

ICE CORE FEATURE DETECTION WITH HYPERSPECTRAL IMAGING

By

Roisin Rumsey

B.A. University of Maine, 2023

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Quaternary and Climate Studies)

The Graduate School

The University of Maine

May 2026

Advisory Committee:

Andrei V. Kurbatov, Associate Professor of Earth and Climate Sciences, Advisor

Mark Royer, Assistant Professor of Computer Science, UMPI

Daniel Breton, Research Geophysicist, Cold Regions Research and Engineering Laboratory

UNIVERSITY OF MAINE GRADUATE SCHOOL LAND ACKNOWLEDGMENT

The University of Maine recognizes that it is located on Marsh Island in the homeland of Penobscot people, where issues of water and territorial rights, and encroachment upon sacred sites, are ongoing. Penobscot homeland is connected to the other Wabanaki Tribal Nations— the Passamaquoddy, Maliseet, and Micmac— through kinship, alliances, and diplomacy. The University also recognizes that the Penobscot Nation and the other Wabanaki Tribal Nations are distinct, sovereign, legal and political entities with their own powers of self-governance and self determination.

ICE CORE FEATURE DETECTION WITH HYPERSPECTRAL IMAGING

By Roisin Rumsey

Thesis Advisor: Andrei Kurbatov

An Abstract of the Thesis Presented
In Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Quaternary and Climate Studies)
May 2026

Visible features preserved in ice cores provide a wealth of information to climate scientists and can be analyzed using a variety of methods and techniques. Hyperspectral imaging presents a novel, minimally destructive approach to maximize information extraction by sampling across the visible light spectrum and into the near-infrared, enabling a richer and more systematic characterization of ice core properties.

This thesis advances the analysis of physical ice core features through two interconnected investigations.

Chapter 1 focuses on air bubbles and stratigraphic layer flagging within a shallow 35-meter section of Antarctic blue ice core ALHIC2302. A pipeline for the detection, mapping, analysis, and removal of air bubbles is developed along with a means of layer flagging through calculation of the Dust Loading Index.

While the resulting algorithms require further calibration and validation against established methodologies, they demonstrate meaningful progress toward automated feature extraction.

Chapter 2 presents a proof-of-concept machine learning framework for the automatic classification of visible ice core features — an approach that could substantially reduce the manual labor and processing time currently associated with ice core analysis. Although the proof of concept is promising, classification performance is constrained by the limited availability of labeled training data, pointing to a clear and addressable bottleneck for future work.

Together, these chapters lay groundwork for a more automated, data-rich approach to ice core analysis, with implications for paleoclimate reconstruction and the broader application of hyperspectral imaging in glaciological research.

DEDICATION

To my parents, whose support, sacrifices, and wisdom made this possible. Their guidance has shaped not only this work, but the person I have become. I am endlessly grateful.

And to my sisters, who stood by me every step of the way.

ACKNOWLEDGEMENTS

Thank you to my advisor, committee, and mentors for their guidance throughout. Thank you to my family and friends for their unwavering support.

I'd like to thank Sara Akiba for her countless hours of work creating annotated datasets.

I'd like to thank Rebecca Addressi and Betty Lee for all their administrative help.

I would like to acknowledge my collaborators: John Fegyveresi, TJ Fudge, Geoff Hargreaves, Valens Hishamunda, John Higgins, and Mikhail Zhizhin.

This work was supported by the next generation U.S. National Science Foundation-Ice Core Facility (NSF-2149519) and U.S. National Science Foundation Center for Oldest Ice Exploration (NSF COLDEX), an NSF Science and Technology Center (NSF 2019719). We thank the U.S. National Science Foundation-Ice Core Facility (NSF-2041950) for ice core sampling assistance and curation. We thank the NSF Office of Polar Programs, the NSF Office of Integrative Activities, and Oregon State University for financial and infrastructure support, and the NSF Antarctic Infrastructure and Logistics Program, the NSF Ice Drilling Program, and the Antarctic Support Contractor for logistical support. We thank the NSF Ice Drilling Program for support activities through NSF Cooperative Agreement 1836328.

I'd like to thank Curt Labombard, Richard Nuun, Theo Carr, and Megan Erskine for ice core sampling assistance at the US-NSF-ICF.

TABLE OF CONTENTS

DEDICATION.....	iii
ACKNOWLEDGMENTS	iv
LIST OF TABLES.....	vii
LIST OF FIGURES	viii
INTRODUCTION	1
1. ADAPTATION OF GAUSSIAN BACKGROUND SUBTRACTION ALGORITHMS FOR PHYSICAL PROPERTY DETECTION AND MAPPING IN HIGH RESOLUTION ICE CORE IMAGES	3
Background.....	3
Methods	5
Ice core data	8
Data processing	8
Results and Discussion	12
Layer flagging	17
Conclusion.....	26
2. TESTING FEASIBILITY OF USING A UNET SEGMENTATION MODEL FOR VISIBLE FEATURE CLASSIFICATION IN HIGH RESOLUTION ICE CORES	27
Background.....	27
Methods	28
Results and Discussion	31
Conclusion.....	36
3. REFERENCES	37
4. APPENDICES	41

Appendix A. Supplementary Materials	41
Appendix B. Modulation Transfer Function	44
Appendix C. ALHIC1901	46
Appendix D. Denali.....	48
Appendix E. ALHIC2416.....	50
5. BIOGRAPHY OF THE AUTHOR	51

LIST OF TABLES

Table 2.1.	Loss and mIoU sample	31
------------	----------------------------	----

LIST OF FIGURES

Figure 1.1.	Standards	6
Figure 1.2.	Bubble Mask	10
Figure 1.3.	Bubble Detection Zoom	11
Figure 1.4.	Bubble Properties Histogram.....	12
Figure 1.5.	ALHIC2302 Bubble Properties by Depth	13
Figure 1.6.	Light Settings Bubble Detection Comparison.....	16
Figure 1.7.	Surface vs Subsurface Bubble Mapping.....	17
Figure 1.8.	Bubble Removal Visualization	18
Figure 1.9.	ALHIC2302 Pixel Intensity by Depth	19
Figure 1.10.	ALHIC2302-42 Grayscale Pixel Intensity	20
Figure 1.11.	ALHIC2302-29 Dynamic Time Warping	21
Figure 1.12.	ALHIC2302-27 Dust Loading Channels.....	22
Figure 1.13.	ALHIC2302 Dust Loading NDIs	22
Figure 1.14.	ALHIC2302-29 High vs Low Intensity Signals	23
Figure 1.15.	Normalized Spectral Radiance	24
Figure 1.16.	ALHIC2302-29 Dust Loading Index Stats.....	25
Figure 1.17.	ALHIC2302-29 Dust Loading Index Visual	26
Figure 2.1.	Visual Features Classifications	28
Figure 2.2.	Annotated Ice Core Image.....	29
Figure 2.3.	Initial Prediction and Ground Truth Masks.....	31
Figure 2.4.	Jetstream Training and Validation Loss	32
Figure 2.5.	Jetstream mean Intersection over Union	32
Figure 2.6.	Derecho Training and Validation Loss.....	34
Figure 2.7.	Derecho mean Intersection over Union	34
Figure 2.8.	Derecho Prediction and Ground Truth Masks.....	35
Figure 2.9.	Comparing Validation Core Results	35
Figure A.1.	Bubble Detection and Analysis Flowchart.....	41
Figure A.2.	Bubble Remove and Layer Flagging Flowchart.....	41
Figure B.1.	MTF Results by Wavelength.....	44
Figure C.1.	ALHIC1901-228_4 Bubble Properties by Depth	45
Figure C.2.	ALHIC1901-228_4 Bubble Properties by Index.....	46
Figure D.1.	DIC1 Dust Loading NDIs.....	48
Figure D.2.	DIC1 Dust Loading Index Stats	48

INTRODUCTION

This study of ice cores offers a novel method to measure physical properties of ice, providing an opportunity to capture these properties with minimum destruction to the samples. By studying ice cores researchers gain a better understanding of climate drivers. However, to properly understand this complex system, it is important to identify all the variables present within the ice and be able to align them across various studies. For example, bubbles can reveal information about temperature and snow accumulation (Lipenkov, 2000), dust can indicate seasonal changes and finding the geological source of the dust can reveal past wind patterns (Nagatsuka et al. 2021), visual stratigraphy offers a means of annual layer counting that can be crucial in understanding the core's chronology.

Hyperspectral imaging (HSI), or hyperspectral imaging spectroscopy, is a tool designed to characterize different materials based on their optical properties. This is possible due to the difference in how materials reflect and absorb light at different wavelengths which allows spectral signatures to be recorded with hyperspectral sensors. A typical modern day instrument captures hundreds of bands across the visible light electromagnetic spectrum. The term 'hyperspectral imaging' was first introduced in 1985 by Goetz in an article about images to be used for the identification of surface materials by looking at the reflectance values of each pixel (Goetz, 1985). Since then, HSI has been successfully used for remote sensing of the Earth using space and airborne platforms as well as extraterrestrial bodies. Most recently HSI has been introduced to other industries as well as propagated lab settings. In this thesis I tested several HSI applications in ice core research.

Chapter 1 has a focus on feature detection within ice cores. It is mainly focused on manual algorithms adapted to segment (Yu et al. 2023) trapped air bubbles. Once bubbles are

detected and their boundaries are identified, geometrical properties of the bubbles (size, shape, orientation) can be automatically analyzed. The bubbles can be digitally masked (removed) which allows for the future study of spectral signatures (fingerprinting) of ice cores. Chapter 2 is a proof of concept that uses supervised machine learning algorithms with manually labeled ice core datasets to automate the feature segmentation in ice core samples.

This thesis uses ice core samples collected from the Arctic and Antarctica and demonstrates how hyperspectral imaging could be used for the development of continuous ice core based time series of physical properties, with minimum destruction of the material.

CHAPTER 1

ADAPTATION OF GAUSSIAN BACKGROUND SUBTRACTION ALGORITHMS FOR PHYSICAL PROPERTY DETECTION AND MAPPING IN HIGH RESOLUTION ICE CORE IMAGES

Background

Hyperspectral imaging, or HSI, has been developed to characterize different materials based on their unique optical properties. HSI has been used in a variety of fields and industries (Bhargava, 2024; Mukhtar, 2025), here I discuss how this technology is utilized in several ice core specific research problems.

Ice core paleoarchives are well known for their robust annual layers derived from visual stratigraphic feature counting. Counting layers is one of the oldest established methods of developing ice core time scales (see summary in Alley et al., 1997). In ice cores collected from carefully selected locations (Andersen et al., 2006), these layers are well preserved with clear and cloudy alternating bands representing seasonal changes, making it possible to visually identify and count these layers. In ice core sites with low accumulation rates (e.g., East Antarctica), annual layers are thinner, therefore, the ability to detect these layers is reduced as the ice core gets older (Shackleton et al., 2025) and more compressed with depth (Winstrup, 2011).

Visual properties of ice cores have been studied to discern these annual layers and stratigraphy (e.g., Benson, 1962; Gow, 1965; Langway, 2008; Winstrup, 2011). Optical measurements were introduced to quantify dust deposits, snow firnification rates, and stratigraphic unconformities (Svensson, 2005; McGwire, 2008). Takata et al. (2004), by studying the intensity of light scattering in polar ice determining that intensity variations offer a means of measuring annual layer thickness. These light scattering intensity changes can be attributed to the

variations of ice grain size (Stoll et al., 2025), bubble number density (Askebjør et al., 1997), ash content (Bramall et al., 2005), and other impurities in the ice (Alley et al., 1986; Stolz and Ram, 2005). The variations in light intensity allow identification of annual layer thickness and annual accumulation rates. Svensson et al. (2005) introduced a high-resolution line scan camera on the NorthGRIP ice core samples to delineate annual layering. The technique captures a single line scan across the ice core's surface as either the core itself or the camera system moves incrementally along its length. The lines are then combined to create a 2D image of the ice core. By recording differences in light scattering and absorption, line scan imaging effectively records physical variations within the ice.

Another significant feature that is often visible in ice cores at certain depths is trapped atmospheric air bubbles (Bendel, 2013). Air bubbles appear in shallower ice, but turn into clathrates once the increasing pressure and decreasing temperature destabilize the bubbles (Neff, 2014). The concentration, spatial distribution, size, and orientation of these bubbles have been studied to understand snow to ice transition or the physical settings during the snow accumulation (Gow, 1968). Lipenkov's (2000) study of bubbles in the Vostok Ice Core revealed that the quantity and size of bubbles served as an indication of temperature and snow accumulation. The sample preparation and analysis of thin bubble sections can be potentially subjective and labor intensive (Fegyveresi, 2011) since manual inspection is required. Even though a new method has very recently been introduced, using Micro-CT, the process of acquiring bubble data to analyze still takes multiple hours. By utilizing line scan imaging technology, Morcillo et al. (2020) were able to analyze bubbles in the gray-scale images, showing that the bright pixels "could be used as a proxy for bubble concentration".

By adapting hyperspectral line scan imaging methods for bubble detection our research aims to expand existing bubble identification and characterization methods, while capturing additional spectral data that can be used for impurity fingerprinting in ice core samples. The hyperspectral imaging system (IceSpec) used in this paper is hosted at the NSF-Ice Core Facility (NSF-ICF), Denver CO (Kurbatov et al., 2025). IceSpec captures 224 bands from across the visible spectrum and into the Near-Infrared (NIR) range, 0.4-1 μ m. Scanning an ice core section that is approximately one meter results in a 2 gigabyte, multi-dimensional hyperspectral data cube (Garzonio et al., 2018), containing information that can effectively act as a material 'fingerprint'. Because bubbles or ice fractures scatter light these features have high intensity signals, effectively dominating other low intensity parts of the image. Therefore, before the spectral data can be used for ice core material characterization, the bubbles, fractures or other dominant features have to be digitally masked.

Methods

Ice core sections are cut into slabs and planed on both sides to ensure there is a flat and even surface for optimal imaging. The need for an ultra flat (microtomed in past studies) surface is negated, however, by focusing the camera 5mm below the sample surface. Samples are imaged using the FX10e Specim camera. Currently the IceSpec system can accommodate sample slabs that are between 2-3.8cm thick. The sample is placed on a tray that moves incrementally beneath the camera. This linescan system has a variable actual resolution across the different spectral bands; calculated using the Modulation Transfer Function (Koren, 2007) methodology adapted by Rumsey et al. (2025) for ice core studies.

The camera acquires the hyperspectral data in ENVI format and metadata is recorded for each image, to capture information like light sources, intensity, exposure, sampling frame rate,

ice core sample ID, core section depths, and ice quality. The information is stored in a tabulated CSV text file. See processing section for more details on how it is used for automatic processing.

IceSpec is equipped with four interchangeable LED or halogen light sources (Kurbatov et al., 2025), two positioned above and two below the sample, each angled at 45 degrees, towards a singular focus area about 5mm below the surface of the sample. This light and camera geometry creates a dark field illumination setting introduced to ice core studies by Svensson (2005). The setup improves contrast of light scattering impurities present in ice.

Every ice sample acquisition is accompanied by the measurement of reflective spectral standards. Three Spectralon panels: 20, 50 and 100% reflectance, are placed at the beginning of each sample scan. These standards are used for radiometric calculation, normalizing digital numbers across days and lighting conditions (Specim, 2025). We use the 20% reflectance panel as our white reference to avoid signal oversaturation from bright ice and to increase the signal to noise ratio of impurities. The spectral standards have materials with known spectral signatures (Figure 1.1). For the dark reference, the camera shutter is closed and about 100 frames are recorded automatically, to

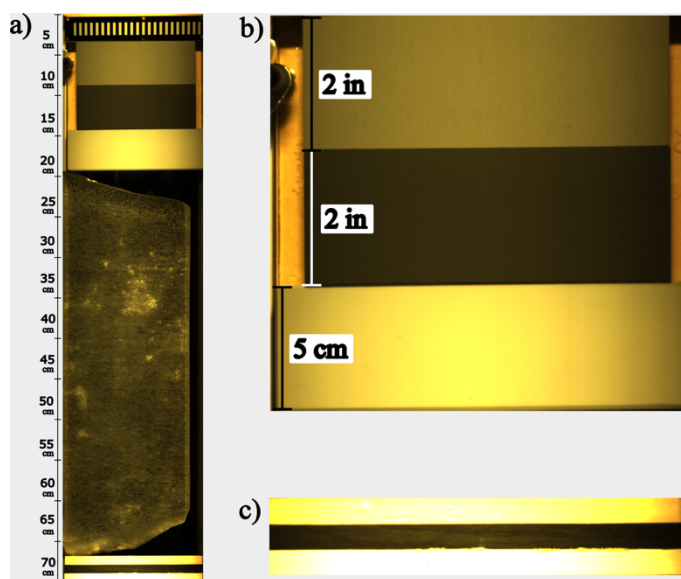


Figure 1.1 Standards used for hyperspectral line scans. Panel A) Example of a fully captured image of a 46cm section of ice bracketed by the lighting standards at the top, and the spectral standards at the bottom. Panel B) Depicts the 3 spectralon panels from the top of Panel A with 20% (top), 50% (middle) and 100% (bottom) reflectance. Panel C) Depicts the spectral standards from the bottom of Panel A.

measure sensor noise. The white and black references are used for image normalization calculation (Eq. 1).

Three different light settings were used to capture images for each ice core scan: 1) with top lights at full intensity, bottom lights off, 2) and 3) with top lights at half intensity, bottom lights at full intensity. Different exposures and scan speeds were used to maximize the recorded dynamic range for each sample with these light settings. Images captured with a higher exposure result in the oversaturation of bright objects, like bubbles and fractures, but the background ice signal will be more prominent after radiometric calculations. The second image light setting has a higher exposure, to look at the background ice, and the third image has a low exposure time to look at the brighter artifacts.

The first image lighting is reflective, using the conventional hyperspectral imaging technique that aims to record materials optical properties using reflective light settings. Traditionally, this is performed on opaque samples. Because ice is usually transparent for visible light this method has limited applicability but is usable for identification of bubbles or embedded dust layer geometries when used in combination with transmitting light settings.

Measuring ice cores in both transmitted and reflected light modes is beneficial because the two geometries reveal complementary information about impurities, microstructure, and optical properties of the ice. In practice, most modern high-resolution ice-core optical systems use transmitted settings because of the very low level of impurities in polar ice core samples. We found that images captured with a transmissive light setting illuminates more features in the measured ice samples (Figure 1.4).

We compared how the different light settings affected the pixel intensities and found the reflectance lighting had a weaker signal, but the overall trend was the same. This is likely due to

the samples captured in July of 2024 using the same reflective white reference for both light settings. This means that the transmitted light isn't normalized for these samples. For later sample acquisitions, we added a transparent sanded glass that could be used for normalizing the bottom lights.

Captured images were processed with Python computing language (version 3.12.2) scripts that use several open source libraries: Spectral Python (SPy), OpenCV, scikit-image, pandas, and NumPy. For processing we use jupyter notebooks (JupyterLab 4.3.4), running on a Debian 12 Linux OS server with an Intel(R) Xeon(R) Gold 6140 CPU and a total 188GB of memory.

Ice core data

For this research we analyzed 39 sections of Allan Hills Ice Core ALHIC2302 recovered by the COLDEX team in Antarctica in 2023/24 season (-76.74543198 N, 159.375134 E). Only samples between 10 and 45 meters in depth had suitable quality for imaging. We also scanned sections of Denali DEN13A (62.93333 N, -151.08333 E) ice core collected in Alaska, several sections of West Antarctic Ice Sheet (WAIS) Divide Deep Core A WDC06A (-112.1115 N, -79.481 E), and samples from the Greenland Ice Sheet Project 2 Core GISP2D (72.5833333 N, -38.466667 E). Results will focus on ALHIC2302.

Data processing

The workflow (Figure A.1) begins with loading the ice core image files, utilizing associated metadata, and converting the image into a NumPy array from the proprietary ENVI format, using SPy (Spectral Python, 0.23.1) The image is manually cropped, so only the representative sample area is included for processing (Appendix A). The pixel-to-ice core depth

conversion is calculated using the recorded depths for the core section, and the sample's corresponding pixel dimensions in the linescan image.

To normalize the scanned samples between different imaging sessions we utilize the spectralon panels mentioned above to derive reflectance values. These reflectance values are used to adjust each pixel using Equation 1.

$$correctedPx = \frac{(px - darkRef)}{(whiteRef - darkRef)} \text{ (Eq.1)}$$

Where px is the original image pixel intensity digital number, $darkRef$ is the average digital number in the corresponding pixel column from the dark image, and $whiteRef$ is the average digital number in the corresponding pixel column from the white reference.

This correction counteracts uneven lighting conditions between different imaging sessions and allows data reproducibility. Since the data on either end of the captured spectrum is often noisy, we remove the first 19 bands and the last 13 bands for each image, by utilizing the delete function from numpy (version 1.26.4) which returns a new array with the selected elements along a specific axis removed. The signal to noise ratio (SNR) of each band is calculated using Equation 2.

$$SNR = \frac{\sigma_{whiteRef}}{Noise_{data}} \text{ (Eq.2)}$$

$Noise_{data}$ is the noise estimate in the ice core data obtained using the Immerkaer method (a convolution-based filter for Gaussian noise variance), and $\sigma_{whiteRef}$ is the standard deviation of radiances over a uniform white reference panel. This allows for a per-channel estimation of quality and usability. Unlike hyperspectral images captured with drones or other aircraft, the SNR equation here can assume a stable environment as these images were captured in a laboratory setting. Higher SNR values indicate lower noise and higher channel performance.

Spectral channels that fall below the SNR threshold of 3.5 or above 10 are removed to avoid any false detections and over saturations.

Once this preprocessing step is complete, the next step isolates and tags the bubbles that are present in the ice. The bubbles can be identified as high-intensity regions in the image, so to isolate them we apply the Background2D function in the photutils package (Bradley et al. 2024) to estimate and subtract the background of one spectral band of the image. A box size of (5, 5) and an exclusion percentile of 2, and MedianBackground is used. Then we calculate the standard deviation of the corresponding band's Full Width at Half Maximum(fwhm) matrix, found in the image metadata, for a gaussian distribution. The band with the background subtracted is convolved with a gaussian kernel, which was calculated using the *fwhm* standard deviation and the *Gaussian2DKernel* function from astropy using the standard deviation and a size of 5. The kernel was then used in the convolve function with the subtracted band array. This data is then used to create a segmentation mask. This process repeated for each spectral band in the image. The masks are then stacked into a single 2D boolean mask, using OR logic; showing bubbles that were detected in any channel as white regions and the surrounding ice as black regions (Figure 1.2).

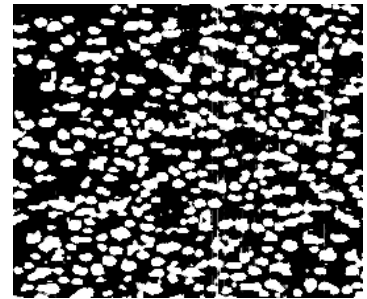


Figure 1.2 An example of a 2D Boolean mask showing bubbles as white regions.

This mask is applied to the preprocessed image, leaving only the pixels in which bubbles were detected. Then we create a radiance based elevation map from the image, by finding regions that have significantly higher intensity than the median. We utilize the *peak_local_max* function from the scikit-image library (Walt, 2014. Version 0.24.0) to do that, with a square footprint of 3 and a threshold of 1. From here, we use the *watershed segmentation* function

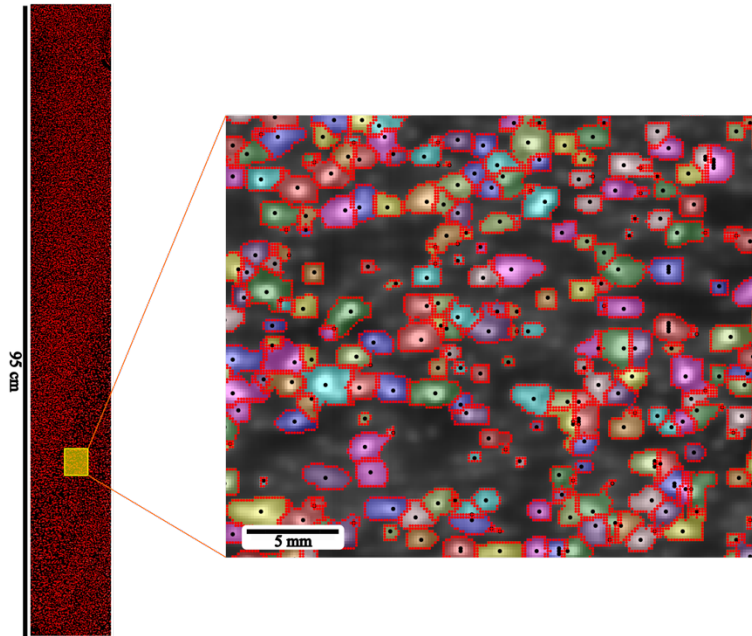


Figure 1.3 A zoomed-in section from ice core sample ALHIC2302-29 with identified bubbles outlined in red after being processed with the bubble detection script.

(`segmentation.watershed`) from the `scikit-image` library, on the radiance based elevation map of the image in order to differentiate any bubbles that are close together; the bounds of each bubble are then mapped (Figure 1.3), using *`find_boundaries`*. As Figure 1.3 shows, not every bubble will be detected. Very small bubbles might be ignored as only bubbles ten pixels ($\sim .5\text{mm}^2$) and larger are

detected in an attempt to avoid small artifacts on the surface of the ice that may become false positives, or they just might be too far below the surface of the ice, and not bright enough to be detected.

Bubble properties are quantified within 5x5 cm subsections in order to get a statistically significant number of bubbles. To do this, the center-most 5 cm in each image is calculated for each depth, and every 5 cm of depth is segmented, creating a 5x5 cm square. In each subsection the number of bubbles with a centroid in that square is counted, and each bubble is logged into a table with its corresponding characteristics (size, elongation, orientation, etc.). The characteristics are acquired with *`skimage regionprops`*. After each square-subsection is analyzed, the mean, median, and standard deviation of each characteristic per subsection are calculated and logged into a table.

Results and Discussion

The bubble detection algorithm was tuned on the 39 sections of ALHIC2302 ice core samples that were scanned with the IceSpec system in July 2024. These sections spanned from 10m to 45m in depth. The histograms in Figure 1.4 show data from the total number of bubbles throughout the ALHIC2302 ice core samples. The histogram data shows that most of the bubbles are small and have an orientation of 0 degrees. It's worth noting that round bubbles are likely to be registered as having a 0 degree orientation. This concern isn't overwhelming, as the last graph in Figure 1.4 shows that while there is a large number of round bubbles, most of the bubbles are elongated.

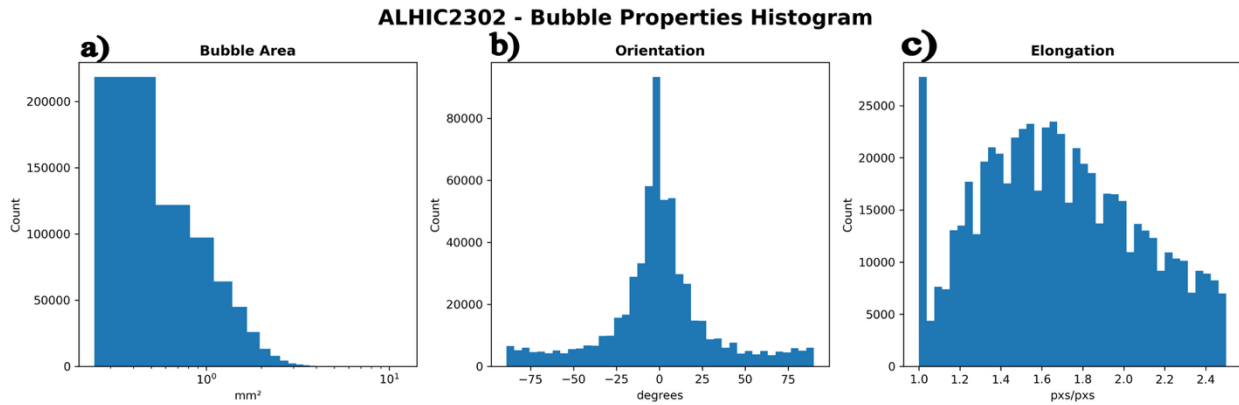


Figure 1.4 Histogram plots for bubble properties throughout the ~10m-45m of ALHIC2302. Panel A) depicts the bubble area, showing that most of the bubbles are quite small. Panel B) depicts the bubbles' orientation from -90 to 90 degrees. Panel C) depicts the bubbles' elongation, showing that while there are a large number of round bubbles (elongation 1), there are also many bubbles that are quite elongated (elongation > 1.5).

Figure 1.5 shows the resulting stats from the 5x5 cm subsections. The number of bubbles, median bubble size, median elongation, and median orientation is shown on plots a, b, c, and d, with the 5th and 95th percentiles shown with the medians. The number of bubbles mostly stays between 15 and 25 bubbles per cm², but has a minimum of 8 and a maximum of 29. Several fluctuations with depth are visible at 16, 19, 23, 27, 33, 38, 42, and 44 meters. We observed a slightly inverse relationship between bubble size and bubble count, most prominent in the interval of 30 to 35m.

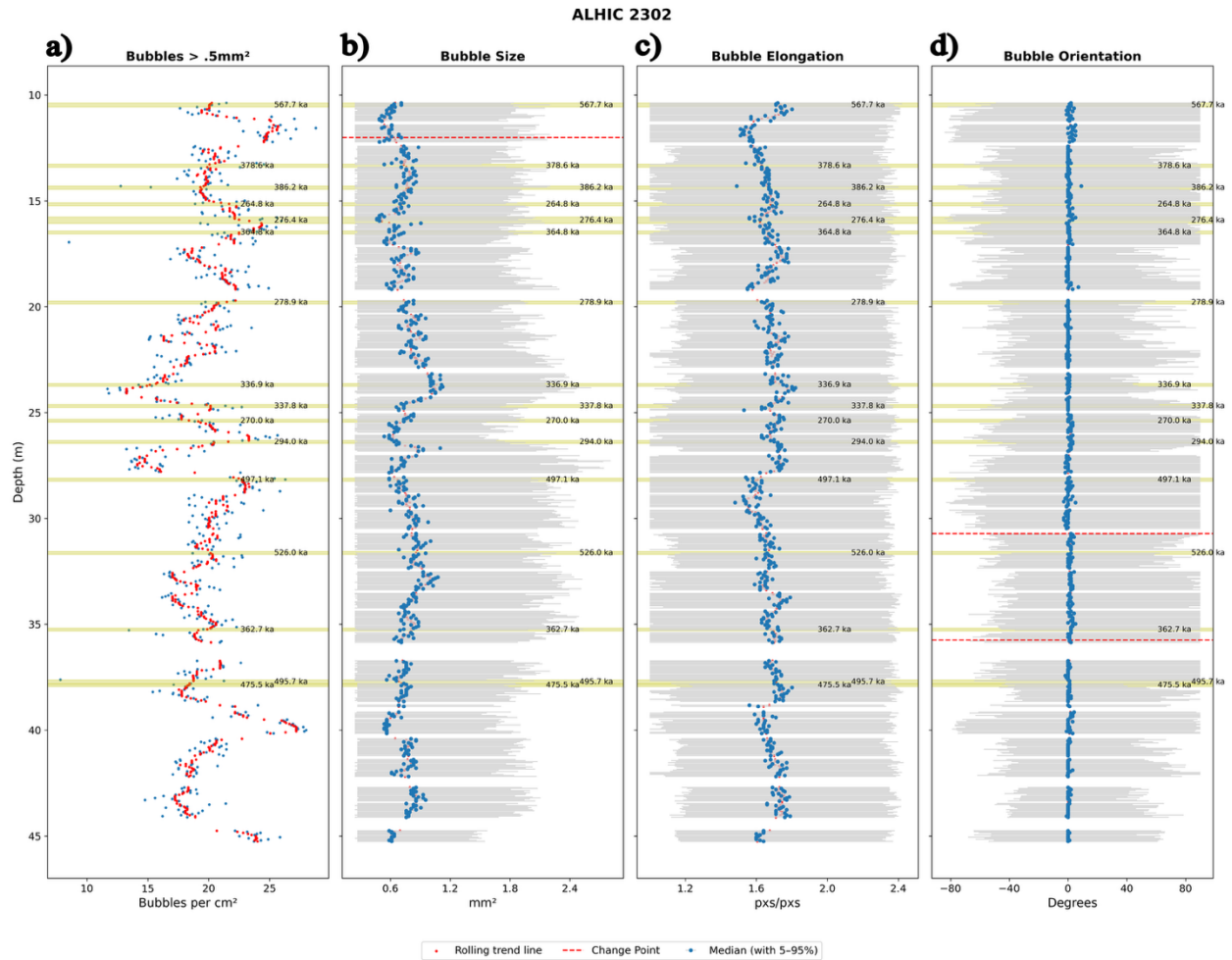


Figure 1.5 The plots for bubble properties in ~10m-45m of ALHIC2302. Each dot graphed represents the summary statistics of a 5x5 cm subsection. Panel A) Depicts the number of bubbles in each 5x5cm subsection in blue and a rolling median trendline with a window of 10 in red. Panel B) Depicts the median bubble size in each 5x5cm subsection in blue with gray bars on either side of each median point extending to the 5th and 95th percentile. Panel C) Depicts the median elongation of the bubbles in each 5x5cm subsection in blue with gray bars on either side of each median point extending to the 5th and 95th percentile. Panel D) Depicts the median orientation of bubbles in each 5x5cm subsection in blue with gray bars on either side of each median point extending to the 5th and 95th percentile. Detected change points are shown with red dashed horizontal lines. Argon dates are shown for regions highlighted in yellow; this preliminary data was shared through personal correspondence by Higgins and Hishamunda from Princeton University.

The mean and standard deviation elongation of the bubbles at each 5x5cm box is also calculated. The standard deviation shows a greater variation of elongated bubbles in the shallower part of the core, with the variation getting smaller at about 20 meters and remains relatively stable in the deeper section of the core. We interpret near the surface variability as a possible impact of surface seasonal temperature changes.

The median bubble orientation shows the angle of the bubble's major axis in degrees, with 0 being fully vertical and 90 being horizontal. The orientation plot in Figure 1.5 shows wide variability from -90 to 90 degrees. We interpreted bubble orientation as an indication of the dip in the ice layers. Since there is a high likelihood that we do not always have a perfect cross section profile of the bubble as it is dipping, the orientation cannot be considered an exact measurement of the layer dipping profile, but the maximum angle of orientation captured for some bubbles is probably close to a layer dip.

The graph in Figure 1.5 also shows captured boundaries established using the Change Points algorithm implemented in the Ruptures Pelt python library (ruptures.Pelt, version 1.1.10) (Truong, 2020). The change points typically occurred at the breaks between different ice core sections.

The changes in bubble orientation may also be an indication of core azimuth changes. As the core is drilled and sections are pulled up, the orientation of the core may be lost in the field. The orientation graph shows spikes in the median around 12, 19, 27, between 30 and 33, 35 and 39m. There's also a change point detected at 24m. We know the core azimuth was lost at sections 22 (bottom depth - 19.57m), 26 (bottom depth - 24.14m), 33(bottom depth - 30.66m), and 42 (bottom depth - 38.84m). These core depths align closely with the orientation changes seen in Figure 1.5.

The method of bubble detection using bright spots as indicators is not 100% accurate but can potentially be calibrated to other methodologies, like thin section studies.

Since the system typically scans ice core slabs that are roughly 3cm thick, it is expected that it captures more bubbles than is typically recorded in the thin section studies used to acquire bubble number density (Fegyveresi, 2016). We observed that the bubbles are evenly distributed throughout the ice when we compared the same ice samples imaged with reflective light and transmitted light. Under the transmitted light, more bubbles are illuminated below the surface compared to the reflected light settings. Figure 1.6 confirms these assumptions.

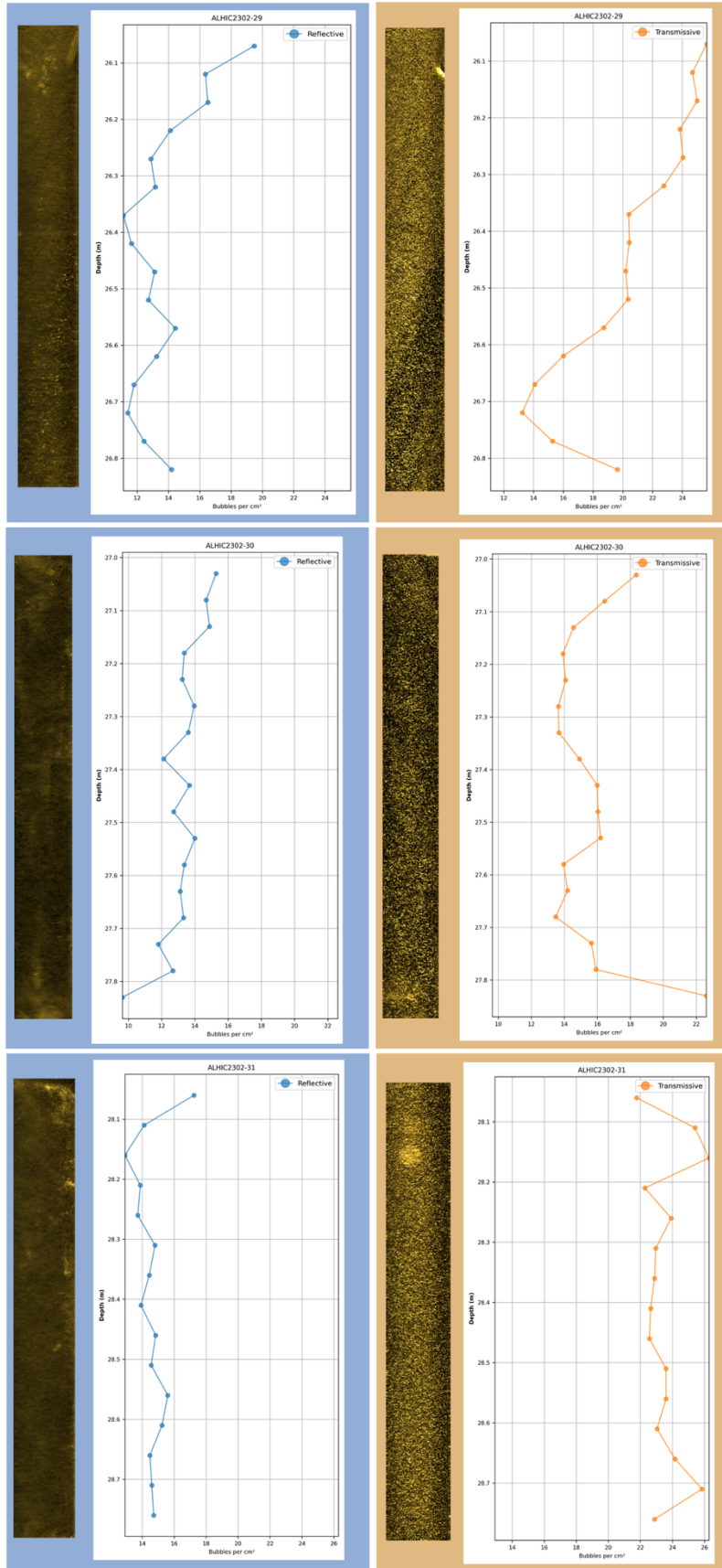


Figure 1.6 Comparison of bubble detection for reflective versus transmissive light settings for ALHIC2302 sections 29 (A), 30 (B), and 31 (C) with bubbles per cm² along the x-axis and depth in meters along the y-axis.

Based on light propagation in ice (Warren, 2019), we compared bubble sizes at different wavelengths of light. Since light scattering travels farther at 400nm than at 1000nm, we hypothesized that a bubble that appears smaller at 1000nm is likely deeper in the ice, while a bubble that appears bigger is likely at the surface of the ice sample, and bubbles that remain roughly the same size are at a mid - depth within the sample. This hypothesis was shown for ALHIC1901 section 228_4-1 as a color map in Figure 1.7. However, this theory still needs to be experimentally tested for confirmation.

Layer Flagging

The large number of bubbles seen in the Allan Hills Blue Ice Area samples complicates stratigraphic layer detection (flagging). We applied the preprocessing steps outlined above to detect and mask bubbles before digitally removing all masked areas. The empty regions that previously contained bubbles are digitally



Figure 1.7 A grayscale image of ALHIC1901_228_4-1 with an overlaid color map of various bubbles showing the depth of the bubble as determined by a change in size across channels (.4 μ m, .6 μ m, and 1 μ m).

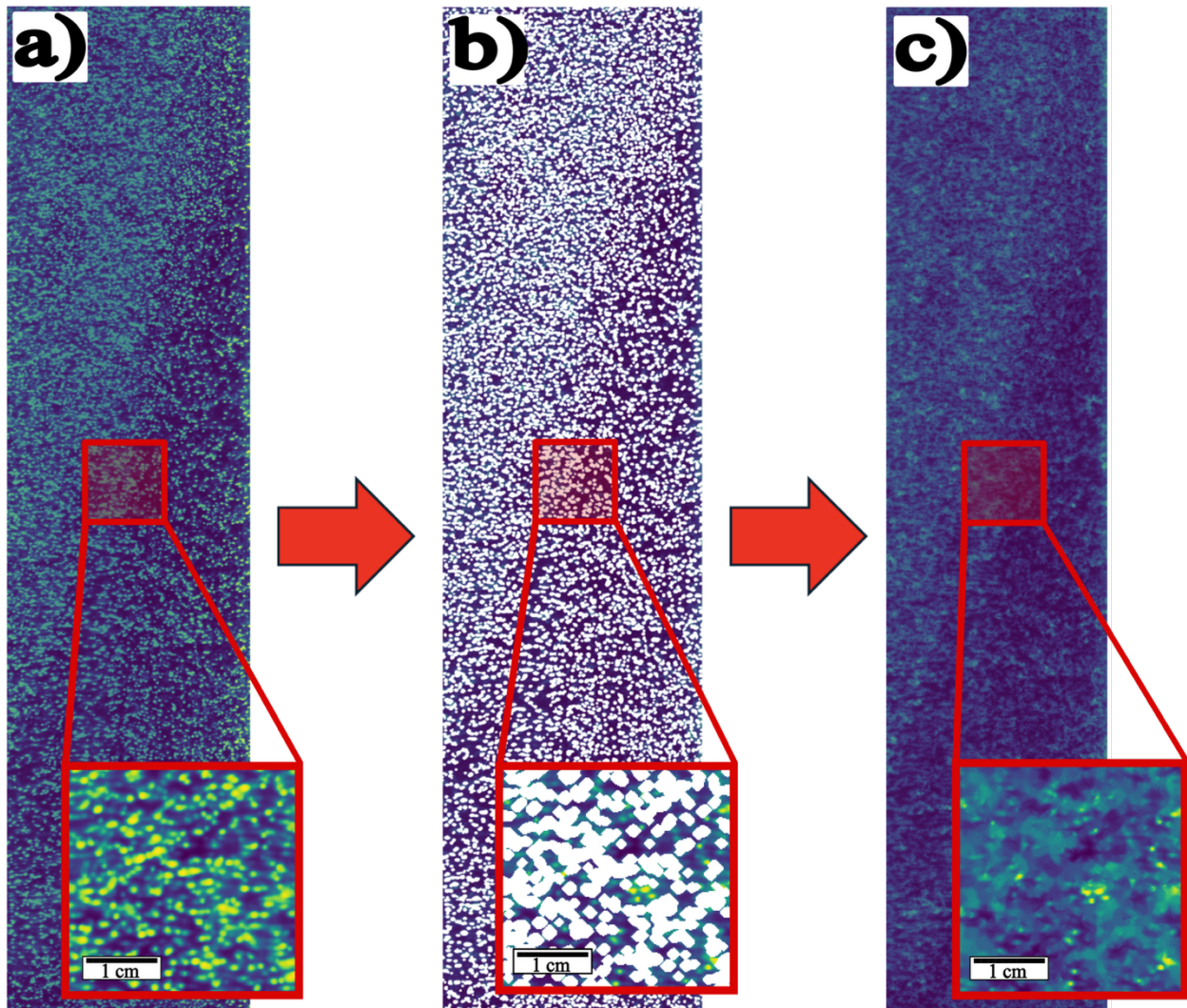


Figure 1.8 Demonstrates the iterative process of bubble detection. Panel A represents the original image. Panel B represents the image post bubble detection algorithm application and removal of detected bubbles. Panel C represents the image post inpainting algorithm application.

smoothed using the surrounding pixels' digital values using an inpainting method (cv2.inpaint, version 4.10.0) (Bradski, 2000). The results of the bubble removal and inpainting are shown in Figure 1.8.

The value for each horizontal pixel row is combined across all channels to calculate the mean, median, and cumulative intensity values. This transformation reduces the original hyperspectral cube to one intensity of values per depth increment creating intensity profiles that can be used to flag potential layers in ice core samples.

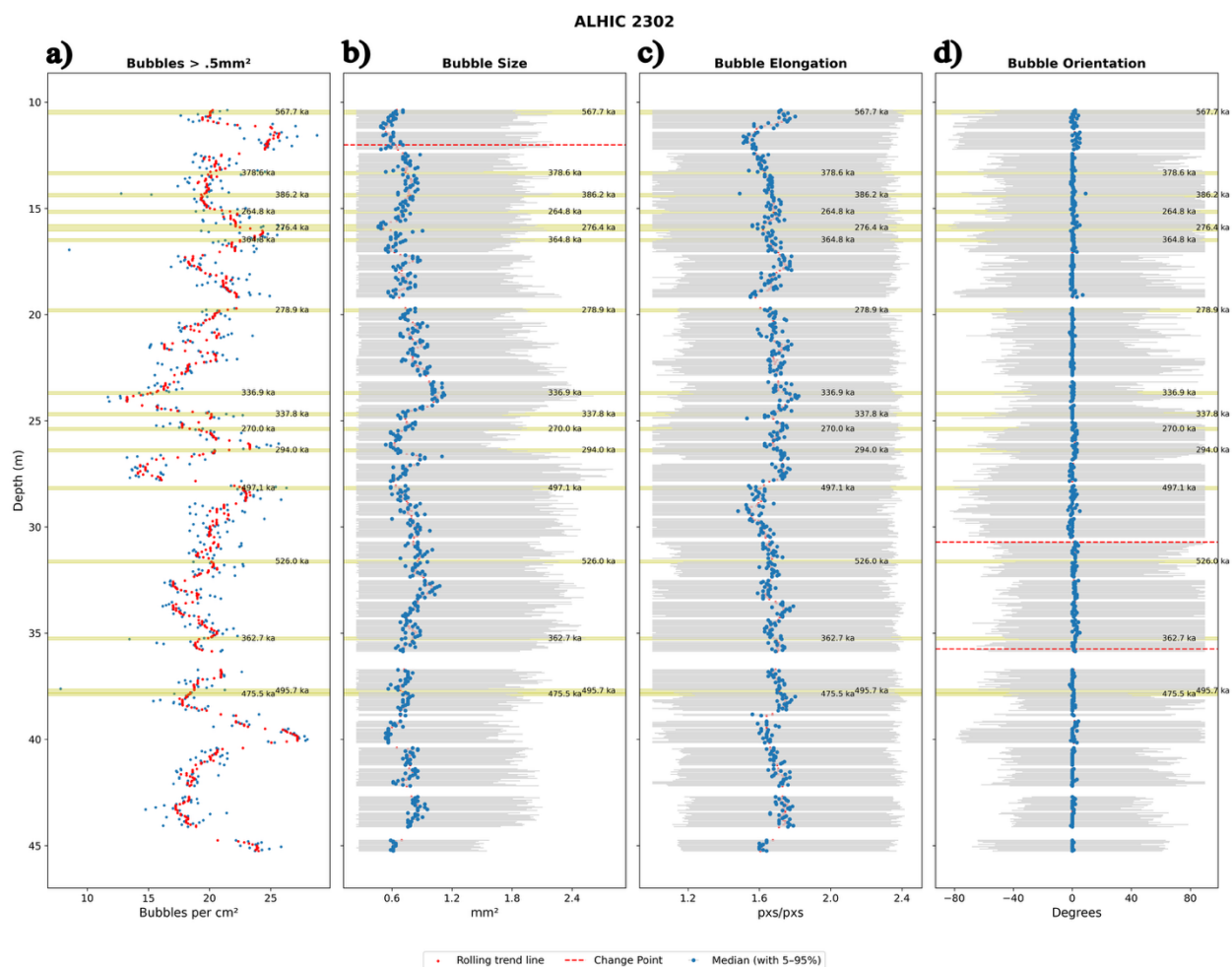


Figure 1.9 The plots for row-wise pixel intensity in ~10m-45m of ALHIC2302. Panel A) Depicts the number of bubbles in each 5x5cm subsection in blue and a rolling median with a window of 10 in red, as seen in Fig 3. Each blue dot graphed in panels B, C, and D represents the summary statistics of a single pixel row of an image after the bubbles were removed and inpainted (see Fig 5). Panel B) Depicts the sum total (Cumulative) digital intensity of the pixel row. Panel C) Depicts the median digital intensity of the pixel row, with the 5th and 95th percentile shown using gray bars. Panel D) Depicts the mean digital intensity of the pixel row, with +/- 1 standard deviation shown using gray bars. This digital intensity is expected to show dust scattering, however this is complicated due to core quality. In an ideal core, this change in intensity can show glacial/interglacial cycles.

Figure 1.9, depicting the cumulative, median, and mean spectral signal for each pixel row, shows relative variation in intensity. In Continuous Flow Analysis (CFA) systems, only annual layers

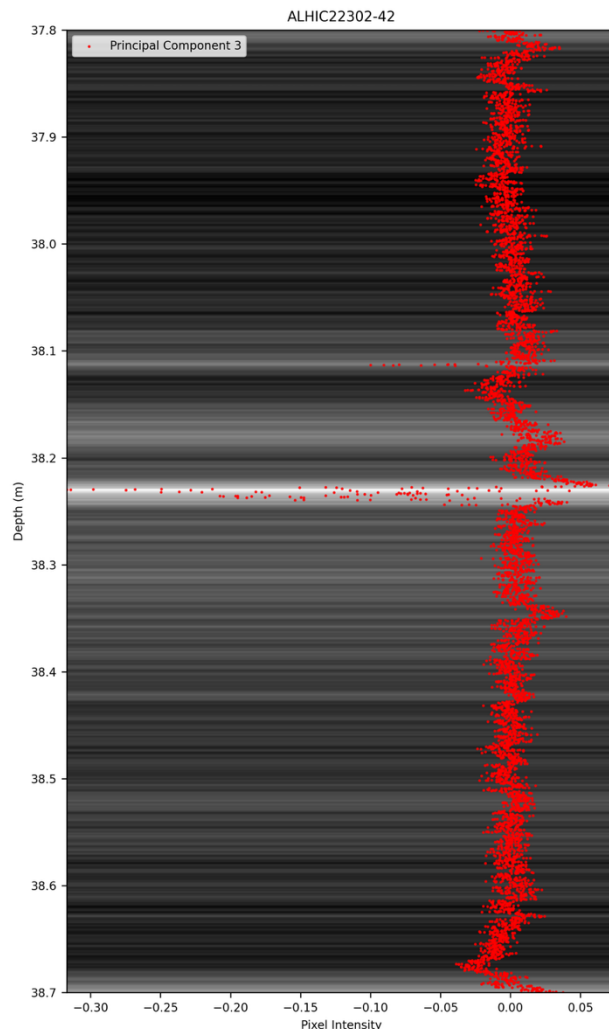


Figure 1.10 3rd principal component plotted for ALHIC2302 section 42 overlaid on a grayscale representation of the section's row-wise cumulative intensity.

that are thicker than 1-2 cm are detectable (Cole-Dai et al., 2006; Osterberg et al., 2006; Grieman et al., 2022). Here we can see the variation per pixel.

In Figure 1.10 we zoom in on ALHIC2302 section 42 (37.8 to ~38.7m). By focusing in on just one meter, we can see the high frequency of fluctuations, and the grayscale profile helps to visualize the prominence and thickness of the pixel intensity changes. These changes can potentially map ice core stratigraphy layering and unconformities (Figure 1.9). This figure also has a principal component of the sample plotted by depth over the grayscale profile;

this plot doesn't increase and decrease with the intensity profile, showing that there are

other variables not being accounted for in this layer flagging method. We also compared how the different light settings affected the pixel intensities and found the reflectance lighting didn't capture as strong of a layer signal, but the overall trend was the same.

Figure 1.11 shows 2 intensity signals, one from the left side of the ice core section ALHIC2302-29, the other from the right side of ALHIC2302-29. By comparing the signals and applying an algorithm called Dynamic Time Warping (Giorgino, 2009), we can determine if the stratigraphy is uniform or slanted. The results show a considerable dip, especially in the second half of the sample, interpreted already to have dipping in the layers (Kirkpatrick et al., 2025).

Since we know light propagation changes in ice at different wavelengths, we also compared the stratigraphic signal at $.42\mu\text{m}$ (violet/blue), $.52\mu\text{m}$ (green), and $.98\mu\text{m}$ (NIR). Light absorption is nearly zero at $.4\mu\text{m}$ and scattering from bubbles and impurities cause blurriness, $.42\mu\text{m}$ was chosen for this study since the signal at $.4\mu\text{m}$ was fairly noisy. The absorption at $.98\mu\text{m}$ is much higher compared to $.4\mu\text{m}$, so light scattered by bubbles or impurities is absorbed, meaning that any light captured is from the camera's focal point, creating higher contrast. All three wavelengths also appeared when capturing just the principal components from the images. The comparison for these wavelengths is shown for ALHIC2302 section 27 in Figure 1.12.

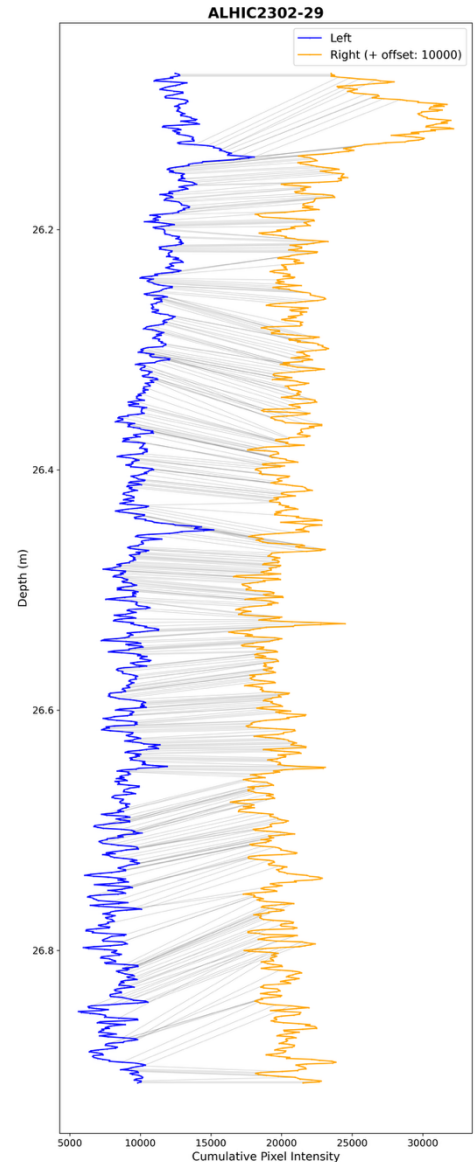


Figure 1.11 Dynamic Time Warping comparing the cumulative intensity of a 50 pixel wide strip from the left of ALHIC2302-29 and the cumulative intensity of a 50 pixel wide strip from the right of ALHIC2302-29. This comparison is expected to show potential layer dipping in the core.

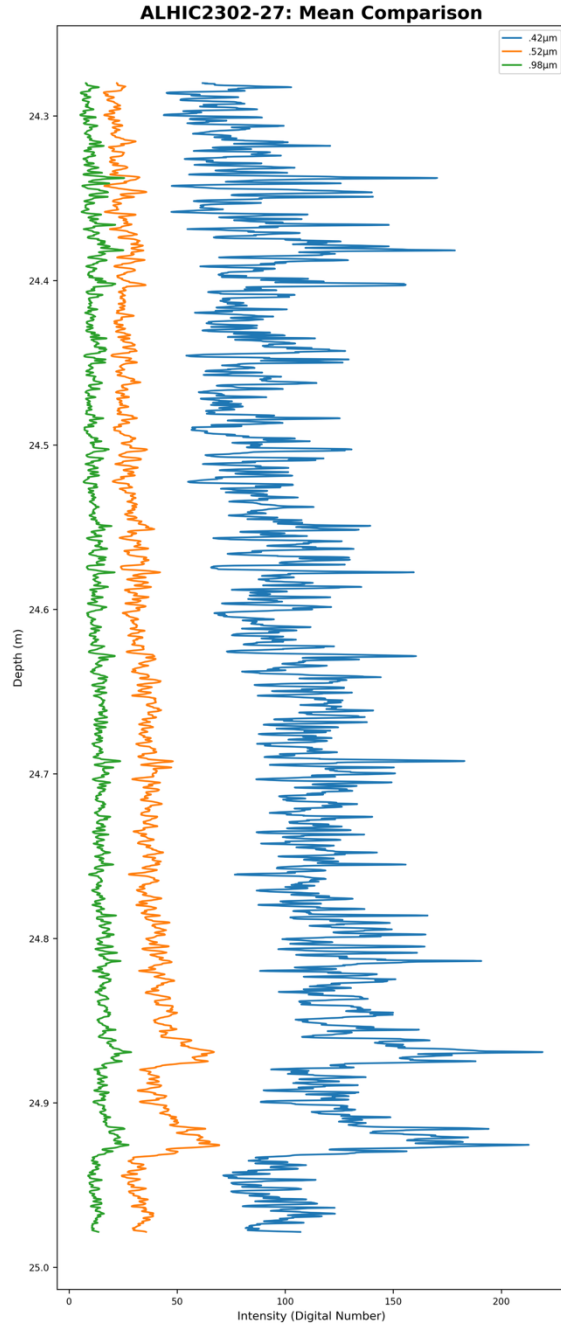


Figure 1.12 This plot depicts dust loading estimated using three distinct channels .42μm, .52μm, and .98μm in ALHIC2302-27.



Figure 1.13 This plot depicts the Blue-Green (.42μm and .52μm) Normalized Difference Index in yellow and the Visible-NIR (.52μm and .98μm) Normalized Difference Index in brown. The x-axis for VN-NDI is reversed for a mirrored comparison.

These signals were compared in order to

analyze the dust content within the core. The blue-

green normalized difference index (BG_NDI) and the visible-NIR normalized difference index (VN_NDI) were calculated using the given equations:

$$BG - NDI = \frac{Signal_{.42\mu m} - Signal_{.52\mu m}}{Signal_{.42\mu m} + Signal_{.52\mu m}} \text{ (Eq. 3)}$$

$$VN - NDI = \frac{Signal_{.52\mu m} - Signal_{.98\mu m}}{Signal_{.52\mu m} + Signal_{.98\mu m}} \text{ (Eq. 4)}$$

The BG_NDI is used to look at blue-specific absorption (soot/organics) and fine-particle effects, whereas the VN_NDI looks at strong scattering signatures common for mineral dust.

These differences are graphed for ALHIC2302 in Figure 1.13.

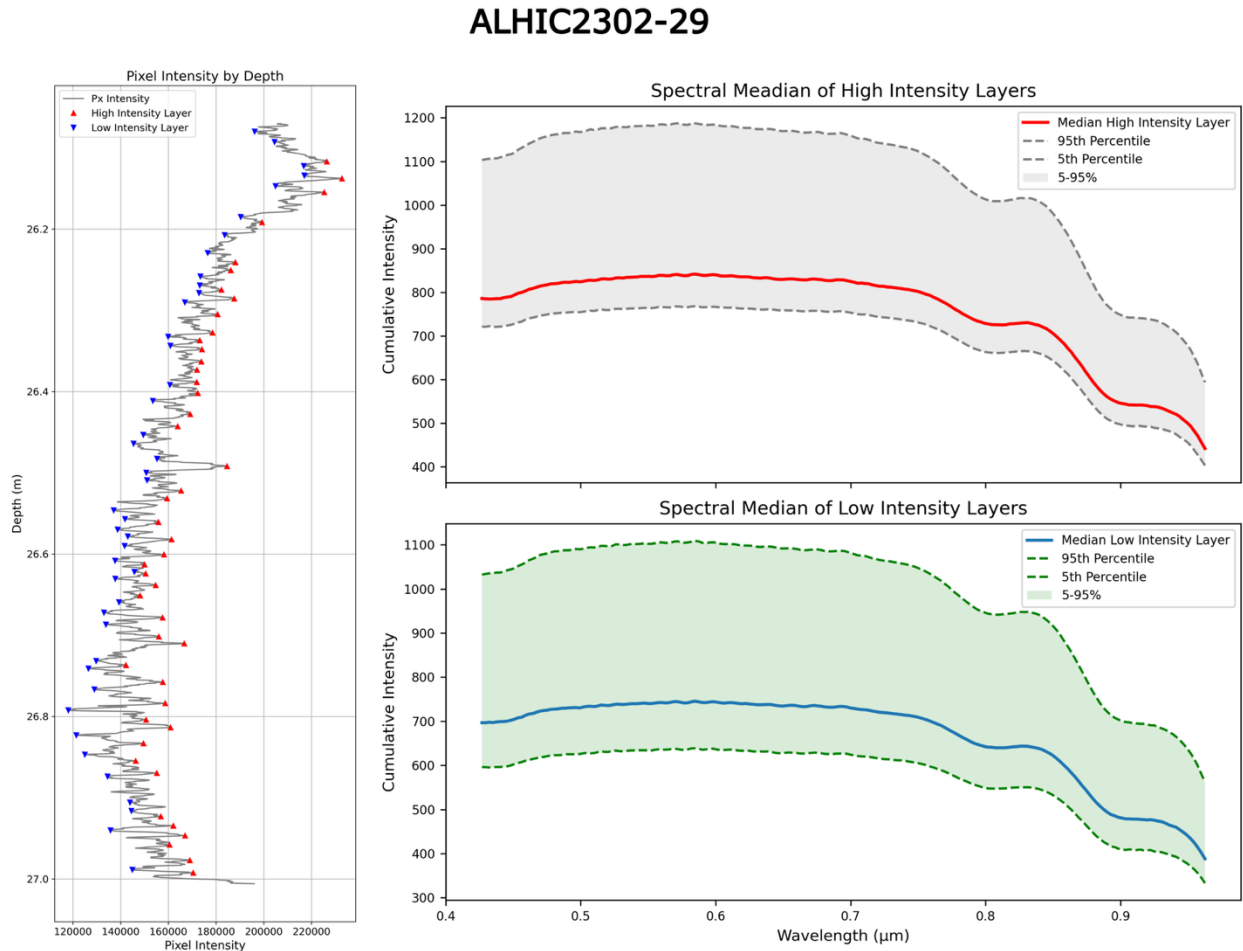


Figure 1.14 ALHIC2302-29 row-wise pixel intensity by depth, with bright (high intensity) and dark (low intensity) layers identified (left). A spectral graph of the median cumulative intensity for bright layers, with the 5-95 percentile range (top right). A spectral graph of the median cumulative intensity for dark layers, with the 5-95 percentile range (bottom right).

We compared the median spectral signal (Figure 1.14) of bright (red dot) and dark (Blue dot) layers in ALHIC2302-29. In both spectral plots, a drop in signal intensity was observed starting around .7 μ m and continuing through 1 μ m. This trend is expected as it follows the light transmitting properties of ice at different wavelengths. In an effort to determine where the bright and dark layers differed the most across spectra, both median spectral signals were normalized to .6 μ m using Eq.5. The normalized median spectral signals were subtracted at each wavelength, the resulting difference is shown in Figure 1.15.

$$\text{Normalized Signal} = \frac{\text{Signal}}{\text{Signal}_{.6\mu\text{m}}} \text{ (Eq. 5)}$$

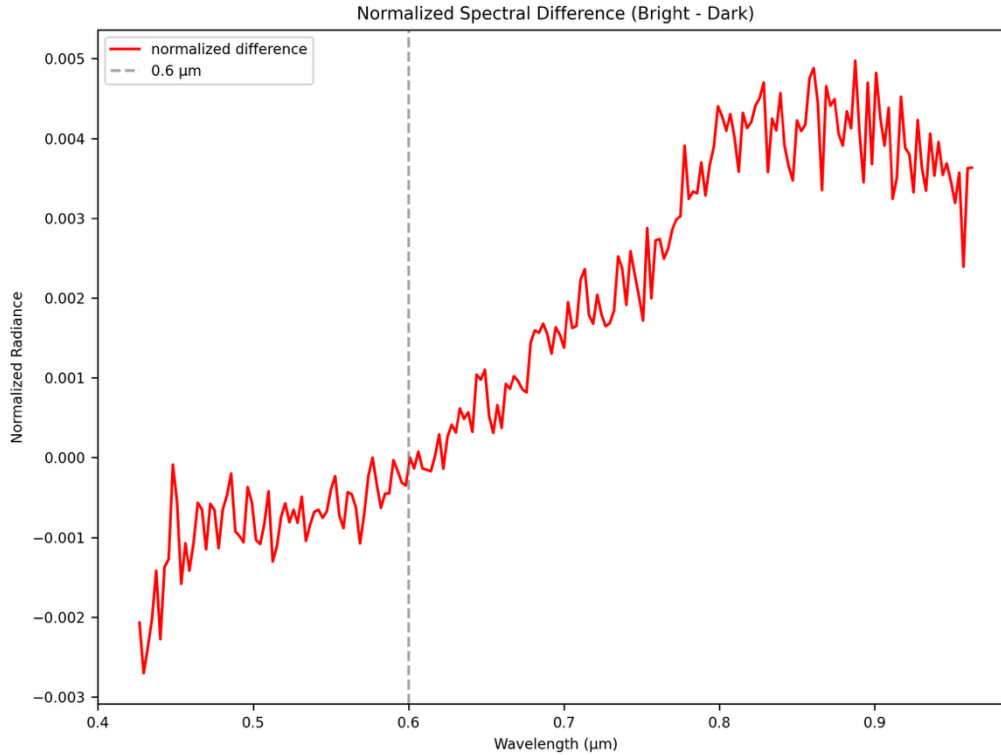


Figure 1.15 This plot depicts the difference between the normalized spectral median intensities for the bright and dark layers (Figure 13) after being normalized to .6 μ m.

Since the greatest difference in the bright and dark layers' spectral signal is observed to be around .8 μ m - .9 μ m, .85 μ m was compared with .6 μ m. Using these two wavelengths, the Dust Loading Index (DLI) was calculated using Eq. 6 for ALHIC2302-29.

$$DLI = \frac{Signal_{.6\mu m} - Signal_{.8\mu m}}{Signal_{.6\mu m} + Signal_{.8\mu m}} \text{ (Eq. 6)}$$

The resulting mean and median DLI signal from ALHIC2302-29 after bubble removal and inpainting is shown in Figure 1.16. Figure 1.17 shows the visual results of the DLI on ALHIC2302-29 before bubble removal, and while there are still many bright spots visible, we can also see the darker regions indicating potential layers.

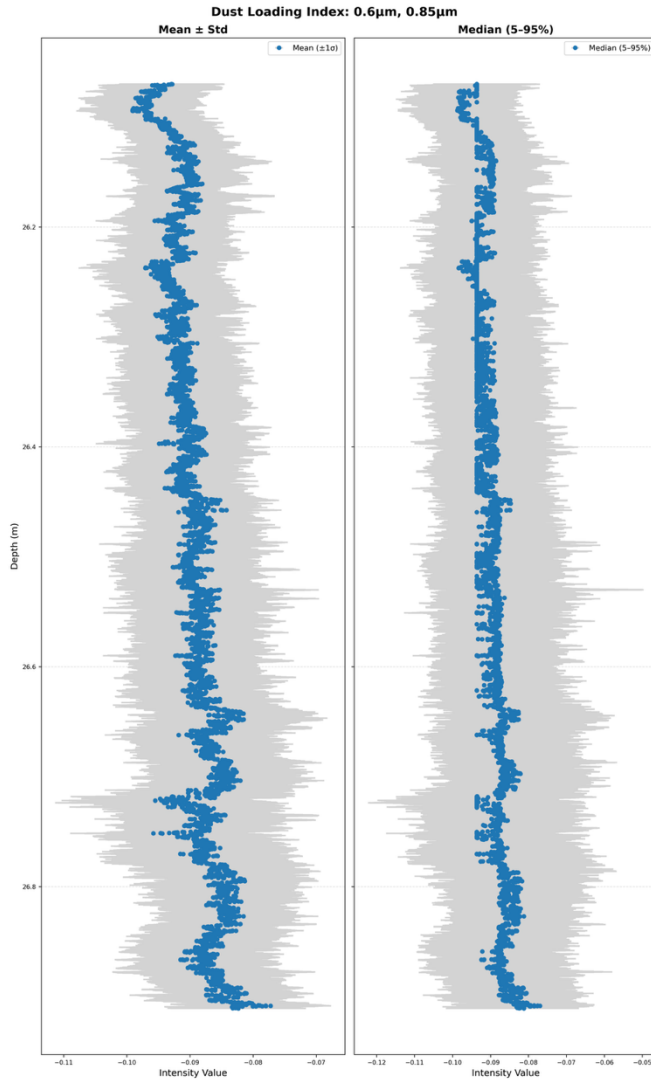
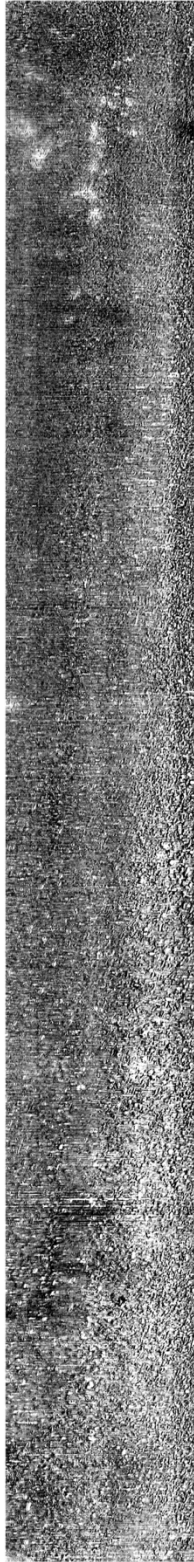


Figure 1.16 The mean (left) and median (right) row wise pixel intensities of ALHIC2302-29 by depth, after bubble removal, inpainting, and DLI application.



Conclusion

We developed an algorithm to detect bubbles in hyperspectral images of ice cores, based on machine vision software that maps regions of high pixel intensity. The developed algorithm isolates these bubble regions, counts and characterizes them.

As demonstrated above, hyperspectral images captured by the IceSpec system are fully capable of replicating layer flagging and bubble detection previously conducted on color or grayscale images captured with line scan technology. This study also shows suitability of images for indications of core azimuth changes in bubble orientations. Additionally, IceSpec images also preserves spectral information. With the bubbles cleaned up from the image, more traditional HSI methods can be used for material identification, opening a future opportunity to further investigate the impurities within the ice matrix.

Figure 1.17 A grayscale image of ALHIC2302-29 after applying the DLI.

CHAPTER 2

TESTING FEASIBILITY OF USING A UNET SEGMENTATION MODEL FOR VISIBLE FEATURE CLASSIFICATION IN HIGH RESOLUTION ICE CORE IMAGES

Background

In Chapter 1, I demonstrated how bubbles can be detected by identifying peaks in brightness throughout the hyperspectral images used in this research. In this chapter I explore a new method focused on accounting for and avoiding false positives caused by quality issues within ice cores (e.g., crack, fractures, etc.). A supervised learning algorithm trained on a custom dataset of labeled ice core images from the IceSpec system was used with two goals in mind. The first was to increase the accuracy of bubble detection as relying on pixel brightness thresholds resulted in many bubbles being overlooked, and the second was to classify other features of interest (e.g., ice edges and reference targets) to automate image trimming and pixel dimension calculations.

Machine learning is the development of computational algorithms that can learn and adapt without explicit instruction (Sarker, 2021). These algorithms are designed to automatically analyze data, recognize patterns, and draw inferences, making their use extremely beneficial.

There are many approaches to build custom machine learning pipelines using existing algorithms to train other algorithms, but all include using labeled datasets. Labeling data is essential for machine learning algorithms, especially those that utilize supervised learning as this is what is used in their training; annotations (e.g., labels) on existing data are used to predict the classification of new data. This differs from unsupervised learning, where the algorithm highlights patterns that it recognizes and does not require pre-existing labeled data.

Machine learning has proven its applicability to ice core research. Wheatley et al., (2012) used an automated process for determining annual layers in NGRIP by looking at the chemical signals by depth, a method that involves classifying sections of the signal as Peak, Descending, Trough, Ascending, or Issue. By understanding how the chemical signals reflect seasonal cycles, they could automatically highlight annual layers. Maffezzoli et al., (2023) coupled flow imaging microscopy with a deep neural network to automate classification of particles in ice core melt water. Their algorithm was trained to classify particles as dust, tephra, pollen, or contaminants and was able to achieve 96.8% accuracy. Winstrup (2011) applied an unsupervised learning algorithm based on Hidden Markov Modeling to automate visible layer detection in the NGRIP ice core. At the time of publication, their algorithm was unable to provide accurate chronology, but has since become a widely accepted method in layer counting and was used to date the WAIS Divide deep ice core (WDC06A) (Sigl et al., 2016). Additionally, an adjacent implementation by Ghanbari et al., (2022) used hyperspectral imaging on sediment cores and deployed a convolutional neural network to map particle size.

Methods

The training dataset library is organized in a structured hierarchy of nested directories using ice core Region/Project/Core/CoreSection. Region represents the geographic location where the ice core was extracted and collected, Project represents the name of the project that collected the ice core, Core represents the name of the core being analyzed, and CoreSection represents a specific section of the core being analyzed.

Each CoreSection folder had several hyperspectral images stored in PNG format using compatible red, green, and blue channels. These images were created using different light settings as the reflective and transmitted lighting illuminated the ice and its visible features in

different ways. These images were loaded into an open source Python based interactive viewer, napari (Sofroniew et al, 2025). For each CoreSection, digital images were aligned by the top pixels and then manually labeled using built in tools. This was achieved by zooming-in until individual pixels were identified and using a click-and-drag tool to draw polygons around each visible feature and reference target (see Figure 2.2).

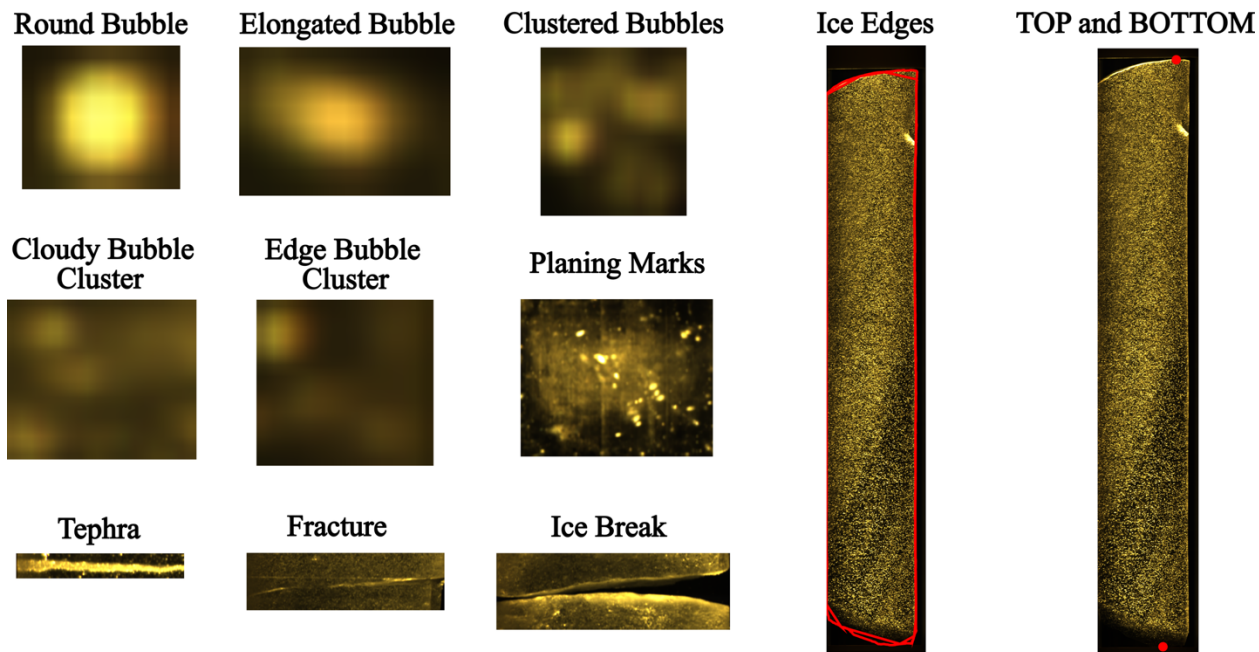


Figure 2.1 Examples of visible features seen in ice core images.

Visible features were classified as RoundBubbles, ElongatedBubbles, IceEdges, ClusteredBubbles, Tephra, CloudyBubbleClusters, Fractures, EdgeBubbleClusters, PlaningMarks, IceBreakBoundary, and Misc. Examples of these features are shown in Figure 2.1. The top and bottom of the ice core sections were labeled with single points, called TOP and BOTTOM. The standards visible in the image were labeled 20%White, 50%White, 100%White. The spectral standards (Figure 1.1 panel c) were labeled as KnownStandardsTrifecta, and the focus chart (Figure 1.1 top of panel a) was labeled as LensTestChart. Each of these labels were output into Comma Separated Values (.csv) files, one for each classification. Each .csv file contains polygon vertices in the format (index, shape-type, vertex-index, axis-0, axis-1), where

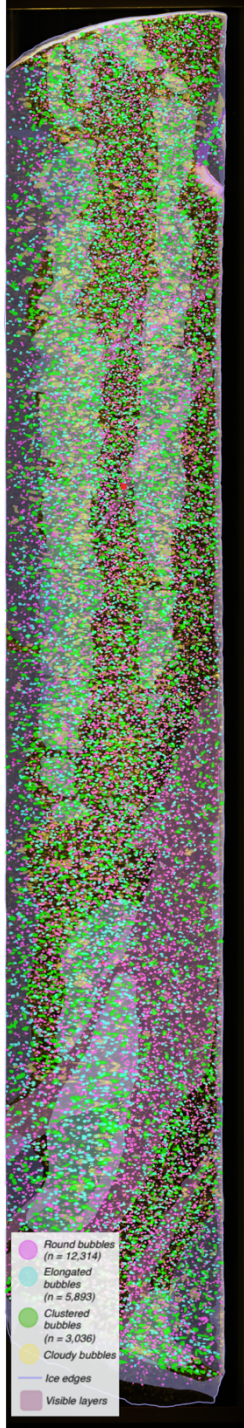


Figure 2.2
ALHIC2302-29
after annotation.

axis-0 and axis-1 represent pixel coordinates as (y, x) pairs. These files were organized into the proper CoreSection folder along with the corresponding images. This was completed for eight core sections from various ice core sites; Denali: DEN13A-227, DEN13A-228, Allan Hills: ALHIC2302-29, ALHIC2302-46, South Pole: SPC14-1317, WAIS Divide: WDC06A-3054-A, WDC06A-3054-B, and Greenland Ice Sheet 2: GISP2D-2546.

As in Chapter 1, for processing Python computing language (version 3.12.2) and jupyter notebooks (JupyterLab 4.3.4) running on a Debian 12 Linux OS server (referred to as jetstream) with an Intel(R) Xeon(R) Gold 6140 CPU and 188GB of main memory were used. For machine learning, PyTorch (version 2.6.0) and torchvision (version 0.21.0) were utilized. The jetstream server was also equipped with 2 RTX A5000 NVIDIA GPUs, each with 24GB of GPU memory.

To programmatically load the dataset, I developed a customized function, `index_samples()`, to traverse the training datasets. For each CoreSection folder, `index_samples()` pulled the raw images and any associated .csv annotation files and checked the PNG images for matching sizes. If images were mismatched by a couple pixels, cropping or padding at the end of the line scan was implemented. The images were converted into grayscale and stacked into three channels, padding one channel for core sections that only have two images. Next the .csv files were parsed using a

custom `parseLabels` script that sorts vertices by the vertex-index and converts the pixel

coordinate (y, x) pairs into (x, y) for drawing. The polygons drawn around each visible feature were grouped based on the global CLASS_MAP, which assigns each CSV filename a numerical class ID. For each CoreSection the polygons get rasterized into a semantic mask, a single-channel integer image where each pixel stores the class ID of the feature covering it, or 0 if the pixel is unlabeled background. This rasterization was made deterministic and consistent using the Python Imaging Library (PIL) polygon fill operations.

Using a random 80/20 split, I used six core sections to train the algorithm and two for validation. The training and validation sets were wrapped into their own DataLoaders for batching and parallel reading. Since there was not enough space to load the full images, the training used a random crop for each tensor, with a size of 1000x1000 pixels.

For segmentation, the model used was UNet with a ResNet-18 backbone from torchvision. It was initialized using ImageNet pretrained weights and then replacing the classifier head to match the number of ice core visible feature classes (background + all csv types in CLASS_MAP). However, these pretrained weights proved to be insufficient (see discussion) and custom weights were implemented.

Training used a standard UNet segmentation loop where images and masks were forwarded through the model, the prediction tensor was compared to the ground-truth mask using a weighted dice loss, and gradients were backpropagated using mixed precision via GradScaler. After each epoch, the model was evaluated in no-grad mode, computing loss and mean Intersection-over-Union (mIoU) using a confusion matrix over all classes. We looped over 100 epochs and the training loss, validation loss, and mIoU was output to measure performance (Terven et al., 2023). The predicted masks and ground truth masks of select epochs were displayed for inspection.

Results and Discussion

A training loss of 2 or more with a very low mean Intersection-over-Union implies the model is randomly guessing and not actually learning. While this is fine, and even expected, for the first epoch, there should be a noticeable decrease in loss and increase in mIoU as later epochs are approached. Initial results from the jetstream server (Table 2.1) demonstrate a near stagnation

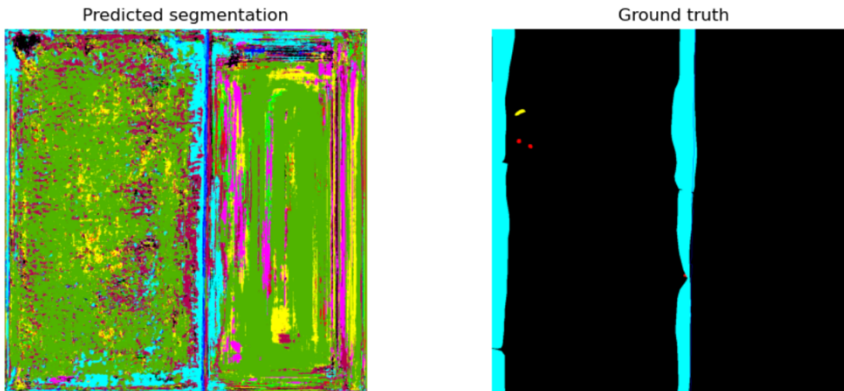


Figure 2.3 The prediction mask (left) and the ground truth mask (right) after looping 100 epochs of the initial machine learning algorithm (before properly weighting the background classification).

in training loss and mIoU.

While there was a decrease in validation loss, those results proved that the model was failing to learn. Inspection of the prediction mask (Figure 2.3) led to the conclusion

that background space needed to be heavily weighted above the physical features.

With this in mind, I adjusted the background to be weighted at 2 and the rest of the feature classes at 1. In doing so, the jetstream was able to produce more stable results.

Epoch	Training Loss	Validation Loss	mIoU
01	2.3153	4.1773	0.024
02	2.2951	3.5040	0.030
03	2.2937	2.8431	0.031
04	2.2789	2.3081	0.039
05	2.3085	2.3272	0.029
06	2.3115	2.3658	0.033
65	2.3035	2.7293	0.030
90	2.2433	2.7053	0.031
100	2.3032	2.7694	0.029

Table 2.1 Sample of the machine learning model's results before proper class weighting.

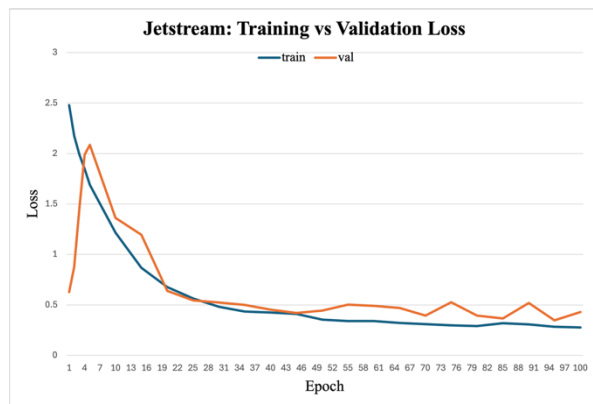


Figure 2.4 Graph of training loss vs validation loss on the machine learning algorithm on the jetstream server through 100 epochs.

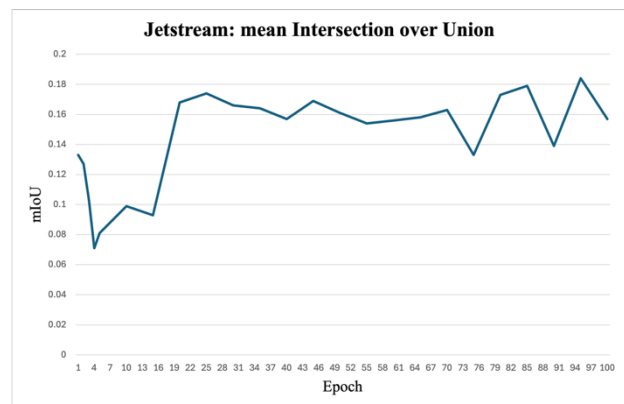


Figure 2.5 Graph of mean Intersection over Union (mIoU) on the machine learning algorithm on the jetstream server through 100 epochs.

Randomly cropped sections from ice cores DEN13A-228, SPC14-1317, ALHIC2302-46, GISP2D-2546, ALHIC2302-29, and WDC06A-3054-B were used for training while ice cores DEN13A-227 and WDC06A-3054-A were used for validation. Figures 2.4 and 2.5 show the

graphs of the algorithm's training versus validation loss and mIoU, respectively. These results show a noticeable decrease in loss and increase in mIoU as epochs are looped through with the training loss decreasing as low as 0.277 and the average mIoU getting as high as 0.179. This indicates that some learning is happening, but large spikes associated with mIoU towards the end of the epochs indicates an issue. Additionally, it is noteworthy that the mIoU does not reach above 0.2, with 0.5 being the common threshold for an acceptable Intersection over Union, this suggests the model still requires tweaking.

However, these results were obtained from small, randomly cropped sections of each labeled core (see Methods) leaving unanswered the question of if training on full images would improve the model's results. To remedy this, UCAR (University Corporation for Atmospheric Research) allocated the use of 2,500 core hours on the derecho supercomputer's GPU. Derecho has 82 GPU nodes with 512 GB DDR4 memory per node. The code base and training datasets were uploaded and once ensuring that the processing would remain the same as it was on the jetstream server, the allocated core hours were spent to loop over 100 epochs. The training loss, validation loss, and mIoU were output into a .log file and the predicted and ground truth masks were output as images.

The first run of full image processing on the Derecho GPU demonstrated results with both loss functions creating decreasing logarithmic curves (Figure 2.6). Validation loss plateaus

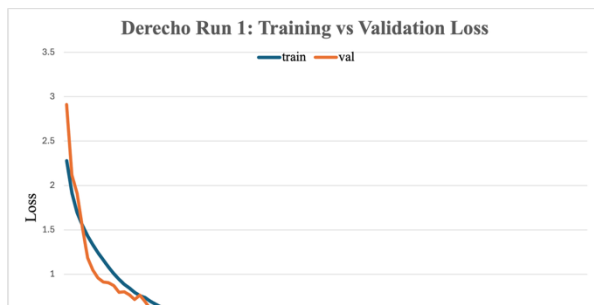


Figure 2.6 Graph of training loss vs validation loss on the machine learning algorithm on the derecho server through 100 epochs.

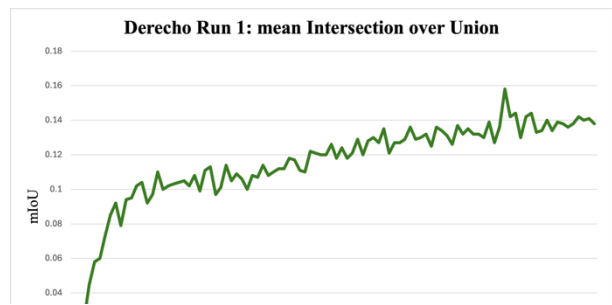


Figure 2.7 Graph of mIoU on the machine learning algorithm on the derecho server through 100 epochs.

but still produces results as low as 0.372, and training loss results as low as 0.141. The mIoU for this run has a more stable shape than the jetstream was able to produce (Figure 2.7). However, the highest point, a spike at epoch 83, was only 0.158; still well below the threshold for acceptable. Figure 2.8 shows the prediction and ground truth masks of epoch 5 and 85 for this run.

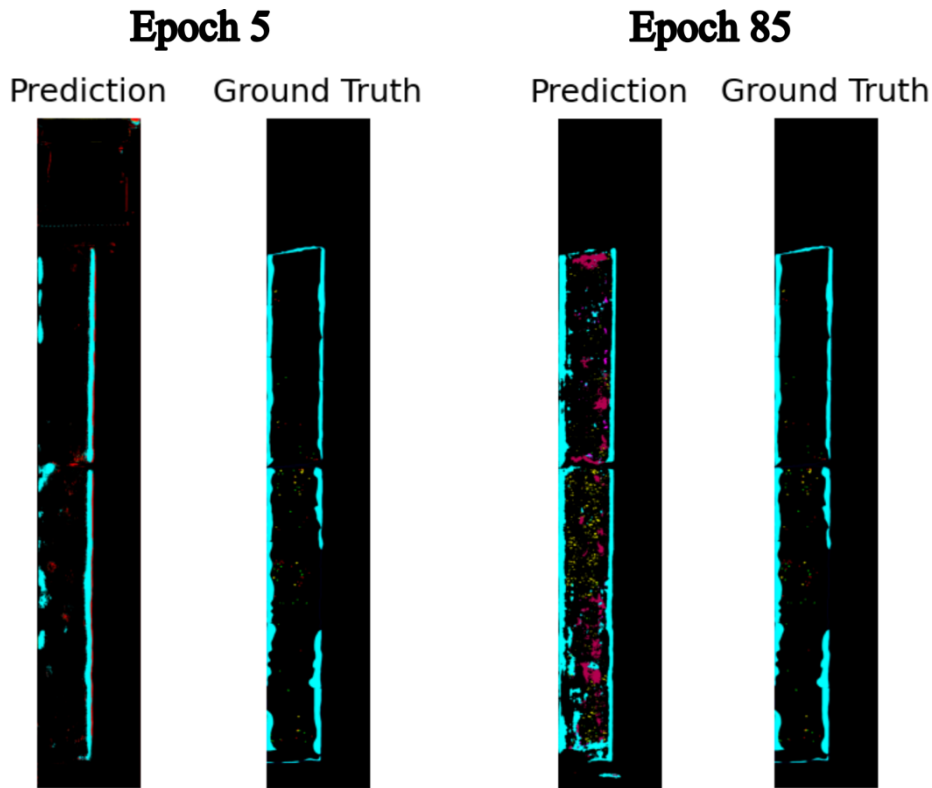


Figure 2.8 The prediction masks and the ground truth masks after looping 5 (left) and 85(right) epochs of the machine learning algorithm on the derecho server.

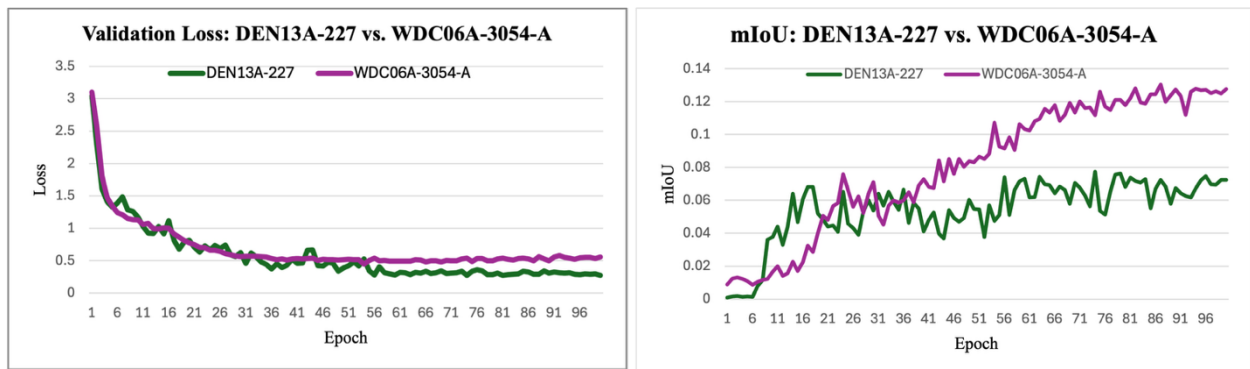


Figure 2.9 Graphs comparing validation loss and mIoU for WDC06A-3054-A and DEN13A-227 on the machine learning algorithm on the derecho server through 100 epochs.

When looking at loss and mIoU for the two validation sets separately, WDC06A-3054-A has a greater increase in mIoU but the validation loss plateaus earlier than DEN13A-227 (Figure 2.9).

The prediction and ground truth masks have shown that this system becomes quite adept at learning the ice edges, shown with light blue in Figure 2.7. This was expected as machine learning quickly understands and recognizes straight lines.

Conclusion

This machine learning pipeline is still evolving, but the proof of concept has been demonstrated. One main limitation was the small size of the available training library. There were only 8 full cores annotated, while most machine learning libraries have hundreds of annotated datasets.

Continuing to add more labeled datasets, will allow the algorithm to become more refined and eventually be able to automatically detect visible features in ice cores with greater accuracy. However, since the 8 annotated datasets took months to complete, future work might include creating annotated datasets of smaller, cropped images to ensure more variation and a greater dataset library.

Doing machine learning on the full HSI datasets, rather than on PNG images, may also prove beneficial in future work, as more channels could provide more data for the algorithm to learn from. This would likely require greater computer resources, due to the very large datasets that would need to be processed by the GPU.

REFERENCES

- Alley, R. B., J.H. Perepezko, and C.R. Bentley. “Grain Growth in Polar Ice: I. Theory.” *Journal of Glaciology* 32, no. 112 (1986): 415–24. <https://doi.org/10.1017/s0022143000012120>.
- Alley, R. B., C. A. Shuman, D. A. Meese, A. J. Gow, K. C. Taylor, K. M. Cuffey, J. J. Fitzpatrick, et al. “Visual-Stratigraphic Dating of the GISP2 Ice Core: Basis, Reproducibility, and Application.” *Journal of Geophysical Research: Oceans* 102, no. C12 (November 30, 1997): 26367–81. <https://doi.org/10.1029/96jc03837>.
- Andersen, Katrine K., Anders Svensson, Sigfus J. Johnsen, Sune O. Rasmussen, Matthias Bigler, Regine Röthlisberger, Urs Ruth, Marie-Louise Siggaard-Andersen, Jørgen Peder Steffensen, and Dorte Dahl-Jensen. “The Greenland Ice Core Chronology 2005, 15–42ka. Part 1: Constructing the Time Scale.” *Quaternary Science Reviews* 25, no. 23-24 (December 2006): 3246–57. <https://doi.org/10.1016/j.quascirev.2006.08.002>.
- Askebjerg, P., S. W. Barwick, L. Bergström, A. Bouchta, S. Carius, E. Dalberg, B. Erlandsson, et al. “UV and Optical Light Transmission Properties in Deep Ice at the South Pole.” *Geophysical Research Letters* 24, no. 11 (June 1, 1997): 1355–58. <https://doi.org/10.1029/97gl01246>.
- Bhargava, Anuja, Ashish Sachdeva, Kulbhushan Sharma, Mohammed H. Alsharif, Peerapong Uthansakul, and Monthippa Uthansakul. “Hyperspectral Imaging and Its Applications: A Review.” *Heliyon* 10, no. 12 (June 30, 2024): e33208. <https://doi.org/10.1016/j.heliyon.2024.e33208>.
- Boggs, Thomas. “Spectralpython/Spectral: Python Module for Hyperspectral Image Processing.” GitHub, April 13, 2025. <https://github.com/spectralpython/spectral>.
- Bradley, L., Brigitta Sipőcz, Thomas Robitaille, Erik Tollerud, Zé Vinícius, Christoph Deil, Kyle Barbary, et al. “Astropy/photutils: 2.0.2”. Zenodo, October 25, 2024. <https://doi.org/10.5281/zenodo.13989456>.
- Bradski, G. (2000) The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*, 120; 122-125.
- Bramall, N. E., R. C. Bay, K. Woschnagg, R. A. Rohde, and P. B. Price. “A Deep High-Resolution Optical Log of Dust, Ash, and Stratigraphy in South Pole Glacial Ice.” *Geophysical Research Letters* 32, no. 21 (November 1, 2005). <https://doi.org/10.1029/2005gl024236>.
- Cole-Dai, Jihong, Drew M. Budner, and Dave G. Ferris. “High Speed, High Resolution, and Continuous Chemical Analysis of Ice Cores Using a Melter and Ion Chromatography.” *Environmental Science & Technology* 40, no. 21 (October 4, 2006): 6764–69. <https://doi.org/10.1021/es061188a>.
- Computational and Information Systems Laboratory. 2023. Derecho: HPE Cray EX System (University Community Computing). Boulder, CO: NSF National Center for Atmospheric Research. doi:10.5065/qx9a-pg09.

- Fegyveresi, J.M., R.B. Alley, M.K. Spencer, J.J. Fitzpatrick, E.J. Steig, J.W.C. White, J.R. McConnell, and K.C. Taylor. “Late-Holocene Climate Evolution at the WAIS Divide Site, West Antarctica: Bubble Number-Density Estimates.” *Journal of Glaciology* 57, no. 204 (2011): 629–38. <https://doi.org/10.3189/002214311797409677>.
- Garzonio, Roberto, Biagio Di Mauro, Sergio Cogliati, Micol Rossini, Cinzia Panigada, Barbara Delmonte, Valter Maggi, and Roberto Colombo. 2018. “A Novel Hyperspectral System for High Resolution Imaging of Ice Cores: Application to Light-Absorbing Impurities and Ice Structure.” *Cold Regions Science and Technology* 155 (July): 47–57. <https://doi.org/10.1016/j.coldregions.2018.07.005>.
- Ghanbari, Hamid, and Dermot Antoniades. 2022. “Convolutional Neural Networks for Mapping of Lake Sediment Core Particle Size Using Hyperspectral Imaging.” *International Journal of Applied Earth Observation and Geoinformation* 112 (July): 102906. <https://doi.org/10.1016/j.jag.2022.102906>.
- Giorgino, Toni. “Computing and Visualizing Dynamic Time Warping Alignments InR: ThedtwPackage.” *Journal of Statistical Software* 31, no. 7 (2009). <https://doi.org/10.18637/jss.v031.i07>.
- Gow, A. J. “Research Report 177 - Snow Studies in Antarctica.” ERDC. Cold Regions Research and Engineering Laboratory (U.S.), November 1965. <https://erdc-library.erdcresearch.org/items/81b728f7-6b83-4ef8-e053-411ac80adeb3/full>.
- Grieman, Mackenzie M., Helene M. Hoffmann, Jack D. Humby, Robert Mulvaney, Christoph Nehrbass-Ahles, Julius Rix, Elizabeth R. Thomas, Rebecca Tuckwell, and Eric W. Wolff. “Continuous Flow Analysis Methods for Sodium, Magnesium and Calcium Detection in the Skytrain Ice Core.” *Journal of Glaciology* 68, no. 267 (July 9, 2021): 90–100. <https://doi.org/10.1017/jog.2021.75>.
- Kirkpatrick, Liam Reed, Austin Joseph Carter, Julia Marks-Peterson, Sarah Shackleton, and T.J. Fudge. 2025. “Three-Dimensional Multitrack Electrical Conductivity Method for Interpretation of Complex Ice Core Stratigraphy.” *Journal of Glaciology* 71 (2025): e105. <https://doi.org/10.1017/jog.2025.10081>.
- Koren, Norman. “Understanding Resolution and MTF.” www.normankoren.com, February 26, 2007. <https://www.normankoren.com/Tutorials/MTF.html>.
- Lipinkov, Vladimir Ya. “Air Bubbles and Air-Hydrate Crystals in the Vostok Ice Core.” *Physics of Ice Core Records* (Hokkaido University), January 1, 2000, 327–58.
- Maffezzoli, Niccolò, Eliza Cook, Willem G. M. van der Bilt, Eivind N. Støren, Daniela Festi, Florian Muthreich, Alistair W. R. Seddon, et al. 2023. “Detection of Ice Core Particles via Deep Neural Networks.” *The Cryosphere* 17 (2): 539–65. <https://doi.org/10.5194/tc-17-539-2023>.

- McGwire, Kenneth C, Geoffrey M Hargreaves, Richard B Alley, Trevor J Popp, David B Reusch, Matthew K Spencer, and Kendrick C Taylor. “An Integrated System for Optical Imaging of Ice Cores.” *Cold Regions Science and Technology* 53, no. 2 (July 2008): 216–28. <https://doi.org/10.1016/j.coldregions.2007.08.007>.
- Morcillo, G, Sérgio H Faria, and Sepp Kipfstuhl. “Unravelling Antarctica’s Past through the Stratigraphy of a Deep Ice Core: An Image-Analysis Study of the EPICA-DML Line-Scan Images.” *Quaternary International* 566-567 (November 1, 2020): 6–15. <https://doi.org/10.1016/j.quaint.2020.07.011>.
- Neff, Peter D. “A Review of the Brittle Ice Zone in Polar Ice Cores.” *Annals of Glaciology* 55, no. 68 (2014): 72–82. <https://doi.org/10.3189/2014aog68a023>.
- Osterberg, Erich C, Michael Handley, Sharon B Sneed, Paul Andrew Mayewski, and Karl J Kreutz. “Continuous Ice Core Melter System with Discrete Sampling for Major Ion, Trace Element, and Stable Isotope Analyses.” *Environmental Science & Technology* 40, no. 10 (April 8, 2006): 3355–61. <https://doi.org/10.1021/es052536w>.
- Sarker, Iqbal H. 2021. “Machine Learning: Algorithms, Real-World Applications and Research Directions.” *SN Computer Science* 2 (3): 1–21. <https://link.springer.com/article/10.1007/s42979-021-00592-x>.
- Shackleton, S, V Hishamunda, L Davidge, E Brook, J Marks Peterson, A Carter, S Aarons, et al. 2025. “Miocene and Pliocene Ice and Air from the Allan Hills Blue Ice Area, East Antarctica.” *Proceedings of the National Academy of Sciences* 122 (44): e2502681122–22. <https://doi.org/10.1073/pnas.2502681122>.
- Sigl, Michael, Tyler J. Fudge, Mai Winstrup, Jihong Cole-Dai, David Ferris, Joseph R. McConnell, Ken C. Taylor, et al. “The WAIS Divide Deep Ice Core WD2014 Chronology – Part 2: Annual-Layer Counting (0–31 ka BP).” *Climate of the Past* 12, no. 3 (March 30, 2016): 769–86. <https://doi.org/10.5194/cp-12-769-2016>.
- Sofroniew, Nicholas, Talley Lambert, Grzegorz Bokota, Juan Nunez-Iglesias, Peter Sobolewski, Andrew Sweet, Lorenzo Gaifas, et al. “Napari: A Multi-Dimensional Image Viewer for Python.” Zenodo (CERN European Organization for Nuclear Research), October 16, 2025. <https://doi.org/10.5281/zenodo.17367124>.
- Specim. “Optimizing Laboratory Hyperspectral Imaging: Best Practices - Specim.” Specim, June 18, 2025. <https://www.specim.com/technology/optimizing-laboratory-hyperspectral-imaging/>.
- Stoll, Nicolas, Ilka Weikusat, Daniela Jansen, Paul Bons, Kyra Darányi, Julien Westhoff, María-Gema Llorens, et al. “Linking Crystallographic Orientation and Ice Stream Dynamics: Evidence from the EastGRIP Ice Core.” *The Cryosphere* 19, no. 9 (September 15, 2025): 3805–30. <https://doi.org/10.5194/tc-19-3805-2025>.

- Stolz, M. R., and M. Ram. “Using Laser-Light Scattering to Measure Impurities, Bubbles, and Imperfections in Ice Cores.” *Journal of Geophysical Research: Atmospheres* 110, no. D11 (June 10, 2005). <https://doi.org/10.1029/2004jd005589>.
- Svensson, Anders. “Visual Stratigraphy of the North Greenland Ice Core Project (NorthGRIP) Ice Core during the Last Glacial Period.” *Journal of Geophysical Research* 110, no. D2 (January 21, 2005). <https://doi.org/10.1029/2004jd005134>.
- Takata, Morimasa, Yoshinori Iizuka, Takeo Hondoh, Shuji Fujita, Yoshiyuki Fujii, and Hitoshi Shoji. “Stratigraphic Analysis of Dome Fuji Antarctic Ice Core Using an Optical Scanner.” *Annals of Glaciology* 39 (2004): 467–72. <https://doi.org/10.3189/172756404781813899>.
- Terven, Juan, Diana M. Cordova-Esparza, Alfonzo Ramirez-Pedraza, Edgar A. Chavez-Urbiola, and Julio A. Romero-González. 2023. “Loss Functions and Metrics in Deep Learning. A Review.” *ArXiv.org*. July 5, 2023. <https://doi.org/10.48550/arXiv.2307.02694>.
- Truong, Charles, Laurent Oudre, and Nicolas Vayatis. “Selective Review of Offline Change Point Detection Methods.” *Signal Processing* 167 (February 2020). <https://doi.org/10.1016/j.sigpro.2019.107299>.
- Walt, Stéfan van der, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. “Scikit-Image: Image Processing in Python.” *PeerJ* 2 (June 2014): e453. <https://doi.org/10.7717/peerj.453>.
- Warren, Stephen G. 2019. “Optical Properties of Ice and Snow.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 377 (2146): 20180161. <https://doi.org/10.1098/rsta.2018.0161>.
- Wheatley, J. J., P. G. Blackwell, N. J. Abram, J. R. McConnell, E. R. Thomas, and E. W. Wolff. 2012. “Automated Ice-Core Layer-Counting with Strong Univariate Signals.” *Climate of the Past* 8 (6): 1869–79. <https://doi.org/10.5194/cp-8-1869-2012>.
- Winstrup, Mai. 2011. “An Automated Method for Annual Layer Counting in Ice Cores -and an Application to Visual Stratigraphy Data from the NGRIP Ice Core.” <https://nbi.ku.dk/english/theses/phd-theses/mai-winstrup/>.

APPENDIX A

SUPPLEMENTARY MATERIALS

Code Availability

The full analysis pipeline will be available on Github and Zenodo.

The repository is organized as:

- Notebooks/: Jupyter notebooks for interactive exploration, visualization, and prototyping of analysis workflows.
- icespec/: Sample acquisition and preprocessing algorithms. Handles raw hyperspectral data ingestion, calibration, and preparation for downstream analysis.
- bubbles/: Algorithms for bubble detection and bubble data processing. Identifies and characterizes air bubbles within ice core images.
- Sums/: Algorithms for basic layer flagging. Detects and marks stratigraphic layers based on spectral summation methods.
- Dust/: Algorithms for detecting and quantifying dust loading in ice cores.

Data Availability

All hyperspectral datasets are hosted by the US Antarctic Program Data Center or the Arctic Data Center.

Workflow Overview

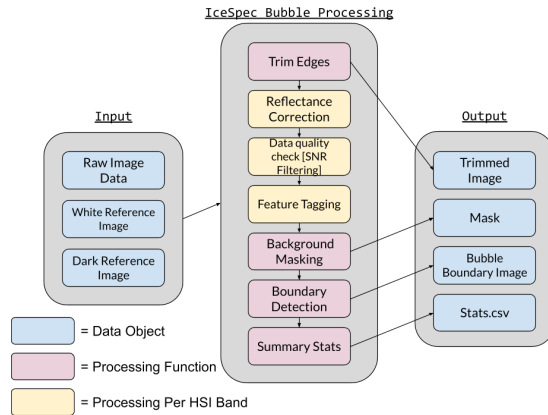


Figure A.1 Hyperspectral bubble detection and analysis pipeline.

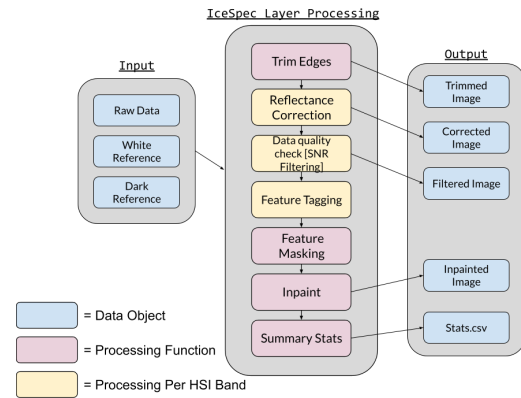


Figure A.2 Hyperspectral bubble removal and layer flagging pipeline.

Bubble detection and analysis pipeline stages:

1. Load hyperspectral cube and references
2. Trim the edges
3. Apply reflectance correction
4. Filter noisy spectral bands for data quality
5. Perform feature tagging
6. Mask background ice
7. Feature boundary detection
8. Extract summary statistics of geometric properties (area, elongation, orientation)

Bubble removal and layer flagging pipeline stages:

1. Load hyperspectral cube and references
2. Trim the edges
3. Apply reflectance correction
4. Filter noisy spectral bands

5. Perform feature tagging
6. Mask features
7. Image inpainting
8. Row-wise summary statistics of image

APPENDIX B

MODULATION TRANSFER FUNCTION

Project Overview

Modulation Transfer Function (MTF) characterizes the transfer of contrast from an object to an image and has been used by many photographers as a metric of resolution over the less accurate measurement of pixel per given space. MTF is especially relevant in the context of spatial frequencies associated with fine details in ice core structures, like layers and microcracks. The hyperspectral imaging of ice cores necessitates a careful consideration of MTF to ensure accurate representation of features corresponding to climate markers and impurities. The goal of this project was to graph the level of image sharpness a camera can capture, allowing for confidence in analyzing fine details and impurities in ice cores imaging.

Code Availability

- DOI: Rumsey, Roisin, Andrei Kurbatov, and Daniel J. Breton. “IceSpec Modular Transfer Function”. Zenodo, October 10, 2025. <https://doi.org/10.5281/zenodo.17315747>.

Methods / Algorithms

- `sfrcalc()`: Calculate SFR or MTF from an Excel-formatted text data file from ImageJ.

Results

Figure B1 shows the results of the hyperspectral camera’s spatial frequency response (SFR) across the captured spectral range. The sine wave shows pixel intensity, with black lines creating peaks and white lines creating valleys. As the lines on the lens test chart get closer together, it becomes difficult for the camera to process space between the lines and the intensity falls. When

contrast falls below a certain threshold, the data becomes inaccurate. MTF falls below 20% at roughly the same point for each spectral point, about 80-85 lp/mm. The low spectral frequency seems to struggle the most with spatial frequency.

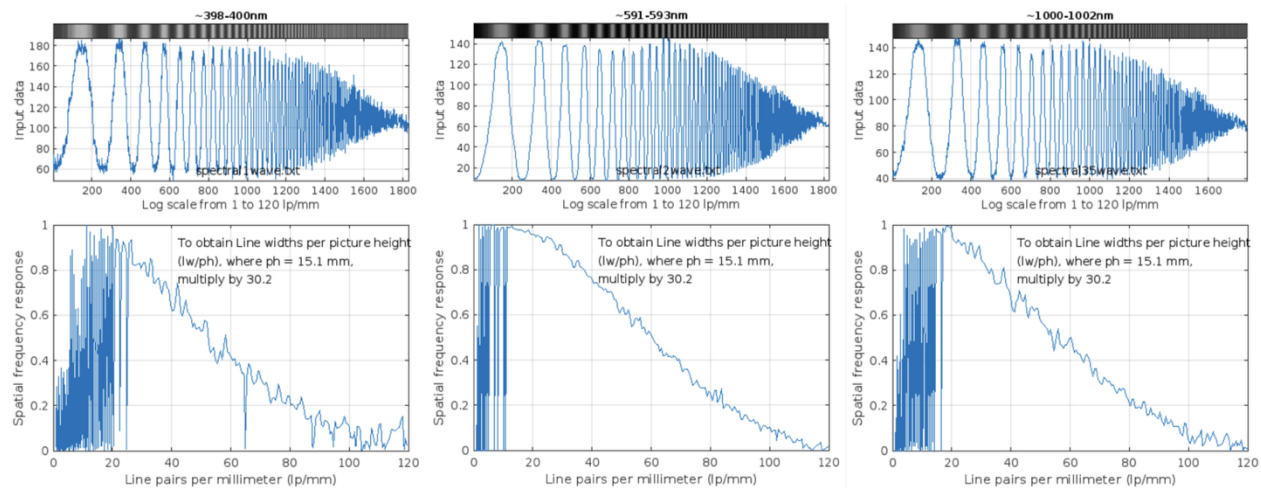


Figure B.1 The image shows the pixel intensity sine wave graphed above the spatial frequency plot for 398-400nm (left), 591-593nm (center), and 1000-1002nm (right).

Future Work / Data Plans

This data served as a foundation for much of the work done in this thesis as it established a proper metric for image resolution, indicated the need to rely on higher spectral wavelengths to resolve finer details, and provided a level of confidence in analyzing fine details and impurities in ice cores imaging. The spatial frequency response for the lowest wavelengths captured by the hyperspectral camera also showed the need for filtering out those bands as image processing occurs.

APPENDIX C

ALHIC1901

Project Overview

This project focused on very small (7cm by 3cm) samples of ice from ALHIC1901-228_4. The goal was to determine whether the bubble analysis could suggest a dip in the stratigraphic layers that has been seen by collaborators (Kirkpatrick et al., 2025)

Data Availability

- Dataset: ALHIC1901 Hyperspectral Scans will be available on the US Antarctic Program Data Center

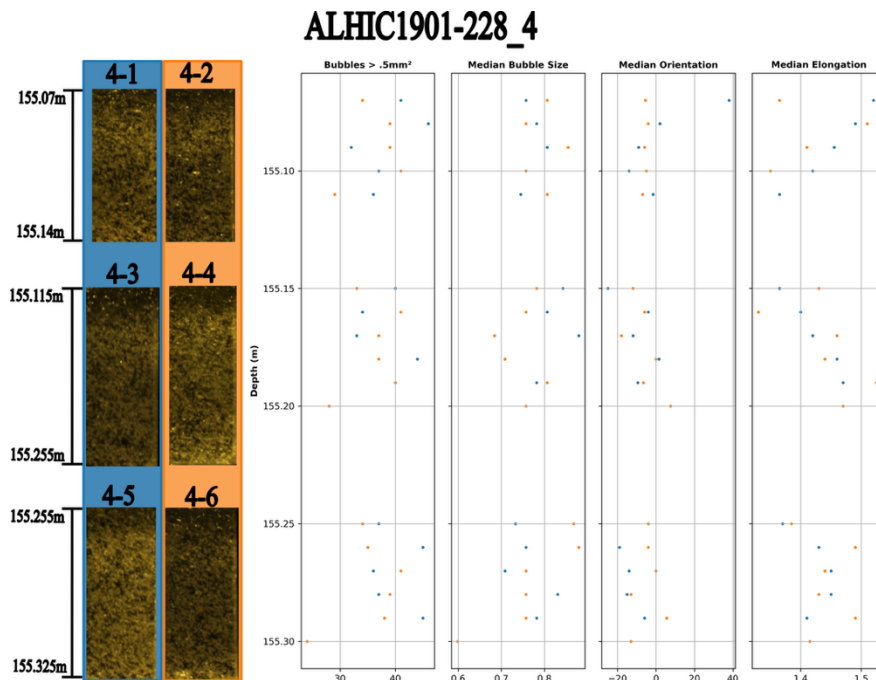


Figure C.1 A visual layout of ALHIC1901-228_4 (~155.07m-155.325m) and the plots for bubble properties. On the left, the 6 small samples are laid out in side-by-side pairs, since this is the way they were cut from the ice core. The plots to the right of the sample images depict the bubble properties by depth, color coded by left and right samples.

Results

The small samples of ALHIC1901-228_4 were analyzed with the bubble detection algorithm. Bubbles as small as 0.5mm^2 and as large as 3.9mm^2 were captured and their properties graphed in Figures C1 and C2. Figure C1 shows the bubble properties graphed by depth, whereas Figure C2 graphs ALHIC1901-228_4-1 and ALHIC1901-228_4-6 side-by-side to account for layer dipping in this core section. The comparison of 4-1 and 4-6 show similar bubble properties, and while it may not be exact, it suggests a correlation between the 2 samples that could serve as further evidence of this core's dipping stratigraphy.

ALHIC1901-228_4:Bubble Properties by Sample Index

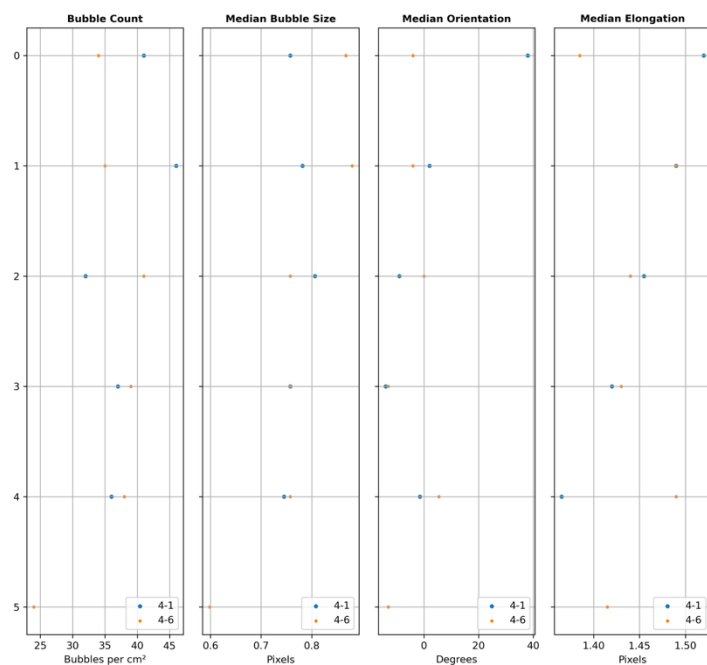


Figure C.2 Graphs showing the plots for bubble properties of ALHIC1901-228_4-1 (blue) and ALHIC1901-228_4-6 (orange) for comparison.

APPENDIX D

DENALI

Project Overview

This project, in collaboration with Emma Erwin and Karl Kreutz, focused on an alpine ice core from Denali, DIC1, between ~150m to ~190m in depth. The goal was to determine whether hyperspectral imaging could detect a change in the ice fabric that was seen on ground penetrating radar around 155m-175m.

Data Availability

- Dataset: DIC1 Hyperspectral Scans will be available on the Arctic Data Center

Methods / Algorithms

- Dust Loading Index (DLI)
- Depth alignment using pixel-to-cm conversion

Results

Working with this core proved to be a considerable challenge, especially the sections within the transition zone, as cracks and fractures were prevalent throughout the ice. This rendered the bubble detection algorithm fairly useless on full section images, as the cracking appears as the brightest part of the core, drowning out any bubbles. The bubble removal and bubble analysis was near impossible on full images, so manual inspection of the core images was required to pick out “clean sections” of ice that could then be analyzed.

Processing the clean sections offered promising results in the BG_NDI vs VN_NDI (Figure D1) and DLI (Figure D2).

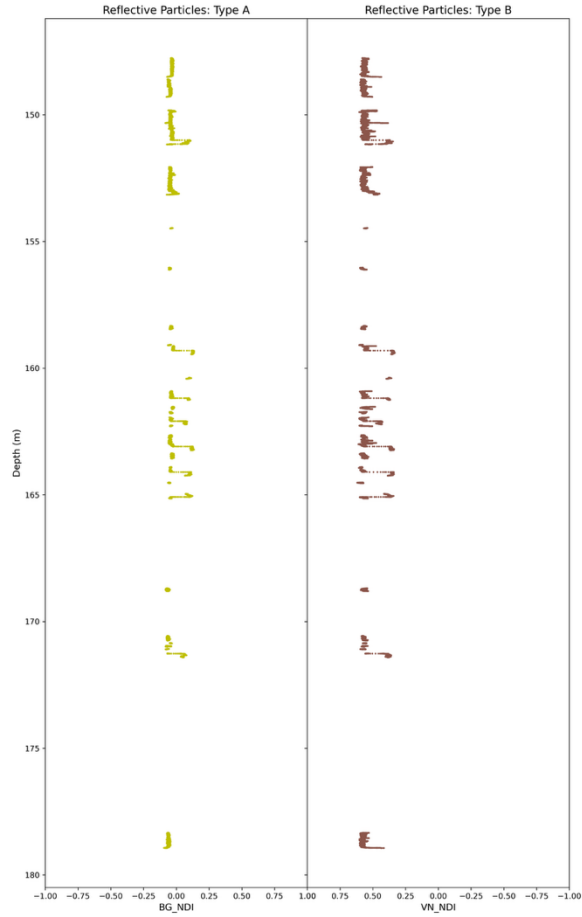


Figure D.1 This plot depicts the Blue-Green (.42 μ m and .52 μ m) Normalized Difference Index in yellow and the Visible-NIR (.52 μ m and .98 μ m) Normalized Difference Index in brown. The x-axis for VN-NDI is reversed for a mirrored comparison.

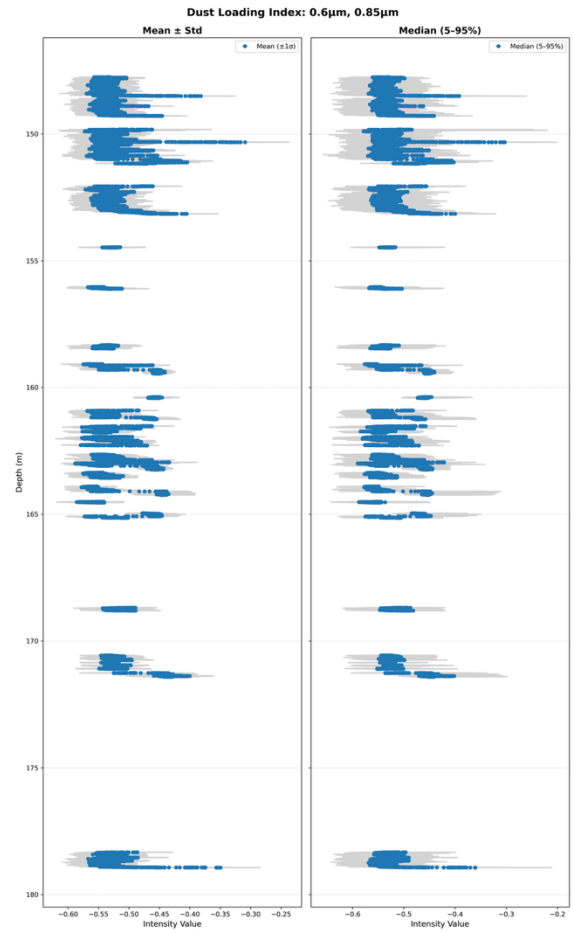


Figure D.2 The mean (left) and median (right) row wise pixel intensities of DIC1 ~150m to ~190m by depth, after bubble removal, inpainting, and DLI application.

Future Work / Data Plans

This work could potentially be integrated into a future publication with collaborators Emma Erwin and Karl Kreutz.

APPENDIX E

ALHIC2416

Project Overview

This project focused on samples of ice from ALHIC2416. The goal was to improve data acquisition methods from ALHIC2302 with the addition of a transparent glass pane to normalize the bottom lights.

Data Availability

- Dataset: ALHIC2416 Hyperspectral Scans will be available on the US Antarctic Program Data Center

Results

Since the method of sample acquisition changed, this dataset has not yet been processed as the light correction algorithm and noise-filtering method will need to be adjusted to account for the new sample settings.

Future Work / Data Plans

Future work includes rewriting the light correction algorithm, the noise-filtering algorithm, and processing the core.

BIOGRAPHY OF THE AUTHOR

Roisin was born in Maine in 2001. After many years out of state, Roisin returned to Maine and graduated from Orono High School. She got a Bachelor's degree in Interdisciplinary Studies, with concentrations in Computer Science and Human-Computer Interaction from the University of Maine in 2023. While working on her undergraduate degree, Roisin worked at the VEMI Lab, a leading lab for research and development of human-technology collaboration. At VEMI, Roisin was a Lead Developer on a prize-winning iOS app for the USDOT, maintained and optimized a fully autonomous vehicle simulator, and aided in research on how future autonomous vehicles can most effectively communicate with passengers. These experiences have helped shape her research focus on implementing creative solutions using technology on some of the world's biggest problems.

Roisin is a candidate for the Master of Science degree in Quaternary and Climate Studies from the University of Maine in May 2026.