizing equation A.1 is intractable, necessitating the EM approach. Let $\{\mathbf{z}_i\}_{i=1}^N$ be a set of latent variables which are "soft" labels for each data point. Each $\mathbf{z}_i$ is a K-dimensional vector whose elements z_{ik} represent the degree of membership of point i to class k. Given an estimate of the component parameters, the values of z_{ik} are obtained as the expected values of the class labels in a process called the E step as shown in equation A.2.

$$z_{ik} = \frac{\pi_k \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{A.2}$$

The expected value of the log-likelihood function for the dataset $\mathbf{X}$ and the expected labels $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^N$ can then be constructed as shown in equation A.3.

$$\log P(\mathbf{X}, \mathbf{Z}|\boldsymbol{\Theta}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \big\{ \log \pi_k + \log \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \big\} \tag{A.3}$$

Unlike equation A.1, the expected log likelihood can easily be maximized with respect to the component parameters. This process is called the M-step, which is shown in equations A.4-A.6.

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} z_{ik} = \frac{N_k}{N} \tag{A.4}$$

$$\boldsymbol{\mu}_k = \frac{1}{N}\sum_{i=1}^{N} z_{ik}\mathbf{x}_i \tag{A.5}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{i=1}^{N} z_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \tag{A.6}$$

The EM algorithm proceeds by repeatedly applying the E step and the M step until the results converge. The expected log-likelihood in equation A.3 is a lower bound for the data log-likelihood from equation A.1, and thus maximizing it in the M step is guaranteed not to decrease the data log likelihood.

Appendix B

Markov Random Fields

B.1 Markov Random Fields for Image Segmentation

Markov random fields (MRFs) are a type of graphical model which can represent statistical relationships between unobserved variables. To apply MRFs to image segmentation, the image is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the edge set. The nodes $i \in \mathcal{V}$ are the image locations which need to be classified, i.e. pixels in a pixel-based approach or regions in a region-based approach. Edges $(i, j) \in \mathcal{E}$ represent statistical relationships between nodes. Usually a nearest neighbour graph is used, where edges exist between all pairs of nodes which share a common boundary. This allows the modeling of spatial correlation effects where the label for a given node is correlated with the labels of its neighbours. Examples of nearest neighbour graphs for the pixel-based case and the region-based case are shown in Figure B.1.

Let the label for node i be represented by ℓ_i, where $\ell_i \in \{0, 1, ..., K-1\}$ and K is the number of classes in the segmentation problem. Let $\boldsymbol{\ell}$ denote the set of labels for all the nodes in the graph. The object is to find the label configuration $\boldsymbol{\ell}^*$ which maximizes the probability $P(\boldsymbol{\ell}|\mathbf{x})$ of the labels given the observed data $\mathbf{x}$. This is accomplished by minimizing an energy function $E(\boldsymbol{\ell})$ as shown in function B.1.

$$\boldsymbol{\ell}^* = \arg\min_{\boldsymbol{\ell}} E(\boldsymbol{\ell}) = \arg\min_{\boldsymbol{\ell}} \left\{ \sum_{i \in \mathcal{V}} U_i(\ell_i) + \sum_{(s,t) \in \mathcal{E}} V_{ij}(\ell_i, \ell_j) \right\} \tag{B.1}$$

The terms $U_i(\ell_i)$ are called the unary potentials and $V_{ij}(\ell_i, \ell_j)$ are called pairwise potentials. The unary potentials represent the cost of assigning a particular label to each node, and

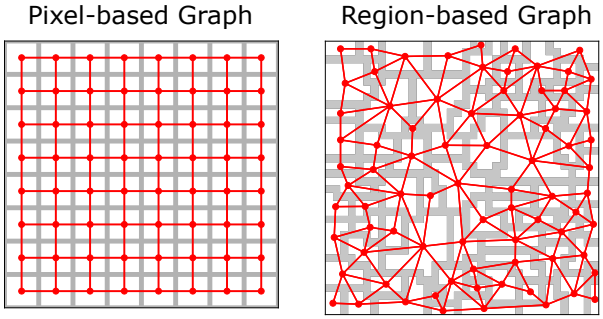

Figure B.1: Examples of pixel-based and region-based nearest neighbour graphs.

they can be derived from the output of essentially any node-level classifier. For example, a common choice is the negative log-likelihood from a mixture model. The pairwise potential $V_{ij}(\ell_i, \ell_j)$ represents the cost of simultaneously labeling node i with label ℓ_i and node j with ℓ_j. The most common pairwise potential in image segmentation is the Potts model, which introduces a discontinuity penalty β when two neighbouring nodes have differing labels as shown in equation B.2.

$$V_{ij}(\ell_i, \ell_j) = \begin{cases} \beta & \ell_i \neq \ell_j \\ 0 & otherwise \end{cases} \tag{B.2}$$

From a Bayesian statistics perspective, the role of the pairwise potential can be interpreted as imposing a prior over label configurations. Minimizing the energy function therefore corresponds to finding maximum a posteriori (MAP) estimates for the node labels.The case where the pairwise potentials are zero corresponds to a uniform prior, resulting in a maximum likelihood estimate where the predictions of the node-level classifier are recovered. From a geometric perspective, the Potts model can be interpreted as a regularization which seeks to minimize the total length of the intra-class boundary in the segmentation result.

When the pairwise potential in a MRF depends on the observed data, it is called a conditional random field (CRF). In image segmentation this is often used to reweight the edge penalty based on local image properties, for example to decrease the edge penalty where the image gradient is large to encourage boundaries to occur along natural edges in the image. Such a weighting scheme for a region-based CRF can be defined following Yu and Clausi [76] as shown in equation B.3. Here $\partial\Omega_i \cap \partial\Omega_j$ represents the set of pixels on the boundary separating region i from region j, ∇_s is the gradient of the image evaluated at pixel s, and $g(\nabla_s)$ is defined in equation B.4.

$$V_{ij}(\ell_i, \ell_j) = \begin{cases} \beta \sum_{s \in \partial\Omega_i \cap \partial\Omega_j} g(\nabla_s) & \ell_i \neq \ell_j \\ 0 & otherwise \end{cases} \tag{B.3}$$

$$g(\nabla_s) = \exp\left(-\left(\frac{\nabla_s}{K}\right)^2\right) \tag{B.4}$$

B.2 Optimization of Markov Random Fields

Finding the minimum energy solution in a MRF is a combinatorial optimization problem which is in general NP-hard, making exact solutions intractable except in special cases. Significant research efforts have been directed towards developing approximate techniques for MRF optimization, and several effective techniques are available such as graph cuts [8], simulated annealing [21], and message passing. In this work I use a variant of the belief propagation (BP) algorithm [52] which is a general message-passing based optimizer. My implementation is inspired by the formulation by Bixler [7] who expressed the belief propagation algorithm using sparse matrix multiplication. This algorithm is very efficient for MRFs on the sparse, irregular nearest neighbour graphs that arise when using a region-based approach tom image segmentation.

The matrix formulation for belief propagation requires encoding the graph structure into a set of sparse binary matrices. The undirected graph is first converted into a directed graph by replacing each undirected edge with two directed edges, one in each direction. Arbitrary but fixed orderings are chosen for the node set and the directed edge set. Using these orderings, let $\mathbf{T}$ be the forward incidence matrix, which maps each directed edge to its destination node. It is a sparse binary matrix of size $2|\mathcal{E}| \times |\mathcal{V}|$. Similarly, let $\mathbf{F}$ be the backward incidence matrix, which maps each directed edge to its source node and is also of size $2|\mathcal{E}| \times |\mathcal{V}|$. Finally, let $\mathbf{R}$ be the direction reversal matrix which maps each directed edge to the corresponding edge in the opposite direction, which is of size $2|\mathcal{E}| \times 2|\mathcal{E}|$.

The unary and pairwise potentials must also be cast into matrix form. Let $\mathbf{U}$ be the matrix of pairwise potentials, a dense matrix of size $|\mathcal{V}| \times K$ as shown in equation B.5 where K is the number of classes.

$$\mathbf{U}_{ij} = U_i(j) \tag{B.5}$$

As above, the rows of $\mathbf{U}$ must follow the fixed node ordering. Similarly, the pairwise potentials are placed in a 3 dimensional array $\mathbf{V}$ of size $2|\mathcal{E}| \times K \times K$ as shown in equation B.6 where i is an index over the ordered set of directed edges, s_i is the source node for edge i, and t_i is the destination node for edge i.

$$\mathbf{V}_{ijk} = V_{s_i,t_i}(j,k) \tag{B.6}$$

Belief propagation employs variables called "beliefs" for each node which represent the likelihood that it is assigned a particular label. The algorithm is an iterative message

passing procedure in which information is propagated between adjacent nodes and used to update the beliefs for each one. To maintain consistency with the convention in computer vision where MRF optimization is cast as an energy minimization problem and the potentials are expressed as costs, I employ a belief propagation variant which operates in the negative log domain. A dense belief matrix $\mathbf{B}$ of size $|\mathcal{V}| \times K$ is introduced where $\mathbf{B}_{ij}$ is the negative logarithm of the belief that node i should take on label j. A message matrix $\mathbf{M}$ is also introduced which carries the information passed between the nodes at each iteration; $\mathbf{M}$ is a dense matrix of size $2|\mathcal{E}| \times K$ where $\mathbf{M}_{ij}$ is the j^{th} element of the message passed between nodes s_i and t_i.

Both $\mathbf{M}$ and $\mathbf{B}$ require initialization before the belief propagation iterations can begin; the algorithm is insensitive to initialization and thus matrices with all zeros are a reasonable choice. The iteration steps are then carried out as shown in equations B.7-B.9. The beliefs are only determined up to a constant, so equation B.8 does not change the result of the algorithm but is useful for avoiding numerical instability. Note that a "broadcasting" operation is required to expand the dimensions of the matrices in equations B.8 and B.9, as described in [7].

$$\tilde{\mathbf{B}} = \mathbf{U} + \mathbf{T}^T\mathbf{M} \tag{B.7}$$

$$\mathbf{B} = \tilde{\mathbf{B}} - \min \tilde{\mathbf{B}} \tag{B.8}$$

$$\mathbf{M} = \min_j \left\{ \mathbf{V} + \mathbf{F}^T\mathbf{B} - \mathbf{R}^T\mathbf{M} \right\}_{ijk} \tag{B.9}$$

The belief propagation has both hard ("min-sum", as shown above) and soft ("sum-product") variants [46]. These variants are essentially analogous to the hard and soft EM variants described in section 3.3.2. The soft variant of BP is obtained by replacing the minimum function in equations B.8 and B.9 with the smooth minimum function shown in equation B.10[1].

$$\text{smoothmin}(\mathbf{x}) = \sum_i x_i - \log \sum_i \exp(x_i) \tag{B.10}$$

[1]The terminology surrounding the various functions related to equation B.10 is unfortunately a bit of a mess. The softmax function should really be called soft-argmax, and the log-sum-exp function is a much better candidate for the name of softmax because it actually provides a soft approximation to the maximum function! The term "softmin" is not available for similar reasons, so I went with "smoothmin" in equation B.10 even though this term is not regularly used.

Once the belief propagation iterations are complete, the final labels are obtained as shown in equation B.11.

$$\ell_i = \arg\min_j \mathbf{B}_{ij} \tag{B.11}$$

Appendix C

Tables of Results

Table C.1: Comparison of polynomial models for the incidence angle effect. Results are presented for the 25 scenes containing both ice and water to enable meaningful conparisons in the unsupervised setting. The best result for each scene is indicated in bold.

Trend Order	1	2	3	4	5
Scene	Pixel Accuracy (%)				
20100510	98.0	98.0	**98.1**	**98.1**	98.0
20100524	**96.8**	95.6	96.0	96.2	96.0
20100605	99.2	98.9	99.2	99.2	**99.3**
20100623	96.4	92.9	97.0	97.3	**97.4**
20100629	95.8	**95.9**	95.6	95.2	95.2
20100704	**97.5**	96.8	94.3	96.5	96.4
20100712	88.4	90.5	90.6	**90.7**	**90.7**
20100721	95.0	95.6	95.7	**95.8**	**95.8**
20100730	90.7	**91.5**	91.3	91.3	91.3
20100807	**89.6**	89.3	89.4	89.4	89.4
20100909	87.3	89.1	89.3	89.3	**89.8**
20101003	**92.0**	91.9	88.8	91.9	91.8
20101017	**73.7**	72.6	72.4	73.4	72.6
20101021	**87.7**	80.7	80.6	81.1	83.2
20101025	**90.3**	87.6	84.9	85.2	85.3
20101027	94.7	94.7	**95.1**	**95.1**	**95.1**
20101114	**94.8**	94.4	94.6	94.1	80.6
20110530	98.5	98.5	98.5	98.5	**98.6**
20110613	98.9	99.0	99.0	**99.2**	99.1
20110627	94.9	**95.5**	95.4	95.1	95.4
20110709	94.8	**95.0**	94.9	94.7	94.7
20110710	94.3	94.4	94.4	**94.6**	**94.6**
20110720	93.7	94.0	**94.1**	**94.1**	93.8
20110725	89.9	89.9	**94.4**	94.3	94.3
20111029	86.6	86.8	87.0	87.1	**87.3**

Table C.2: Pixel accuracy results for the CNN model and the hybrid CNN mixture regression model.

scene	CNN	Hybrid
20100418	99.3	99.9
20100426	99.5	99.9
20100510	99.0	99.1
20100524	98.6	98.7
20100605	99.1	99.2
20100623	98.0	98.8
20100629	93.6	96.6
20100704	95.8	99.3
20100712	84.8	92.5
20100721	94.0	96.5
20100730	92.4	94.7
20100807	93.5	94.2
20100816	93.9	95.0
20100907	99.9	99.7
20100909	95.0	95.2
20101003	94.9	97.0
20101014	99.7	99.9
20101017	94.8	95.8

scene	CNN	Hybrid
20101021	96.1	96.6
20101025	94.8	95.8
20101027	95.8	96.9
20101114	96.5	96.4
20101120	98.6	99.2
20101206	98.1	99.3
20101214	98.8	100.0
20110530	99.3	99.5
20110613	99.4	99.7
20110627	87.9	94.9
20110709	96.5	97.4
20110710	96.3	97.0
20110720	96.4	96.9
20110725	93.5	95.7
20111006	99.9	99.5
20111013	99.3	99.6
20111029	95.0	95.9

Appendix D

Ice Gallery

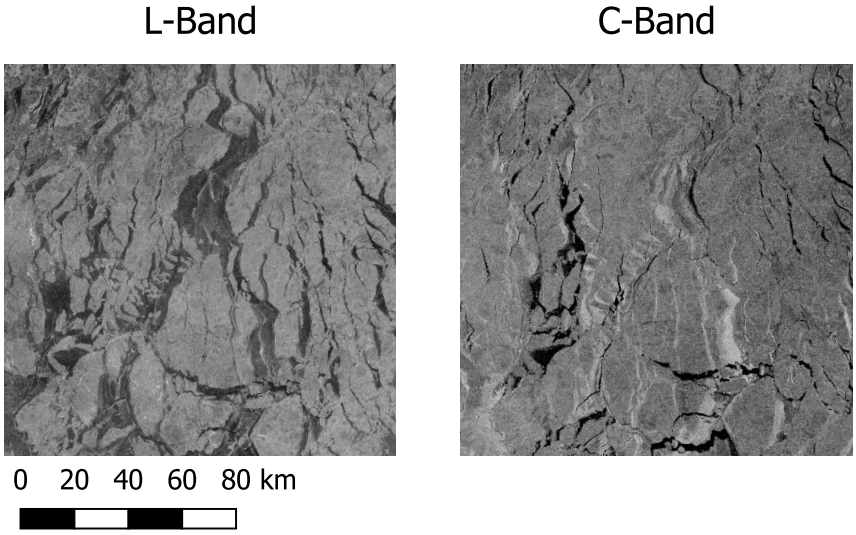

Figure D.1: Comparison of thin ice features in L-band (ALOS PALSAR) and C-band (RADARSAT-2) during freeze-up. Both images are taken with HH polarization. The presence of thin ice is ambiguous in C-band, with some regions appearing dark and others bright, possibly due to frost flowers or other anomalous surface scattering conditions. The deeper penetration depth in L-band resolves this ambiguity.

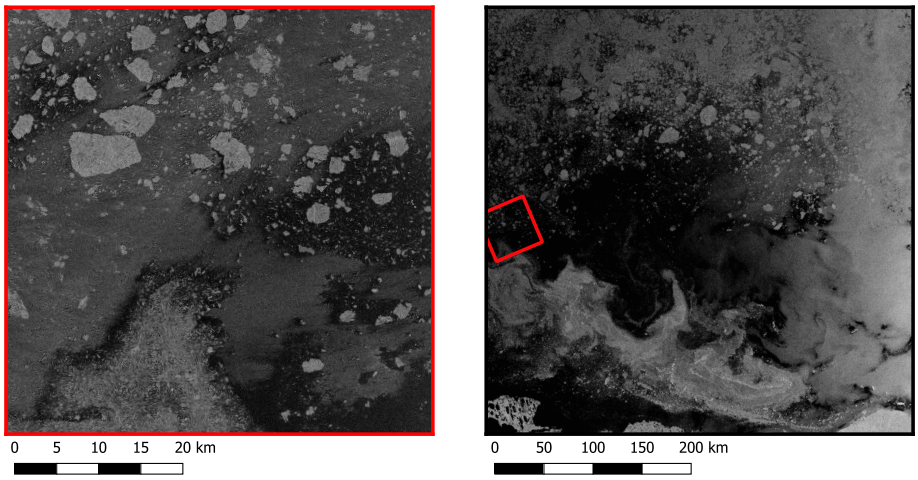

Figure D.2: Example of small ice floes too small to be resolved in wide-swath SAR modes. The left-hand image was acquired in the L-band by the ALOS PALSAR sensor in the fine-beam single polarization mode (HH polarization), with a nominal resolution of 6.25x6.25m. The right hand image is from RADARSAT-2 (C-band, HH polarization) in the ScanSAR Wide beam mode with a nominal resolution of 50x50m. The RADARSAT-2 image was acquired 8.5 hours after the PALSAR image.