Modeling Sea Ice Thickness using Machine Learning and Remote Sensing Modalities

John Baker

Ira A. Fulton Schools of Engineering

School of Computing and Augmented Intelligence

APPROVED:

Dr. Douglas Cochran	______________________,	Director
Dr. Hua Wei	______________________,	Second Reader

ACCEPTED:

____

Dean, Barrett, the Honors College

Arizona State University

Modeling Sea Ice Thickness using Machine Learning and Remote Sensing Modalities

Honors Thesis

Barrett, The Honors College

Author:
John Baker

Committee:

Dr. Douglas Cochran
Dr. Hua Wei

April 2024

Acknowledgement

I hereby declare that I can carry out the present work independently without outside help. and used only the sources and aids indicated. I assure Furthermore, that I have not yet submitted this thesis to any other examination board.

Abstract

Little is known about the state of Arctic sea ice at any given instance in time. The harshness of the Arctic naturally limits the amount of in-situ data that can be collected, resulting in gathered data being limited in both location and time. Remote sensing modalities such as Synthetic Aperture Radar (SAR) imaging and laser altimetry help compensate for the lack of data, but suffer from uncertainty because of its inherent indirectness. Furthermore, precise remote sensing modalities tend to be severely limited in spatial and temporal availability, while broad methods are more accessible at the expense of precision. This thesis focuses on the intersection of these two problems and explores the possibility of corroborating remote sensing methods to create a precise, accessible source of data that can be used to examine sea ice at a local scale.

///////////////////////////

English Abstract here. The abstract should provide a complete but concise description of your work. In brief, you should address the following: motivation, problem statement, approach, results, and conclusions.

Contents

Introduction

A distinct problem arising from attempting to study the exact state of sea ice is that there is limited data available. Data pertaining to sea ice is widely distributed, not correlated, or limited in scope such that it's difficult to draw inferences from. Naturally, the immediate approach to this thesis will first require research into available data sources before conducting any exploration. Exploration of these sources will reveal which sources should be targeted for the remainder of the thesis.

After reviewing the preliminary research, this paper will move forward to a more targeted effort in obtaining and analyzing data to develop a machine learning model. Chapter 2 will describe the process of obtaining the selected data, and Chapter 3 will discuss the methodology of developing the model.

1.1 Preliminary Research

Preliminary research into the availability of ice thickness data yields two notable organizations which may act as sources of data. These two organizations are the National Snow and Ice Data Center (NSIDC), and the National Aeronautics and Space Administration (NASA). Although both organizations offer large amounts of data, they differ in approach. The NSIDC acts as a historic repository, accumulating previous studies into a single location for easy access. NASA offers a similar service, but additionally allows researchers the opportuntity to query information from any of their active orbiting satellites. /// Add the European Space Agency as another portal - as this will be discussed in preliminary exploration. If not, it will surely need to be discussed somewhere, including the process of retrieving it, because it's the data set we're using for the model. ////

NASA

Here is where you should include sources to the ATL10 Data Specification pdf, and the associated CryoSat data specs. It should also be included where these links came from. It also should be included where, if applicable, specific data downloads were from. It should include the process of querying for data, which portals are made accessible, and how flexible they are. Do not forget to mention that they made an IceSat-2 mobile application for iOS.

Include how NASA's Sentinel-2 delivers SAR imaging, but at 50m resolution. This is a modality that can be used for a CNN but the resolution is not meaningfully compatible with any source of ground truth we have. Thus, searching for alternatives led to the European Space Agency and its sponsorship for ICEYE's 1m resolution imaging.

Somewhere you should discuss IceSat-2 and its laser altimetry. ESPECIALLY that it's 17m footprint offers precision at a given location,

NSIDC

Here is where the NSIDC links and data retrieval should be discussed. Limitations of the data or other notes about it's composition should be included according to the details in the abstract saved on file. Preliminary exploration will be where the figures will be uploaded (Or maybe we show one figure showing all of the points, and the exploration will be for the single instance?)

1.2 Preliminary Exploration

Data of interest were the IceSat-2 ATL10 product and the NSIDC In-situ Dataset, selected for their combination of breadth and accuracy. Furthermore, the European Space Agency's requirements for sponsored data need to be explored to pursue the higher resolution SAR imagery hoped for. The in-situ measurements will provide an invaluable source of ground truth while the ATL10 product will drastically expand the flexibility and accessibility of sea ice freeboard data moving forward.

NASA

A single delivery of IceSat-2's ATL10 data product yields a '.h5' file, incompatible with traditional spreadsheets. To access IceSat-2 data in a meaningful way involved developing a script to extract relevant information according to the types enumerated in the data product specification. Columns of interest include the latitude, longitude, time, and calculated freeboard height for each of the three beam pairs. // Consider adding perl and dynamically adding sample rows from the extracted .h5 file. This will give the reader context as to what's being gathered /

NSIDC

The "On-Ice Arctic Sea Ice Thickness Measurements by Auger, Core, and Electromagnetic Induction, from the Late 1800s Onward, Version 2" contains 69,750 rows of data, spanning 5 categorized regions; the Arctic Ocean, the Beaufort Sea, Greenland Coast, Prudhoe Bay, and Russian Coast. Filtering down to the Beaufort Sea region, a region of interest, yields 23 separate studies ranging between 1958 and 2016.

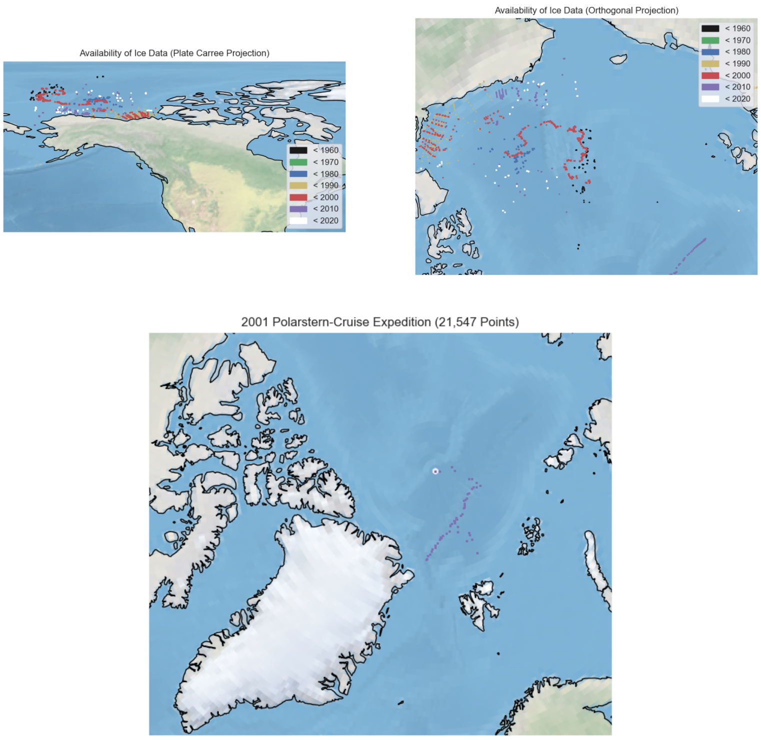

Figure 1.1: In-Situ Data Availability

For a simple analysis, the largest single dataset was chosen to conduct a simple test on the distribution of sea ice thickness. This distribution if normal will reasonably allow for future conducting of statistical tests, and if not normal will give insight as to what a reasonable assumption of ice thickness distribution should be. The largest dataset consists of 21,547 distinct points sourced from direct auger measurements and indirect electromagnetically sensed methods. Given that the majority of the data comes from the remote sensed method, it was validated by plotting the sensed values against the auger values where applicable and seeing how correlated they were. The results of a linear regression test yielded a relationship of 0.92x + 0.18, with an adjusted $R\hat{2}$ value of 0.866 and a P Value of 2.31 * $10\hat{}202$. The adjusted $R\hat{2}$ value and low P value demonstrate there is a high correlation between these variables, and our model captures a significant portion of the variance. The 0 line is plotted as a way to visualize how the residuals are symmetrically clustered around 0.

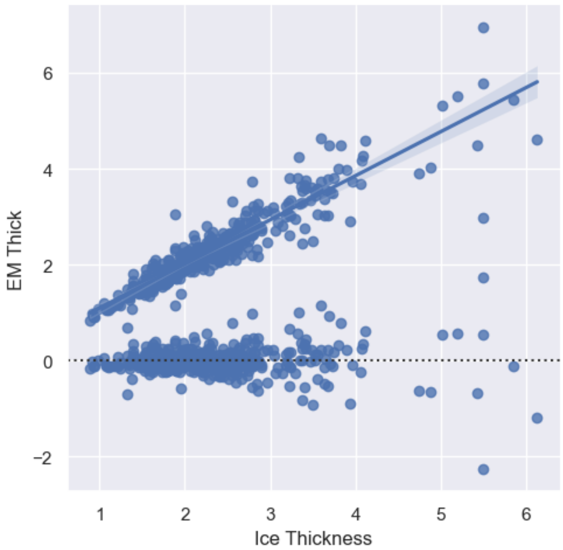

Figure 1.2: Electromagnetic vs Auger Sourced Relationship

After accepting the validity of the electromagnetically sensed data, the points were plotted to visualize the strip of ice that was measured. Given that these data points were taken in a single linear track, the line graph shows a cross-sectional profile of ice in that axis. The distribution reveals a slight tail on the left side of the curve, and a hypothesis test confirms that the distribution is not normal. Moving forward, it is reasonable to believe that ice distribution is not normal, and over a large enough sample there is expected to be a slight tail on either end.

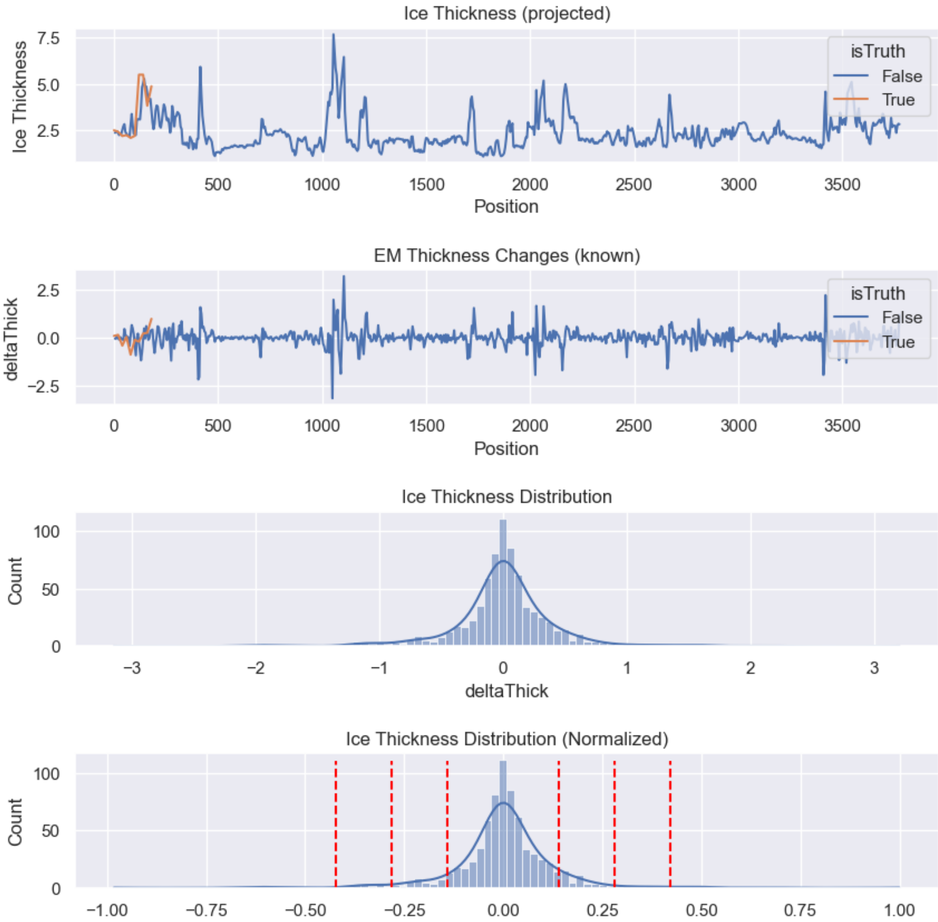

Figure 1.3: In-Situ Ice Thickness Distribution

Data Collection

In this section, include research relevant to the gathering of data. Include the different metrics that may be captured and the different options presented from different avenues. Much research was done earlier based on the opportunities and limitations of NASA satellites. Include documentation you captured about CryoSat if pertinent to this section.

It may be necessary to include information on SAR imaging in the introduction, so even the non-expert reader can get a functional understanding of the topic.

It may be possible to include figures and other data from the NSIDC In-Situ data repository and discuss the viability of that data as a 'ground-truth' and demonstrate that according to those studies, sea ice was non-normal in distribution. Acknowledge that the data from those studies spanned many miles, and the distribution may sheerly be by variance of ice thickness across the collection area.

Here is where we'll discuss the selection of ICEYE's Data for it's high resolution imaging, and IceSat-2 for it's high-accuracy laser altimetry. It may be possible to include figures to demonstrate how these satellites work, but it may not be pertinent to the thesis (although it would help with understanding).

2.1 Laser Altimetry (IceSat-2)

2.2 SAR Imaging (ICEYE)

——Default Text ———-

Provide in brief the background information for your work/field keeping in mind that maybe your readers do not have experience with topics your reference or address in your thesis.

In the second part, provide a review of the state of the art relevant to your thesis. Here you present relevant research that relates to your work.

Experimentation

Here you should present how you aimed to answering your question or solving the problem you identified in the introduction. You should present the methods and the instrument you use, the structure of your study or workplan and how you achieved the desirable outcome. The order is indicative, please feel free to rearrange as you see fit.

3.1 Methods

3.2 Study setup

3.3 Data collection

3.4 Data analysis

Results

Here you present the results of your study (if you carried out one), and any data analysis you may have performed to answer your question. You should consider splitting the results per research question, as these are presented in the introduction

Discussion

In the Discussion section you should elaborate on the following points:

5.1 Research Questions

Here you will answer your research questions, as they appear in the introduction. Answer each question in a different section. Relate your answer to your results. Discuss if your findings support and align with related work or not. Explain why do you think this happens, especially if your findings contradict existing work. Discuss alternative interpretations of your findings.

5.2 Theoretical and Practical Implications

Here you should explain what is your contribution and how it promotes knowledge in the field both in terms of theory and practice. How do you envision your work to change or promote research in the area. How could it be used? How do you envision it to be used?

Conclusion

In the Conclusions section you should elaborate on the following points:
1. Conclusions: summarize your main objective/question/problem and how you solved it

6.1 Limitations

2. Limitations: Present the limitations or shortcomings of your work

6.2 Future Work

3. Future Research: Propose related topics that should be addressed by research in the future or ideas that you had while conducting your thesis.

List of Figures

List of Tables

Insert Appendix Name

A.1 Insert subtitle