GEOPHYSICAL FEATURE EXTRACTION AND SPATIOTEMPORAL ANALYSIS OF POLAR SEA ICE USING HIGH SPATIAL RESOLUTION IMAGERY

by

Dexuan Sha
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Earth Systems and Geoinformation Science

Committee:

___________________________________	Dr. Chaowei Yang, Committee Chair
___________________________________	Dr. Xin Miao, Committee Member
___________________________________	Dr. Ruixin Yang, Committee Member
___________________________________	Dr. Liang Zhao, Committee Member
___________________________________	Dr. Dieter Pfoser, Department Chairperson
___________________________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
___________________________________	Dr. Fernando R. Miralles-Wilhelm, Dean, College of Science
Date: ______________________________	Summer Semester 2021 George Mason University Fairfax, VA

Geophysical Feature Extraction and Spatiotemporal Analysis of Polar Sea Ice Using High Spatial Resolution Imagery

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Dexuan Sha
Master of Sciences
Missouri State University, 2016
Bachelor of Sciences
Hainan University, 2014

Director: Chaowei Yang, Professor
Department of Geography and Geoinformation Science

Summer Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This dissertation is dedicated to my beloved parents - Geli Han, Qingchao Sha, wife - Rencong Wang and daughter - Stephanie Sha. All I have and will accomplish are only possible due to their love and sacrifices.

ACKNOWLEDGEMENTS

I would like to thank all people who have helped and inspired me during my doctorate study. I especially want to thank my advisor, Dr. Chaowei Yang, and committee members, Drs. Xin Miao, Ruixin Yang, and Liang Zhao, for their guidance during my research and study at George Mason University. My special words of thanks should also go to my colleagues at the NSF Spatiotemporal Innovation Center. My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation is simply impossible without them.

This work was supported by the NSF EarthCube and CSSI Programs: [Grant Numbers 1835507, 192027521, 1841520], NSF COVID rapid response [Grant Numbers 2027521] and NSF Innovation Center Membership Projects funded by NASA and Amazon GMU-19-03, GMU-18-02 and STC-15-01.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Advanced Microwave Scanning Radiometer .... AMSR
Advanced Very High-Resolution Radiometer .... AVHRR
Airborne Topographic Mapper .... ATM
Application programming interface .... API
Arctic Cyberinfrastructure .... ArcCI
Arctic High Spatial Resolution .... ArcHSR
Chinese Arctic Research Exploration .... CHINARE
Cyberinfrastructure CI
Comma Separated Values .... CSV
Comprehensive Knowledge Archive Network .... CKAN
Continuous Airborne Mapping by Optical Translator .... CAMBOT
Deep Learning .... DL
Digital Mapping System .... DMS
European Centre through the Medium-Range Weather Forecasts .... ECMWF
European Organization for the Exploitation of Meteorological Satellites .... EUMETSAT
Extract-Transform-Load .... ETL
File Transfer Protocol .... FTP
First-Year Ice .... FYI
Geographical Information System .... GIS
Geographical JavaScript Object Notation .... GeoJSON
Geospatial Data Abstraction Library .... GDAL
Global Positioning System .... GPS
Graphic User Interface .... GUI
Hadoop Distributed File System .... HDFS
Healy-Oden Trans-Arctic Expedition .... HOTRAX
High Performance Computing .... HPC
High Spatial Resolution .... HSR
HyperText Markup Language .... HTML
International Arctic Buoy Program .... IABP
Joint Photographic Experts Group .... JPEG
Sea Ice Lead Detection Algorithm using Minimal Signal .... SILDAMS
The European Centre for Medium-Range Weather Forecasts .... ECMWF
Moderate Resolution Imaging Spectroradiometer .... MODIS
Multiple Altimeter Beam Experimental Lidar .... MABEL
Multiple Linear Regression .... MLR
Multi-Year Ice .... MYI
NASA Team 2 .... NT2
National Aeronautics and Space Administration .... NASA
National Center for Atmospheric Research .... NCAR
National Center for Environmental Prediction .... NCEP

National Snow and Ice Data Center .... NSIDC
Object-Based Image Analysis .... OBIA
Ocean and Sea Ice Satellite Application Facility .... OSI SAF
Open Geospatial Consortium .... OGC
Open Source Sea-ice Processing .... OSSP
Operation IceBridge .... OIB
Out-of-bag .... OOB
Point of Interest .... POI
Portable Document Format .... PDF
Portable Network Graphics .... PNG
Region Adjacency Graphs .... RAG
Relational Database Service .... RDS
Resilient Distributed Dataset .... RDD
Root Mean Square Difference .... RMSD
Scanning Multichannel Microwave Radiometer .... SMMR
Sea Ice Concentration .... SIC
Simple Cloud Storage S3
Special Sensor Microwave Imager/Sounder .... SSMI/S
Special Sensor Microwave/Imager .... SSMI
Surface Heat Budget of the Arctic Ocean .... SHEBA
Synthetic-Aperture Radar .... SAR
Tagged Image File Format .... TIFF
Thin Ice Concentration .... TIC
Unified Modeling Language .... UML
United States Geological Survey .... USGS
Unmanned Aerial Vehicle .... UAV
Web Feature Service .... WFS
Web Graphics Library .... WebGL
Web Map Service .... WMS
Web Map Tile Service .... WMTS

ABSTRACT

GEOPHYSICAL FEATURE EXTRACTION AND SPATIOTEMPORAL ANALYSIS OF POLAR SEA ICE USING HIGH SPATIAL RESOLUTION IMAGERY

Dexuan Sha, Ph.D.

George Mason University, 2021

Dissertation Director: Dr. Chaowei Yang

The Arctic sea ice region has become an increasingly important study area since it is not only a key driver of the Earth's climate, but also a sensitive indicator of climate change. To model and validate sea ice changes, it is crucial to extract geophysical features of sea ice from high-resolution remote sensing data. We collected a large volume of remote sensing images from multiple platforms such as airborne Digital Mapping System (DMS) and Worldview series satellite in the Arctic region during melting season. Processing such a large volume of imagery poses a significant challenge for extracting sea ice spatiotemporal patterns in a timely manner. Additionally, high spatial resolution (HSR) has been largely ignored due to its complex and heterogeneous nature in both space and time, and Arctic operational missions can routinely produce hundreds of gigabytes of data. The advancement of drone technologies keeps adding rapidly to the volume of sea ice aerial-survey-based observations. In summary, processing such big sea ice data

includes challenges such as: 1) the big data challenges in HSR image product, e.g., the big data volume and the heterogeneous formats of a variety of sea ice HSR image data collected by different platforms or agencies; 2) the lack of standard sea ice feature extraction procedure from HSR imagery; 3) the ability for managing, visualizing, and processing HSR sea ice image data, and extracting geophysical properties or attributes. I propose a reliable and effective high-accuracy and high-performance approach to extract sea ice geophysical features from a large amount of HSR remote sensing data to support scientists and allow them to gain new insights from the spatiotemporal analysis on big data process.

The objectives of this research are to 1) develop an efficient geophysical feature extraction workflow based on object-based image analysis (OBIA) method for HSR image data to classify different sea ice features and extract the relevant geophysical parameters such as sea ice leads, sea ice floe, melt pond and ice ridge; 2) design a practice workflow to analyze spatiotemporal patterns of sea ice geophysical features; and 3) design and develop a prototype of an on-demand web service for the cyberinfrastructure, providing a publicly available portal for various data owners and users.

In order to achieve these objectives, an on-demand sea ice HSR imagery management and processing service is developed, and a scientific case study is demonstrated for geophysical feature extraction and spatiotemporal analysis of sea ice leads.

This research on geophysical feature extraction and spatiotemporal analysis of sea ice from high spatial resolution data is innovative for: 1) the practical OBIA classification workflow in a distributed environment for large datasets; 2) the extracted geophysical features could serve as ground references in sea ice research; 3) the developed arctic cyberinfrastructure provides a data service prototype for polar community. The results of this research can be helpful for the understanding of sea ice processing and utilization of climate modeling and verification at different scales.

CHAPTER 1. INTRODUCTION AND OBJECTIVES

1.1 Background

The Arctic region has become increasingly important to study as it: 1) provides significant natural resources, 2) is sensitive to human activities and global environment change, and 3) is a key driver of the Earth's climate. The public and media are also concerned about the wellbeing of the arctic ecosystem and polar wildlife, oil and gas resources exploitation, and ice hazards in ship navigation. Sea ice of Arctic regions acts as both an indicator and an amplifier of climate change, since it is very sensitive to small temperature changes. The past 13 years (2007–2019) have marked the lowest Arctic summer sea ice extents in the modern era, with a record summer minimum (3.57 million km^2) set in 2012, followed by 2019 (4.15 million km^2), and 2007 (4.27 million km^2) (Parkinson, 2019). Some climate models predict that the shrinking summer sea ice extent could lead to the Arctic being free of summer ice within the next 20 to 30 years (Marshall, 2013; Peng et al., 2018). If the trend continues, serious consequences will appear including higher sea level, higher water temperatures, more powerful and frequent storms (Parkinson & Comiso, 2013), diminished habitats for polar animals, increased above-ground biomass (Jeffries et al., 2013), more pollution due to fossil fuel exploitation, and increased ship traffic.

Arctic sea ice modeling and observation are the two basic research approaches to understand the dynamic pattern of sea ice as well as relevant factors. A lot of efforts have been put into Arctic sea ice modeling to simulate sea ice processes. However, these sea ice models were initiated and developed based on limited field surveys and aircraft or satellite image data. From a geographical information system (GIS) and remote sensing perspective, observation approaches provide significant material for data-driven research by helping detect sea ice geophysical parameters and calibrate/validate climate models (National Research Council, 2007). Traditional focused low-resolution observations are advanced in spatial coverage and high temporal frequency which support a long-term polar observation of sea ice extent and concentration. However, low resolution imagery is limited on spatial resolution for the detection of small-scale objects. These object include melt ponds (diameter of 5-20 meters), pools of open fresh water on the surface of Arctic sea ice in the summer which play an important role in Earth's radiation balance due to how they strongly absorb solar radiation rather than reflecting it as snow and ice does.

With the development of sensors and big data techniques supporting large amounts of data processing, many available sea ice aerial photos including Operation IceBridge (OIB) DMS, Chinese Arctic Exploration (CHINARE 2008, 2010, 2012), Surface Heat Budget of the Arctic Ocean (SHEBA) 1998 and Healy-Oden Trans-Arctic Expedition (HOTRAX) 2005, and the newly released declassified intelligence satellite images have become a major source of arctic sea ice research. High spatial resolution (HSR) remote sensing imagery is valuable in Arctic sea ice research since it can be used to verify Earth observation satellite data, extract sea ice geophysical parameters, and calibrate/validate climate models (National Research Council, 2007). Different from low-resolution imagery, HSR images have been largely ignored due to their complex and heterogenous nature in both space and time. It is difficult to weave these small pieces of information into a coherent large-scale picture, which is important for coupled sea ice and climate modeling and verification. This dissertation introduces an effort to develop a reliable and efficient on-demand image batch processing web service CI module and its associated data sets. The developed a data platform Arctic Cyberinfrastructure (ArcCI) is capable of extracting accurate spatial information on open water, bare ice, submerged ice, thin ice, and ridge shadows from a large volume of HSR image data set with limited human intervention.

The spatiotemporal analysis in this research reveals how the extracted geophysical features change as well as the trend of arctic sea ice in the past decade at a ground reference scale. It will also uncover the interaction and correlation between sea ice features and climate factors. A practical object-based image classification workflow is created for a large collection of aerial survey imagery to extract the sea ice lead distribution of the Arctic Ocean from 2012 to 2018. In this scientific case study, the spatiotemporal variations of leads along the Laxon Line are verified by ATM surface height data and correlated with sea ice motion, atmosphere temperature, and wind data.

1.2 Objectives and Contribution

1.2.1 Objectives

This dissertation research is motivated by the challenge of 1) the lack of standard sea ice feature extraction procedures from HSR imagery; 2) the lack of an easy-to-use functionality for exploring, visualizing, and processing a large volume of arctic HSR sea ice image data alongside the extraction of geophysical properties and attributes; and 3) the big data challenge of the HSR sea ice imagery in volume, velocity and variety, for example, the heterogeneous formats of a variety of sea ice HSR image data collected by different platforms and agencies. I delivered a reliable and effective approach for the high-accuracy and high-performance extraction of sea ice geophysical features from a large amount of HSR remote sensing data to support scientists who seek to gain new insights for arctic sea ice change.

The objectives of this research are to 1) develop an efficient geophysical feature extraction system based on the object-based image analysis (OBIA) method for HSR image data to classify different sea ice features such as sea ice floes, sea ice leads, melt pond and ice ridges; 2) design and develop the prototype of an on-demand web service based on the cyberinfrastructure, providing a HSR image analysis and management service for various data owners and users; 3) demonstrate the capabilities of the online system in a case study to analyze the spatiotemporal patterns of sea ice leads from the HSR DMS dataset collected over 7 continuous years over the Arctic Ocean.

1.2.2 Contributions

The main contributions of this dissertation are as follows:

1) Reviews HSR sea ice observation imagery from the past 20 years as well as the available geophysical parameters from the imagery, which provides valuable metadata collection on ground-reference level observations in polar research,
2) Develops a cloud-based web system for managing and analyzing HSR imagery, which allows for high performance batch image processing,
3) Develops and improves a practical object-based image classification workflow capable of extracting geophysical parameters from HSR imagery,
4) Analyzes the correlation between sea ice leads and relative climate and environmental factors from a spatiotemporal perspective.

1.3 Dissertation Overview

The rest of this dissertation is organized as follows: Chapter 2 reviews the literature of related research in terms of high spatial resolution observation in the polar region, geophysical parameter extraction of sea ice, and online data services and framework. Chapter 3 introduces a cloud-based cyberinfrastructure and an on-demand service for managing and processing HSR imagery. Chapter 4 focuses on the spatiotemporal analysis of extracted sea ice features from a time-series of a large data collection in the Arctic Ocean. Chapter 5 concludes the dissertation and proposes potential future works.

CHAPTER 2. LITERATURE REVIEW

2.1 High Spatial Resolution Imagery in Polar Study

There is a long history in arctic sea ice observation activities (Walter N. Meier, 2016). Starting the early 1900s, sea ice has been manually observed by humans at an observation center on the Arctic Ocean national coastline (Watanabe et al., 2006). In the 1970s, satellites were equipped with various microwave and radar sensors for observation from outer space (Cavalieri et al., 2003). At the time, commercial and scientific exploration ships above water and undersea submersibles were widely utilized for in-situ data sensing for sea ice feature extraction (McLaren et al., 1992; Y. Yu et al., 2004).

Within the last 30-40 years, scientists have used satellite remote sensing technology in most observation applications of sea ice density, sea ice extent, thickness, and water density across the Arctic region (Kurtz & Markus, 2012; R Kwok, 2010; Sandven et al., 2006; Teleti & Luis, 2013). The advantages of satellite remote sensing are as follows: 1) long time range of data collection with high temporal resolution; 2) coverage across large areas. General polar satellites can obtain information of the entire polar region and obtain data on the coverage of ice over a wide range; 3) a wide range of applications. For example, the sea ice extent from the 1900s to the present can be estimated through the fusion of historical human eye-observed data, and the ice change trend can be seen in a longer term (Teleti & Luis, 2013).

However, satellite remote sensing is not a panacea. The main limitation of satellite data is its low spatial resolution. For example, the spatial scale of a microwave remote sensing product is 25km and that of a radar product is 5-10 kms (SAR) (Howell et al., 2019; Thomas et al., 2008). Detailed sea ice information such as status, size, fraction, distribution, and melt ponds cannot be directly identified by satellite imagery (Markus et al., 2003; Rösel & Kaleschke, 2011). Features of sea ice including thickness and concentration could be induced with high uncertainty through applying the blurry inversion methodology and data capture perspectives. For example, there are limitations to extracting above water height information from orthographic images as opposed to using oblique Unmanned Aerial Vehicle (UAV) data instead (Kraaijenbrink et al., 2016; M. Wang et al., 2018).

High spatial resolution (HSR) imagery is continuously collected by observers and researchers over many years. Advantages of the high resolution HSR imagery has allowed it to be widely utilized for the validation and assessment of remote sensing results. With the development of new equipment as well as an increase in research activities focusing on climate change in Arctic regions, HSR imagery is collected (Figure 1) from industry or research ships, helicopters, submarines, and various types of UAVs. More applications can be enabled considering the large amount of available congregated HSR imagery data.

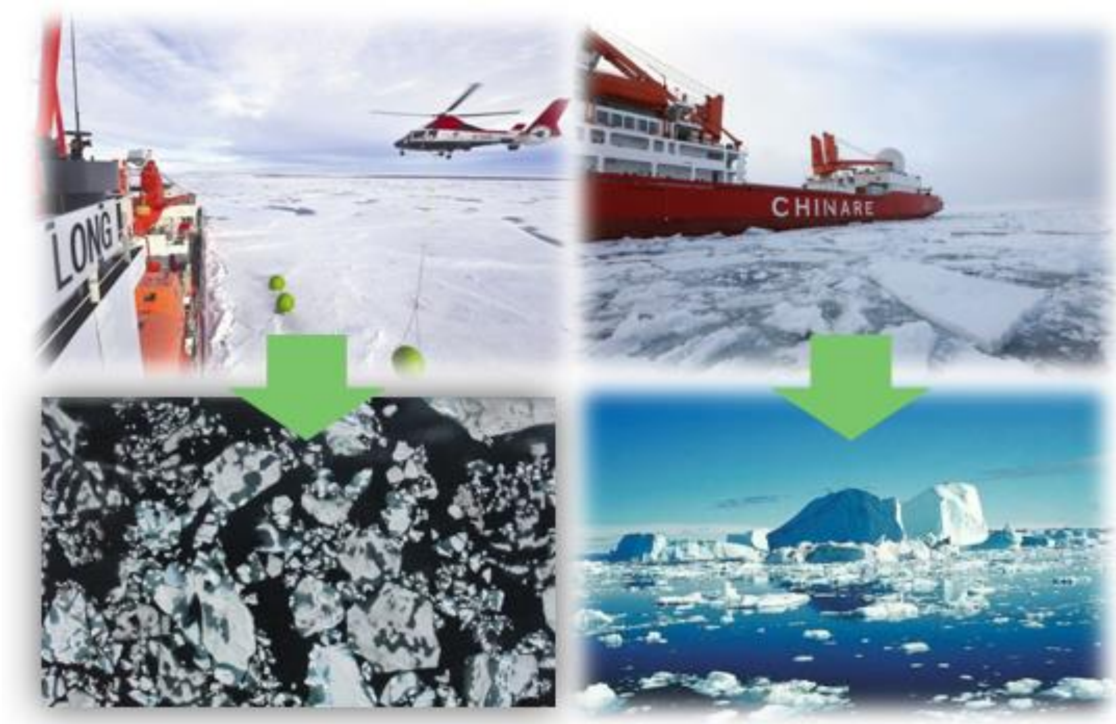


Figure 1. In-situ HSR Imagery Collection Scene from helicopter and ship.

2.1.1 Public Dataset

Public datasets consist of publicly available data usually collected by federal agencies, scientific communities, or non-governmental organizations. They are accessible to all visitors/users and can be discovered through site-wide data catalog in an open science gateway. Public data has three characteristics: (1) the datasets are usually collected by large, funded projects or missions, (2) the data volume is usually at TB (Terabyte) level, and (3) they are usually well-designed and managed and operated in a web server by professional data management teams.

Three public datasets were used to train and build this CI module. First, recently released declassified intelligence satellite images were used. These images are one of the historical high spatial resolution image data sources for arctic sea ice research. In 1995, a group of governmental and academic scientists started to review and advise acquisitions of imagery obtained by classified intelligence satellites as well as recommend the declassification of certain datasets for the benefit of science (Ronald Kwok & Untersteiner, 2011). As a result, numerous declassified HSR arctic sea ice images have become publicly available through the USGS Global Fiducials Library (GFL). The library includes two types of panchromatic images: (1) Literal Image Derived Products (LIDPs) acquired since 1999 at six fiducial sites in the Arctic Basin (Beaufort Sea, Canadian Arctic, Fram Strait, East Siberian Sea, Chukchi Sea, and Point Barrow), with spatial-resolution of 1 m. (2) Repeated imaging of numerous ice floes tracked by data buoys since summer 2009, with a spatial resolution of 1.3 m (Figure 2). The data shows unprecedented value in helping track sea ice/ melt pond evolutions as well as for estimating sea ice ridge heights, ice concentration, floe size, and lateral melting (Ronald Kwok & Untersteiner, 2011).

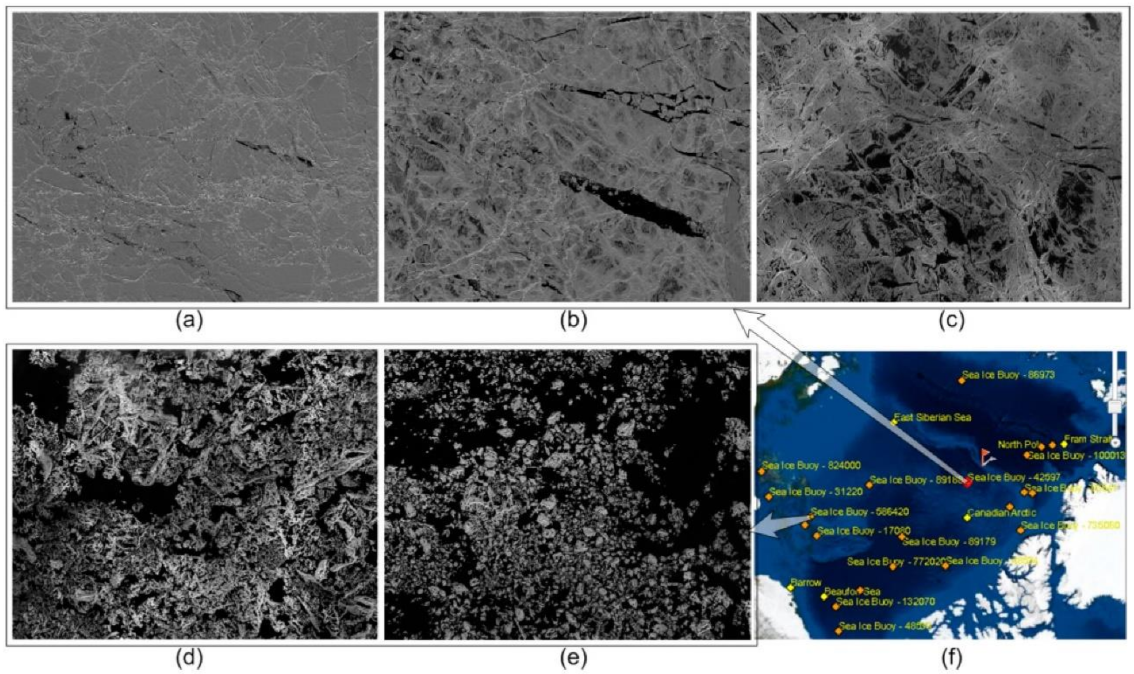


Figure 2. Examples of Global Fiducials Library (GFL) sea ice and melt-pond evolution: images of Buoy 42597 taken on June 6 (a), June 24 (b), and July 1 (c) of 2010, and images of Buoy 586420 taken on August 30 (d) and September 1 (e) of 2010, with the geographic positions of the two buoys shown in (f).

Second, Polar Geospatial Center (PGC) provides National Science Foundation (NSF) funded projects with high-resolution imagery from DigitalGlobe, including WorldView series satellite. WorldView-1, -2, and -3 satellites were launched in 2007, 2009, and 2014, respectively. The most recent WorldView-3 satellite provides one panchromatic image band with a spatial resolution of 0.31 m, and eight multispectral bands with a spatial resolution of 1.24 m. It has become a major source of polar sea ice research.

Finally, Operation IceBridge Digital Mapping System (DMS) is a large collection of digital color aerial photos for polar regions sponsored by the National Aeronautics and

Space Administration (NASA) (Dominguez, 2010). The spatial resolution of DMS ranges from 0.015 to 2.5 m depending on flight altitude and digital elevation model used. DMS data has been used by the sea ice community across a broad range of applications including the detection of leads of open water in sea ice, melt ponds, and other sea ice features. Table 1 shows the characteristics of different image types and spatial resolutions of the three datasets as well as their applications. The publicly available datasets are well-processed and of good quality as they are based on remote sensing data formats and normal geospatial database formats as overseen by professional data labs.

Table 1. Public high spatial resolution (HSR) images.

Dataset (Provider)	Image Type	Spatial Resolution	Applications
Literal Image Derived Products (USGS Global Fiducials Library)	Panchromatic satellite images	1.3 m	Tracking sea ice/melt pond evolutions and estimating sea ice ridge heights, ice concentration, floe size, and lateral melting.
Operation IceBridge DMS (NSIDC)	Multispectral (RGB) aerial photo	0.1 m (0.015 to 2.5 m)	Leads detection of open water in sea ice, melt ponds, and other sea ice features.
WorldView-3 (Polar Geospatial Center)	Panchromatic and multispectral (8 bands) satellite images	0.31 m for Panchromatic, 1.24 m for multispectral	A major source of polar sea ice research with wide spatial coverage.

2.1.2 Longtail Dataset

Longtail datasets are usually collected and managed by independent scientists, research firms, or longtail companies. They can only be accessed by the dataset owner and users with the appropriate sharing permissions. In an operational manner, the longtail or individually captured dataset with available licenses could be archived and located by their metadata in the ArcCI open science gateway. Researchers are also able to contact the owner of the data for access. The ArcCI online service provides data storage and sharing services if the data owner authorizes the platform with a standard open data license. Most longtail datasets have smaller size and/or volume, and are not well documented or published in any data center; they are only mentioned in regional analysis publications. Three different types of longtail HSR sea ice images are used for building the CI module. The first type is aerial photos collected during ship-based expeditions to the Arctic sea ice zone including those from SHEBA (Surface Heat Budget of the Arctic Ocean) 1998 (Perovich, 2003), HOTRAX (Healy-Oden Trans-Arctic Expedition) 2005 (Perovich et al., 2009), and CHINARE (China's Antarctic Research Expedition) 2008, 2010, 2012 (Ruibo Lei et al., 2012; RuiBo Lei et al., 2016; Lu et al., 2010b; Xie et al., 2013). The second type of longtail HSR imagery is time lapse images. An example would be the time lapse images (one taken every 30 min) taken by a fixed camera in Cape Joseph Henry and collected by Christian Haas (Table 2). The images cover two melt onsets, May–July 2011 and May–July 2012, as well as one sea ice onset from August–November 2011.

Table 2. Longtail HSR images collection.

Data	Size	Description
Declassified GFL data	450 GB	The six fiducial sites and repeated images tracking data buoys/floes.
SHEBA 1998 (Perovich)	16.5 GB	Beaufort Sea, 13 flights between May 17, 1998 and October 4, 1998. Also includes a few National Technical Means high resolution satellite photographs.
HOTRAX 2005 (Perovich)	31.3 GB	TransArctic cruise from Alaska to Norway, 10 flights from August 14, 2005 to September 26, 2005.
CHINARE 2008 (Xie)	20.0 GB	Pacific Arctic sector (between 140 °W and 180 °W up to 86 °N), August 17 to Sept 5, 2008.
CHINARE 2010 (Xie)	23.7 GB	Pacific Arctic sector (between 150 °W and 180 °W up to 88.5 °N), July 21 to August 28, 2010
CHINARE 2012 (Xie)	21.2 GB	Transpolar section, (Iceland to Bering Strait), August-September 2012
The time lapse camera (Haas)	40.5 GB	Cape Joseph Henry (82.8N, -63.6W), May 2011 to July 2012.
EM-bird thickness and aerial photos (Haas)	21.2 GB	April 2009, 2011, and 2012, between 82.5 N and 86N, and -60W and -70W.

Longtail HSR imagery is our initial motivation of developing ArcCI. The in-house HSR images are summarized in Table 2. Many other Arctic HSR images are held by different agencies and research teams and will be collected and processed during the operation period.

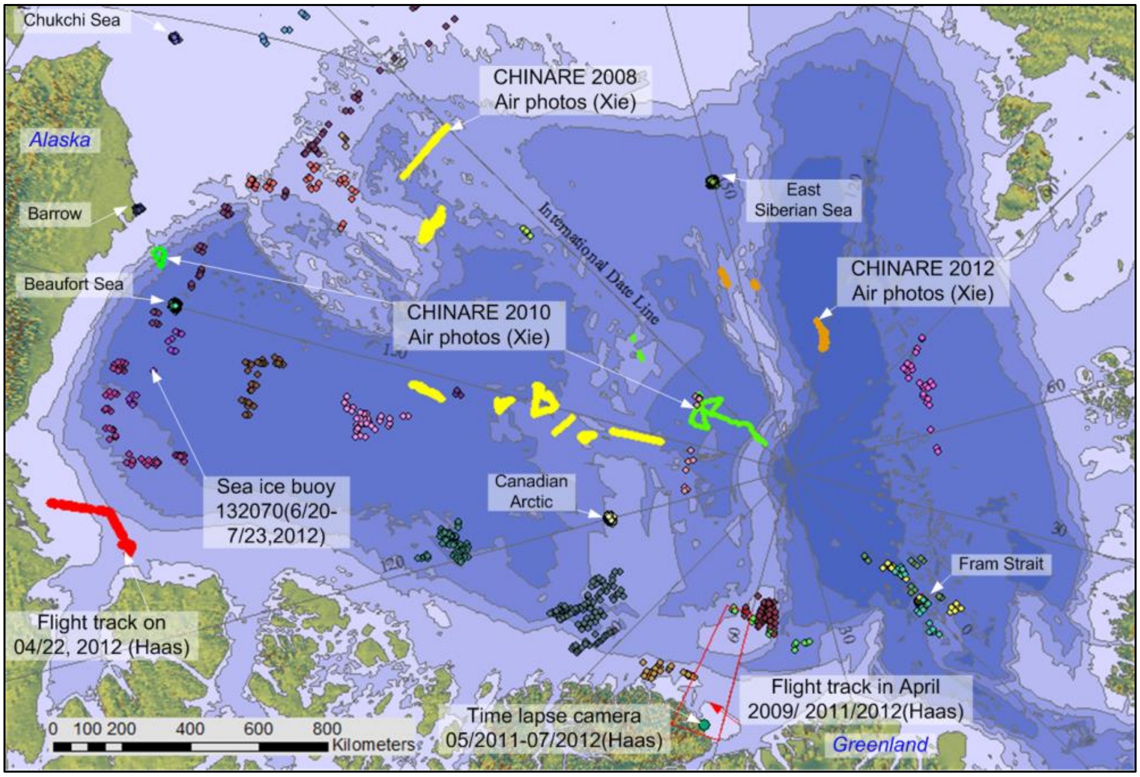


Figure 3. Spatial distribution of declassified images and other survey data.

Generally speaking, there is a large amount of HSR imagery data. Figure 3 shows the spatial distribution of the in-house data through the Arctic region. Four basic characteristics could be identified for HSR imagery: multi-source, heterogenous, discreteness in space, and irregularities in time.

Multi-source and Heterogenous. For the datasets listed above, various transportation carriers with different camera equipment provide different spatial scales and perspectives for HSR imagery collection. Table 1 and Table 2 show the research projects led and contracted by different countries, organizations, scientists, and groups.

These human and project factors allow for specific metadata information to be extracted from HSR imagery files.

Discrete in Space and Irregular in Time. Unlike other moderate or low-resolution satellite images such as MODIS and AVHRR (Scambos et al., 2006; Scharfen et al., 1997), HSR images are both discrete in space and irregular in time: images usually cover only a small area without any overlap, and time intervals vary between a few seconds and several months. Therefore, it is difficult to weave these small pieces of information into a coherent large-scale picture, which is important for climate modeling and verification using coupled sea ice. The time-series images of drifting ice floes do not have fixed geographical positions. Therefore, it is difficult to use conventional geo-registration and change detection approaches to study a specific Arctic region or match results to other mosaicked geophysical data.

Given these unique challenges of HSR images, I am motivated to develop a cyberinfrastructure -based system to collect, manage, process, and share these large amounts of small, scattered pieces of information to understand the sea ice dynamics in the Arctic.

2.2 Geophysical Parameter Extraction of Sea Ice Imagery

Regardless of the image source, all image features and classification results can be used to derive sea ice geophysical properties. Sea ice surfaces undergo substantial changes throughout the season due to the rapid loss of sea ice, increase in duration of sea ice melt, and the decrease in sea ice thickness in recent years (Lu et al., 2010a). Ice surface condition observations are required to understand the underlying causes of the seasonal and spatial evolution of albedo, and identifying key aspects of sea ice surfaces (melt pond coverage, degree of deformation, floe size, and lead distributions) require the evaluation of the surface at meter to decimeter resolution. Figure 4 shows the available sea ice features of the Arctic Ocean in the spring and summer seasons from the HSR imagery. Sea ice leads, pressure ridges, and sea ice floes are abundant in the spring season, and melt ponds begin to appear as the temperature increases in the summer season. In this related work review, sea ice properties from HSR imagery are described and discussed in the first part including sea ice type, sea ice concentration, sea ice thickness, melt pond, surface roughness, and ridge height. The evolution of sea ice extraction approaches, specifically sea ice surface classification, is discussed in the second part.

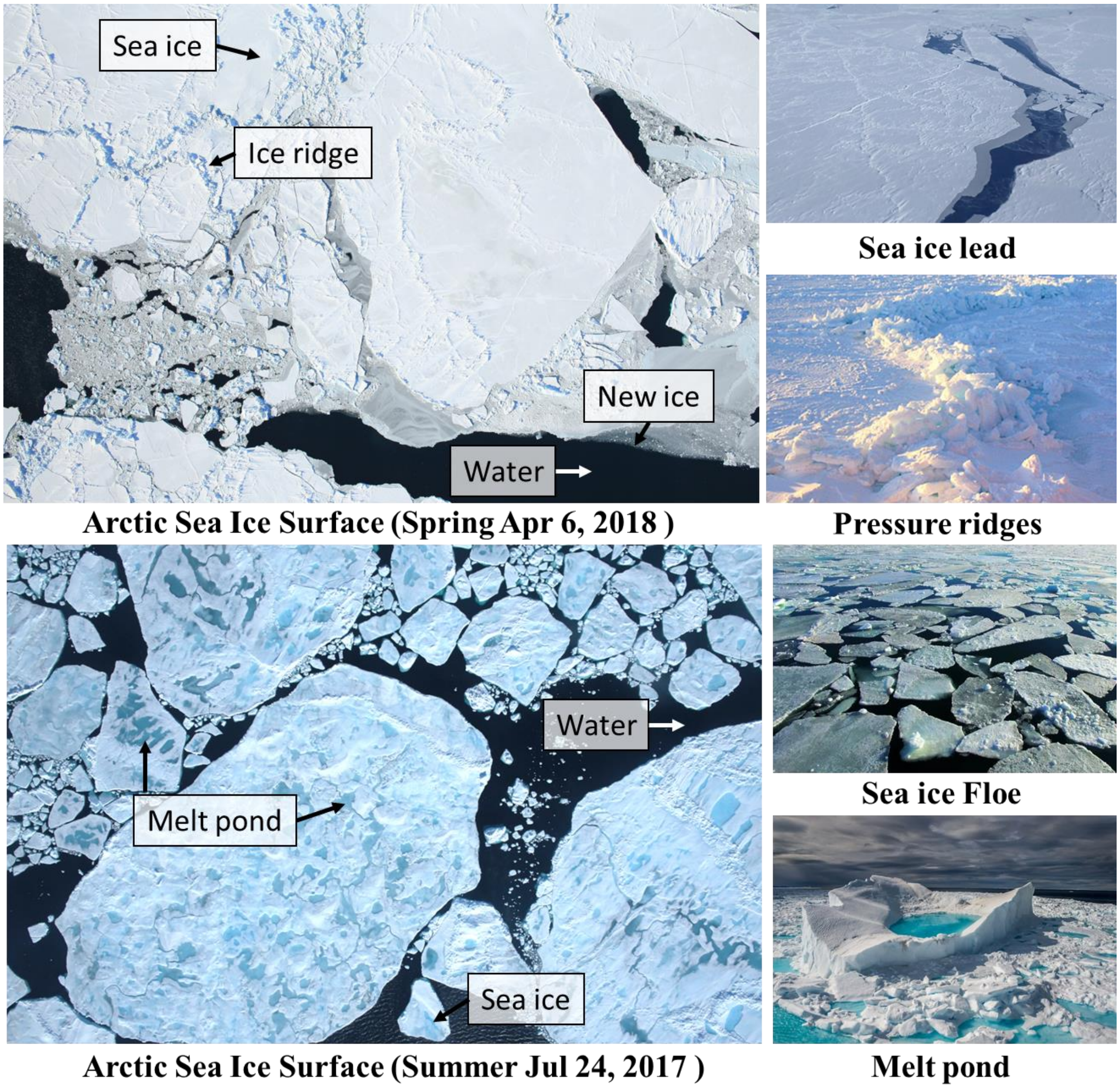


Figure 4. Typical sea ice features extracted from DMS imagery data (NSIDC) and photography work of the Arctic Ocean in the spring and summer season, photo credits to Joe MacGregor (NASA), Clay Machine Gun (Shutterstock), Ted Scambos (NSIDC) and Brian Skerry (National Geographic).

2.2.1 Sea Ice Parameters from Imagery

2.2.1.1 Sea ice type. Sea ice is the most important feature of the Arctic Ocean. There are various types of sea ice and sea ice types classified according to sea ice growth stages, states of motion, and surface size levels. In Figure 5, sea ice types are listed by three different perspectives (WMO standard) including sea ice growth stage, sea ice motion state, and sea ice surface size. Various sub-types also exist within each stage depending on the internal structure of the ice. Sea ice features and types can be identified and recognized by experienced observers in field work or by observing ground truth images. Ice Watch is an observation and data network program providing researchers with the Arctic Shipborne Sea Ice Standardization Tool (ASSIST) to capture sea ice conditions and processes unique to the Arctic, as well as traditional shipborne meteorological and sea surface reporting. They made a detailed user manual considering sea ice types and made a standard to help researchers classify sea ice types through distinctive morphological features (Hutchings et al., 2016).

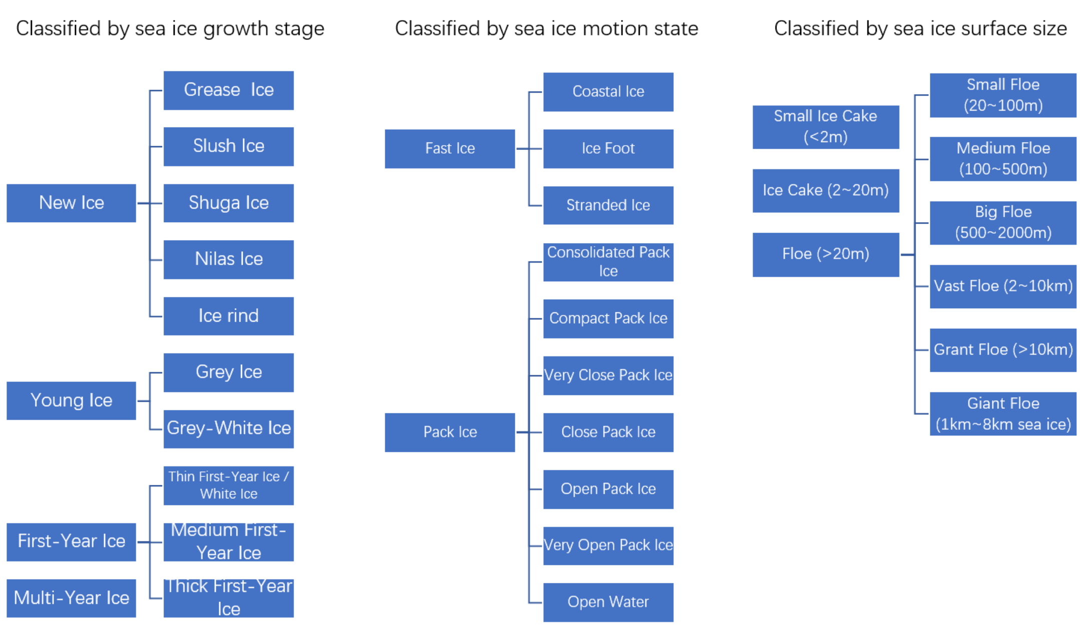


Figure 5. Three sea ice type classification perspectives.

Sea ice concentration, area, extent, and thickness products are the most common characteristics of sea ice and these indicators can be calculated from remote sensing products in multiple scales. Sea Ice Concentration refers to the fraction of ice-cover in a given area. Many sea ice parameters (such as area, extent, sea ice edge lines) can be obtained from sea ice concentration (SIC) results; therefore, the accuracy of sea ice concentration products will affect other studies. Measurements from ships and aircraft used to be based on the results of simply calculating the relative area of ice versus water visible within the scene from human eyes. In recent years, SIC has been estimated by HSR photographs based on automatically generated machine learning classification results with classification schemas of sea ice, water, melt ponds and melt sea ice.

Sea Ice Thickness provides the third dimension of information in sea ice research. Sea ice thickness is necessary for assessing sea ice mass balance, surface energy budget, seasonal and annual sea ice predictions, and changes in the polar climate system. The in-situ collection of ice cores and underwater observation of ice depth by submarines are the most direct approaches to measure ice thickness. Ship-based HSR imagery contains ice ridge height which could help calculate ice thickness when combined with other geophysical parameters such as ice type, density, and the relative scales of reference substances.

Melt ponds are pools of open water that form on sea ice during the spring and summer periods. Melt ponds are usually darker than the surrounding ice which makes them stand out, and their distribution and size is highly variable. Miao (2015) developed an object-based machine learning approach to extract accurate geographic information pertaining to melt ponds, sea water, submerged ice, and bare ice from HSR imagery. Other melt pond related properties shown in Table 3 can be calculated from melt pond results.

2.2.1.2 Sea ice lead. The ocean and atmosphere exert stresses on sea ice which create elongated cracks or openings (ice leads) where the ocean is exposed directly to the atmosphere. Leads cover a small fraction of the surface but dominate the vertical exchange of energy, particularly in winter when turbulent heat fluxes over leads can be orders of magnitude larger than they are over thick ice. Sea ice leads carry a significant importance in climate research and as a requirement for the remote retrieval of sea-ice freeboard and thickness from altimeters (Onana et al., 2013). The accurate identification of leads is critical in the precise estimation of sea level reference, but current numerical models are not capable of providing the required accuracy and spatiotemporal resolution for ocean surface topography (X. Wang et al., 2013). Hence, a local sea level reference or sea surface height derived from Airborne Topographic Mapper (ATM) elevation data must be used to determine snow freeboard and compute ice thickness.

In high spatial scale ice lead extraction studies, Onana et al. (2012) first identified the ice leads by developing and demonstrating an automated algorithm called Sea Ice Lead Detection Algorithm using Minimal Signal (SILDAMS) through the data fusion of the orthorectified optical DMS imagery from Operation IceBridge with altimetry data. Wang et al. (2013) then developed an automated approach to derive the local sea level reference from ATM data while simultaneously deriving snow freeboard for the computation of ice thickness from IceBridge data. Researchers then continued to expand the utility of lead detection methods by further accounting for the shaded areas under different solar illuminations through using dynamic pixel intensity thresholds (X. Wang et al., 2016) .

2.2.1.3 Sea ice freeboard. Sea ice freeboard refers to the height of snow plus sea ice surface above the water level (Kurtz et al., 2013). It allows for the estimation of thickness and volume of sea ice, and hence is the property retrievable by airborne and satellite altimeters. Knowledge of snow depth is useful for precipitation trend and variability studies as well as melt pond coverage, and also plays an important role in the retrieval of sea ice thickness from altimeter data sets. Accurate measurement of sea ice freeboard may be obstructed by factors including snow cover, melt ponds on the surface, and weather patterns, but modeling has been improved over the last decade. Additionally, the inclusion of additional instruments will lead to improved sea ice freeboard retrieval results in subsequent years.

In high spatial resolution extraction and application of sea ice freeboard studies, Kurtz et al. (2013) used DMS and Continuous Airborne Mapping by Optical Translator (CAMBOT) images to identify morphological features on sea ice. Multiple Altimeter Beam Experimental Lidar (MABEL) laser altimetry data was utilized to derive sea-ice freeboard while simultaneously using novel snow depth estimates from IceBridge to assess sea-ice thickness estimates derived from MABEL freeboard data, demonstrating the effectiveness of laser altimetry data alongside snow depth and snow/ice density data for reliably determining freeboard and sea-ice thickness (Farrell et al., 2015).

2.2.1.4 Pressure ridges. Sea ice pressure ridges are defined as thick features accounting for around one-half of total sea ice volume, developing in sea ice cover as a result of stress from currents and winds which causes separate ice floes to move and collide (Miao et al., 2016). Ice pressure ridges are an important sea ice mechanical and mass distribution feature, attributes of which can be used to refine, validate, and improve sea ice and climate models (Miao et al., 2016). These structural deformations are of climatological interest due to their impact on the mass, energy, and momentum transfer of the polar oceans; understanding regional and seasonal distribution as well as the variability of ridges is important for quantifying total sea-ice mass and for improving treatment of sea-ice dynamics in high-resolution numerical models (K Duncan et al., 2018). Sail height represents the raised part of the ridge above local sea-ice surfaces and is dependent on the thickness of the parent ice floe (Duncan et al., 2018). Sail height can provide information on the strength of the parent ice sheet forming the ridge and increasing the quality of sail height measurements will enable estimation of variability in ridge parameters and lead to improved representation in sea-ice models. Recently, the regional and temporal variability in sail height has been important for understanding the observed increase in sea ice drift and observed decline in sea ice thickness (Duncan et al., 2020).

Miao et al. (2016) first developed a batch processing algorithm for the automatic detection of sea ice pressure ridge location and profile from HSR optical imagery based on shadow geometry features, which effectively negated the issue of limited spatial resolution of Synthetic-aperture radar (SAR) data. Based on Miao’s method (2016), Duncan et al. (2018) further developed the identification and mapping of shadows cast by pressure ridges in OIB DMS imagery and derived sail heights by comparing surface elevation measurements from ATM data. In 2020, Duncan applied the derived sail heights in small regions of the Arctic by looking at a larger, Arctic-wide scale using the wide geographical coverage of OIB data.

Table 3. Sea ice geophysical properties to be derived from the HSR images.

#	Name	Description
1	Ice concentration	Fraction of ice-cover in a given area
2	Ice edge	Boundary between an area of ice and open ocean
3	Floe size distribution	Probability Density Function (PDF) of ice floe diameters/ areas
4	Ice freeboard	The height of the snow plus sea ice surface above the water level
5	Ice lead	The narrow, linear cracks in the ice that form by ice floes movement of divergency and shearing
6	Melt pond distribution	PDF of melt pool diameters/ areas
7	Fresh water of melt ponds	Product of the areas and depths of melt ponds
8	Lateral melting	Melt rate of ice at the edges of ice floes
9	Surface roughness	PDF of the elevation of ice above level ice, derived from shadow
10	Ridge height	Height of the (usually) linear features above the surrounding level or undeformed ice
11	Fractional heat transferring into the ocean	Cumulative fraction of solar heat incident on ice/snow, submerged ice, melt ponds, and open water, weighted by the area and transmittance of each component. For actual heat transferring, we will have to use the incident solar irradiance from reanalysis products. This might bring some uncertainties, but it is the best we have (can get) in the present day.

In summary, these sea ice products can be directly used to validate other coarse resolution remote sensing images/products. Furthermore, the above derived sea ice geophysical properties summarized in Table 3 can be analyzed to address scientific objectives such as, but not limited to, (1) analyzing the evolutions of ice concentration and edge, floe size distributions, melt pond distributions, lateral melting processes, surface roughness, and ridge heights, (2) examining the fractional heat transferring into the ocean through leads/water, melt ponds, submerged ice, and bare/snow-covered ice, (3) examining fresh water volume and change based on melt pond distribution, depth and areas, and (4) calibrating/validating sea ice modeling outputs/parameters. Specifically, derived sea ice geophysical properties and sea ice trajectories can be compared to sea ice model results. Currently, it is challenging for scientists to use and evaluate the extraction algorithms and derived parameters in very high spatial scales. This research will provide a use case of sea ice leads extraction with a practical workflow to fill the aforementioned gap.

2.2.2 Evolution of Sea Ice Classification Approach

Remote sensing methodologies using HSR imagery such as aerial captured imagery is essential for the accurate calibration and validation of climate models because arctic sea ice is an important climate change indicator. In the past, most high-resolution sea ice aerial or ship-based photos were analyzed through pixel-based methods (Lu et al., 2010a; Renner et al., 2013; Weissling et al., 2009). Pixel-based methods based on pixel brightness values or spectral values ignores spatial autocorrelation and generates ‘salt-and-pepper’ noise in classification (Liu & Xia, 2010; Xie et al., 2007). In contrast, object-based classification has been developed based on image segmentation, the process of partitioning an image into multiple objects or groups of pixels, thus making it more meaningful and easier to analyze (Hussain et al., 2013; Shapiro & Rosenfeld, 1992). This method not only considers spectral values but also spatial measurements which characterize the shape, texture, and contextual properties of the region to potentially improve classification accuracy (Liu & Xia, 2010).

A random forest object-based classification algorithm was developed to automatically and efficiently extract and classify sea ice (water, general submerged ice, shadow, and ice/snow) and melt ponds from optical imagery, thus alleviating the time-consuming and labor-intensive effort of manually delineating sea ice and melt ponds. However, the workflow of the authors was based on proprietary remote sensing software packages not capable of batch processing in an operational environment (Miao et al., 2015). As Figure 6 shows, commercial software such as ENVI, eCognition, MATLAB and ArcMap were used in image preprocessing, segmentation, feature engineering/classification and post-classification analysis, respectively (Miao et al., 2015). In this phase, the professional license and the knowledge of remote sensing and geographical system is required for competing the production workflow, which become a restriction for users with no professional commercial tools training. Then in 2018, the Open Source Sea-ice Processing (OSSP) package was developed based on Python libraries for detecting sea ice surface features and classifying sea ice surface types from HSR optical products, specifically the OIB DMS and WordView 2/3 satellite images (Wright & Polashenski, 2018). In the OSSP package, a watershed-based image segmentation algorithm and random forest machine learning classifier was implemented through the scikit-image and scikit-learn libraries on the Miao (2018) workflow. This package was then improved and applied to investigate the behavior of meltwater on first-year ice and multiyear ice during summer melting seasons (Wright et al., 2020). The challenge for new users not only needs to aware of remote sensing knowledge, but also have open-source configuration and debugging issue under potential computing environments. The pre-labeled training set may not work well with high accuracy in complex lighting and weather conditions. However, Sha, et al. (2020) implemented and further improved the OSSP module and integrated it into an on-demand web service, ArcCI, with a cloud-based infrastructure capable of operational usage. Dozens of pre-labelled training set is created for DMS image classification under various capture scenarios and users could take advantage the friendly interface to conduct batch production of self-updated imagery. The limitation for all three generation of OBIA classification for sea ice is the leak supporting of all customized image set.

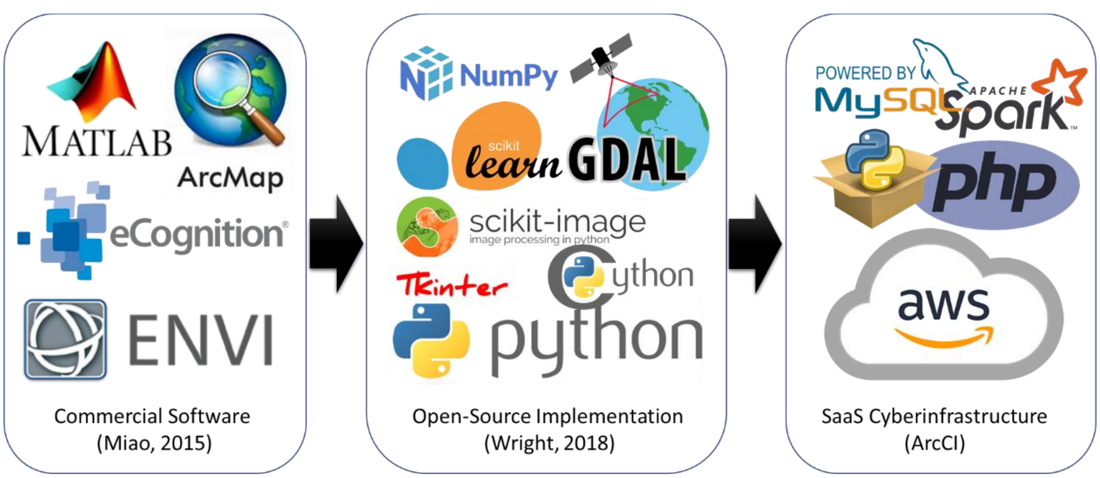


Figure 6. Evolution of Object-based Image Classification Tools Used for Sea Ice Classification.

In the past decade, deep learning (DL) has become an exciting new frontier in machine learning and artificial intelligence (AI) (Goodfellow et al., 2016). It can effectively learn from data and solve complex classification problems with a higher accuracy than traditional machine learning techniques. DL methodologies have been employed for sea ice classification and detection to achieve greater accuracy. Through a combination of active learning and semi supervised learning classification approaches, deep learning techniques achieved a high classification accuracy on hyperspectral and multispectral datasets and proved that active learning is able to find the most informative samples from small labels (Han et al., 2018). However, in order to take advantage of this new paradigm, a large amount of labelled data is required to feed into the deep and complex algorithms, similar to how large fuel reserves are required to power a rocket engine (Ng et al., 2015). As a rough rule of thumb, supervised deep learning algorithms will achieve acceptable performance with around 5,000 labeled samples per category, and will match or exceed human performance when provided with at least 10 million labeled training samples (Goodfellow et al., 2016). Furthermore, the performance on computer vision deep learning tasks increases logarithmically with the volume/size of training data (Sun et al., 2017). Overall, classified sea ice image results from object-based image analysis could be an important training input for deep learning models, increasing the sample volume and decreasing misclassification performance caused by illumination, weather conditions, and other complex situations.

2.3 Cyberinfrastructures and Web Services of Polar Science

The Cyberinfrastructures (CI) and web services of polar science have evolved quickly over the past decade. This section introduces the highlighted products and services employed in polar studies and data services over the past few decades.

2.3.1 Three Generations of Polar Web Services

2.3.1.1 Data archive. The first generation of polar CIs consists of static data infrastructure, focusing on interoperability at the data level and only providing comprehensive data deposits through static web pages. Data archive web services are usually attached under the homepage of the research institution or research project. A data archive is capable of displaying information including metadata and allows users to download stored raw datasets from backend servers, simultaneously providing search, query, visualization, and interactive data discovery functionalities based on attributes of the metadata.

For example, the Arctic Research Mapping Application (ARMAP) was designed to access, query, and browse the Arctic Research Logistics Support Service database (Walker Johnson et al., 2011). The Arctic Data Repository (ACADIS) is a joint effort by the National Snow and Ice Data Center (NSIDC), the University Corporation for Atmospheric Research (UCAR), UNIDATA, and the National Center for Atmospheric Research (NCAR) to provide a portal for Arctic Observing Network (AON) data and is currently being expanded to include all National Science Foundation Applied Research Center (NSF-ARC) data (Jodha Khalsa et al., 2013). The Polar Geospatial Center provides geospatial mapping services and collects Alaska High Altitude Photography, Landsat, and MODIS images. The Norwegian Polar Data Centre provides a dataset service under its homepage with all published and unpublished datasets created by the Norwegian Polar Institute (https://data.npolar.no/) (Norwegian Polar Institute, n.d.). The Ice Archive from the Government of Canada allows users to search for archived charts and data, view individual dataset online, and download zipped files by self-packages through web services.

2.3.1.2 Data portal. The second generation of CIs began to consider intelligent data discovery and provide access through web crawlers, Internet mining, and advanced functionalities of data integration and visualization approaches (Z. Li et al., 2017).

The data portal website not only provides data archiving, indexing, searching, downloading, and other services, but also provides more vivid data visualizations using front-end dynamic interaction and other website development technologies including interactive WebGIS maps and statistical data charts. Data portals interactively display different thematic data in the same area through dynamic map services and provide one-stop query services by aggregating raw data and metadata through a web data portal, collecting and storing more data from researchers.

For example, Arctic Portal (http://portal.inter-map.com/) is one such data portal used by various Arctic-related organizations, affiliations, initiatives, and projects. The Arctic Data Interface is designed to provide retrieval and interfacing services for observational metadata and consequently, data interpretation and access tools for customers on demand. Multiple layers of location-based information are available to flexibly display in a WebGIS interface. Relevant documents, project databases, virtual libraries, events links, and multimedia materials are integrated and posted on this one-stop data portal.

The Swedish Polar Research Portal (https://polarforskningsportalen.se/en/arctic) presents onsite photos, cruise reports, and expedition blogs about polar research expeditions conducted by polar researchers since 1999. This portal gives a unique insight into the work and daily life of researchers during their expeditions in Arctic regions and Antarctica. Researchers could take advantage of this platform as a metadata service and as an index for specific spatiotemporal records.

Ice Watch (https://icewatch.met.no/), coordinated by the International Arctic Research Center, is an open-source portal for sharing shipborne Arctic sea ice observation data and ship-captured images. Extracted geophysical attributes could be uploaded and shared on the web service as well.

The NSF-funded Arctic Data Center (https://arcticdata.io/) allows researchers to document and archive data in diverse formats as part of their normal workflow using a convenient submission tool. This infrastructure comes with a set of community services including data discovery tools, metadata assessment and editing, data cleansing and integration, data management consulting, and user help-desk services based on dataset sharing. The Polar Geospatial Center (PGC) provides geospatial mapping services and collects Alaska High Altitude Photography, Landsat, and MODIS images.

2.3.1.3 Data platform. The emerging third generation of CIs can be defined as a knowledge infrastructure providing rudimentary interactive analysis and reasoning modules. For example, a multi-faceted visualization module for complex climate patterns with an intelligent spatiotemporal reasoning system has been proposed recently (Yang et al., 2015). Knowledge discovery can be implemented through an on-demand cloud computing system, and data processing could be done on the fly in the backend.

Data platform web services extend the functionality of the previous two generations of data web services while providing more possibilities for data analysis and mining. In terms of functionality, users will be able to upload data from a web browser and store them in a backend storage system or database, as well as employ a real-time analysis workflow to discover and share customized analytical results and mined knowledge.

With the advancement of technology, cloud computing has become a new and advantageous computing paradigm to solve scientific problems which have traditionally required large-scale high-performance clusters since it provides a flexible, elastic, and virtualized pool of computational resources (Huang et al., 2013). Cloud computing is suitable for supporting on-demand services of ArcCI with the following advantages: (1) it can manage distributed storage for big data; (2) it leverages scalable computing resources for dynamic on-demand web services, which often causes computing spikes; and (3) it provides a transparent implementation for running models so scientists can focus on research without having to consider underlying computational mechanisms.

Distributed file systems (DFS) and distributed computing frameworks are two core components in big data processing systems. DFS provides the capabilities of transparent replication and fault tolerance to enhance reliability. The backup storage automatically makes a secondary copy (or even more copies) of the data so that it is available for recovery if the original data is damaged (Yang & Huang, 2013). On the other hand, distributed computing techniques enable high-performance computing on big data.

Google Earth Engine (GEE) is a data platform serving remote sensing images. GEE is a cloud-based computing platform allowing planetary-scale analysis capabilities through a combination of a petabyte of satellite imagery and geospatial datasets on a global spatial scale. Scientists, researchers, and developers can get free access for detecting changes, mapping trends, and quantifying differences on various properties of the Earth's surface based on GEE services (Gorelick et al., 2017).

There is no highly specialized Arctic cyberinfrastructure building block which emphasizes (1) HSR sea ice image collection, (2) on-demand value-added services such as automatic batch image classification and geophysical parameter extraction, and (3) spatial-temporal visual analytics of sea ice evolution. This is a core motivation for us to develop a CI building block to serve the Arctic sea ice community as well as the polar sciences community in general.

2.3.2 Advantages Technology in Web Infrastructures

Cloud computing delivers scalable, on-demand, pay-as-you-go access to a pool of computing resources (Mell et al., 2011; Yang et al., 2011). Cloud computing aims to maximize the utilization rate of geophysical resources and provide virtual resources to aid applications and services. Cloud computing relies on several technologies including virtualization, network security, and high availability to provide services over the network. These technologies make it easier, more efficient, and more economical to set up an architecture for big data analysis.

In cloud computing solutions within the commercial market, public cloud and private cloud services are the two major sectors. Public cloud services such as Amazon Web Services, Microsoft Azure, and Google Cloud are the most accessible services provided by big Information Technology (IT) companies; private cloud services are widely implemented and applied by organizations with self-purchased computing hardware and a secure private network. From a cloud-based services perspective, cloud computing can be categorized into three types: Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS). The abundant service ecosystem enables general and domain-specific users to take advantage of the capabilities of computing resources, thus mitigating the limitation of a geophysical location.

High Performance Computing (HPC) is a loosely coupled set of computing resources composed of multiple computing nodes (servers) for the purpose of solving complex scientific analytical and simulation problems. Among the HPC cluster, multiple server nodes are connected through corresponding hardware and a high-speed network. Controlled by software, the complex computing tasks are decomposed and distributed to each computing node. Each node conducts its own process independently in a parallel manner, and the processes can communicate with each other by exchanging data. The overall objective of HPC is to use a parallel computing processing framework to run advanced application programs efficiently, reliably, and quickly. The concept of cloud computing brings a new way of organizing computing resources and provides an innovative application model for high performance computing from on-demand access and scalability of compute and storage resources (Vecchiola et al., 2009). As an example, the prototype design of ArcCI's massive image analysis application requires a TB-level storage and corresponding computing resources. Furthermore, algorithms for image analysis are complex and time-consuming on single server nodes. Thus, it is necessary to utilize high performance computing technology to distribute the data into multiple nodes and conduct algorithms in parallel. However, since the processing demand is irregular when the factor of time is considered, dynamic support of computing sources is more efficient than a traditional static cluster. In an industry-level implementation, Amazon Web Services (AWS) provides an elastic and scalable cloud infrastructure for scientists to run HPC applications beyond the limitations of on-premises HPC infrastructure.

Cloud-based high-performance techniques have rarely been integrated and tested in a HSR image processing system, and the specific implementation of the aforementioned structures and components could provide new possibilities in the market.

CHAPTER 3. DEVELOPING AN ON-DEMAND SERVICE FOR MANAGING AND ANALYZING ARCTIC SEA ICE HIGH SPATIAL RESOLUTION IMAGERY

3.1 Introduction

Remote sensing is a valuable technique in Arctic sea ice research that helps researchers detect sea ice geophysical parameters and calibrate/validate climate models (National Research Council, 2007). Big remote sensing image data are collected from multiple platforms in Arctic regions on a daily basis, thus posing the serious challenge of discovering spatiotemporal patterns from the aforementioned big data in a timely manner (Xindong Wu et al., 2014). This demand is driving the development of data CI, data mining, and machine learning technologies.

Most of the existing Arctic CI systems focus on low spatial resolution imagery without generally including high spatial resolution (HSR) images. Compared to low resolution imagery, HSR can provide incomparable details of small-scale sea ice features. One of these features is melt ponds, which develop on Arctic sea ice due to the melting of snow and upper layers of sea ice in the summer. Once developed, melt ponds have a lower albedo than the surrounding ice, absorbing a greater fraction of incident solar radiation and increasing the melt rate beneath pond-covered ice by two to three times compared to that below bare ice (Flocco & Feltham, 2007). Therefore, an accurate estimate of the fraction of melt ponds is essential for a realistic estimate of the albedo for global climate modeling and improving our understanding of the future of Arctic sea ice. Unfortunately, a typical melt pond cannot be seen in low spatial resolution images due to its relatively small size. Only HRS images can provide detailed spatial distribution information pertaining to melt ponds and other fine sea ice features.

HSR images are difficult to process and manage due to three factors: (1) the data and/or file size is usually very large compared to coarse resolution images; (2) HSR images are collected from multiple sources (e.g., airborne and satellite-borne) with varied spatial and temporal resolutions; (3) HSR usually has a complex and heterogeneous nature in both space and time. Unlike other moderate or low-resolution satellite images such as Moderate Resolution Imaging Spectroradiometer (MODIS) or Advanced Very High Resolution Radiometer (AVHRR), HSR images such as aerial photos usually cover only a small area without overlapping with other images and their time intervals vary between a few seconds and several months. Therefore, it is difficult to weave these small pieces of sparse information into a coherent large-scale picture, which is important for sea ice and climate modeling and verification.

This chapter introduces our efforts to develop a reliable and efficient on-demand image batch processing web service CI module (ArcCI) and its associated data sets. ArcCI as a data platform for extracting accurate spatial information of water, submerged ice, bare ice, melt ponds, and ridge shadows from a large volume of HSR imagery data with limited human intervention. It also has a 3D visualization function to explore the spatiotemporal evolution of sea ice features. Furthermore, the approach can be used in other polar CIs as an open plug-in module.

3.2 Methodology

ArcCI is designed and developed to support on-demand Arctic HSR image processing. We detail each part of the architecture for ArcCI: Section 3.2.1 provides an overview and key techniques used in each layer. Section 3.2.2 describes the methodologies used in data storage and metadata extraction. Section 3.2.3 introduces the workflow and algorithms used in image processing and analysis.

3.2.1 Architecture Design

The ArcCI architecture (Figure 7) consists of three layers. The distributed physical infrastructure layer (bottom layer) provides the physical computing resources for supporting all computing requirements of the system. Above the physical infrastructure layer is a software layer that includes the operating system, cloud software, and database management system, providing cloud advantageous services such as elasticity and on demand. Virtualized machines are utilized to ease the system development, integration, and deployment. The software layer includes the community private cloud computing environment at George Mason University (GMU), and the public cloud computing environment at Amazon, both of which are currently serving the public (Yang & Huang, 2013) through the NSF Spatiotemporal Innovation Center with integration to best leverage the cloud computing environment for sea ice research. The top layer is developed to provide different types of on-demand services including Extract-Transform-Load (ETL) processing and data storage, image processing, parameter extraction, and spatiotemporal visual analyses. This layer also provides a graphical user interface (GUI) for geo-search and query functions, and it can be remotely used by desktop computers or mobile computing devices (Gui et al., 2013), so as to support the data life cycle of generation/discovery, processing, analyses, and visualization for end users (Z. Li et al., 2011). On top of the three-layer architecture, many applications can be customized by end users based on specific polar science research requirements.

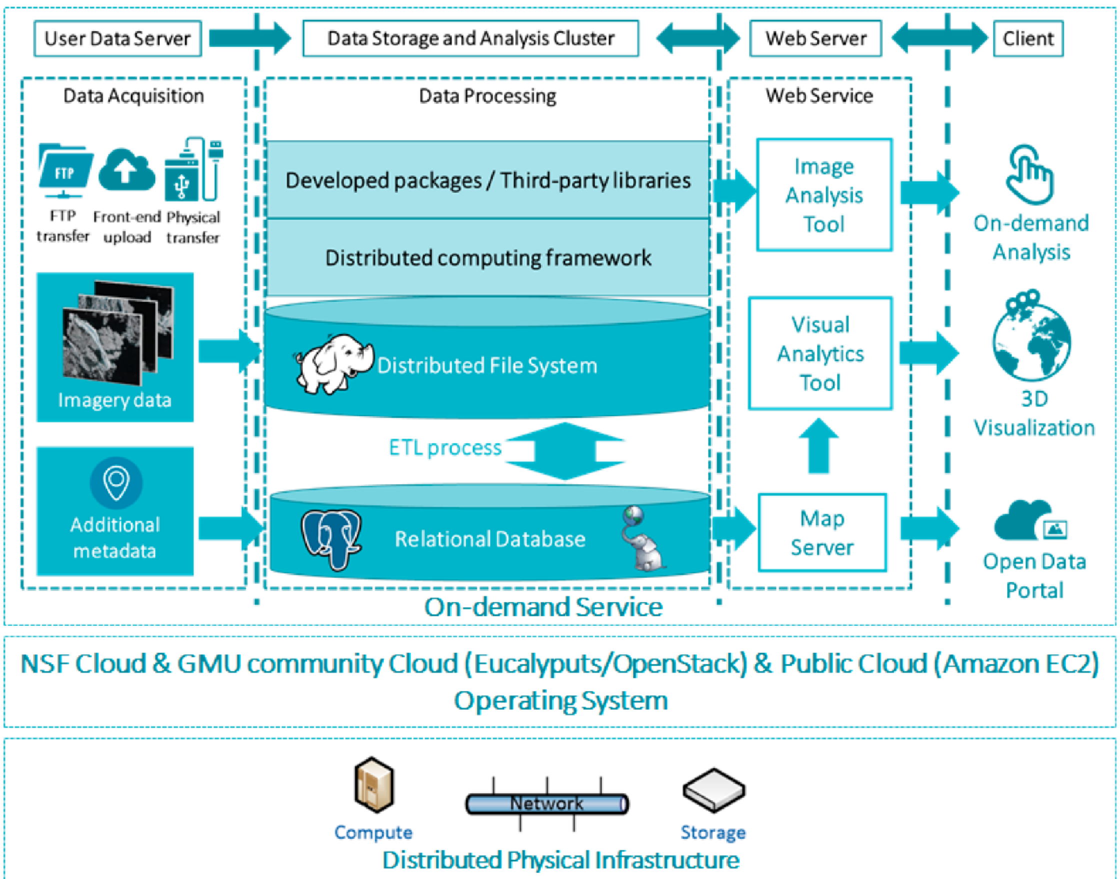


Figure 7. Concept model of ArcCI architecture.

ArcCI hosts a big data platform in the cloud with comprehensive components to support web services. All components were deployed on an elastic number of virtual machines from a resource pool that combine CPU cores, RAM (random-access memory) for computing, and hard drive arrays for data storage. Four key components form the skeleton of ArcCI. The first component is the distributed file system. As a fundamental component of the proposed infrastructure, the distributed data management system provides scalable storage to store large amounts of HSR raster data upon Hadoop Distributed File System (HDFS). The files with Geographical Tagged Image File Format (GeoTiff), Joint Photographic Experts Group (JPEG) and Portable Network Graphics (PNG) formats can be directly uploaded into HDFS without conversion. The second component of ArcCI is Apache Spark, a distributed computing engine used to process large amounts of HSR imagery data. A Resilient Distributed Dataset (RDD) based data frame structure is used to represent image elements in the distributed cluster. RDD is the basic data structure for data transformation, image processing, and image analysis, such as image reading, segmentation, and classification, in Spark. Hadoop distributed file system and Spark are the most popular implementations of distributed file systems and distributed memory-based computing frameworks of the Apache big data ecosystem. The third component of ArcCI is a relational database which is embedded in the proposed framework to store metadata and extract features from HSR imagery. The output results from the distributed computing engine are exported to a relational database, upon which GeoServer will provide Web Map Service (WMS)/Web Feature Service (WFS) APIs for further web services. GeoServer is deployed to serve as an online map server for 3D visualization. PostgreSQL and GeoServer is a mature and popular combination for open source WebGIS projects supporting the Open Geospatial Consortium (OGC) standard and a wide range of users. The fourth component is the web portal and services. A Comprehensive Knowledge Archive Network (CKAN) based open science gateway is deployed on the web server to provide a data landscape for sea ice research. Based on the GeoServer Application Programming Interface (API), a 3D visualization tool is created for visual and interactive exploration of extracted features. Jupyter Notebook, an open-source web application, is set up as a programming platform for developing new workflow or image analysis algorithms requested by users. Distributed computing tasks can be created and shared in a Jupyter-based interactive code editor.

3.2.2 ArcCI Database Design and Data Pipeline

3.2.2.1 Database Design. The ArcCI system is designed for processing multi-source HSR image data for multiple users. Figure 8 demonstrates the Unified Modeling Language (UML) diagram of the database design for the ArcCI system, including metadata for single images and image collections, profile information of users, organizations, and projects. All tables are created and stored in a relational database as shown in Figure 8.

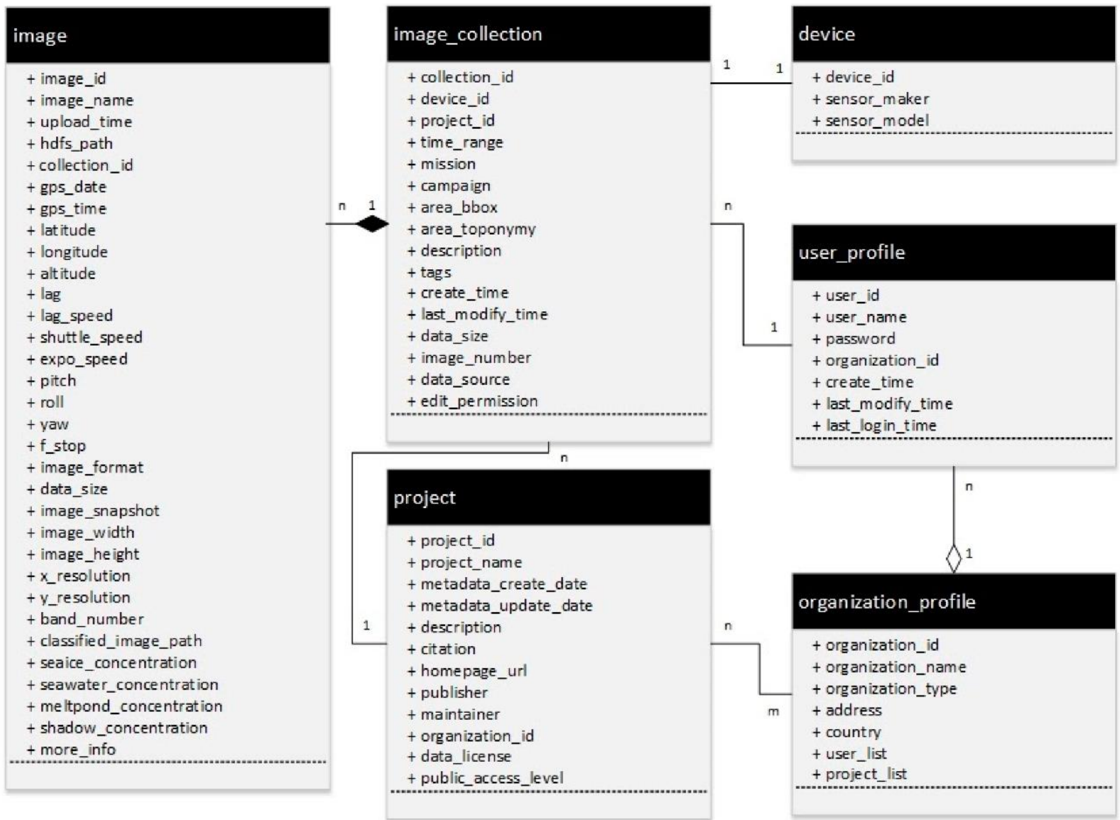


Figure 8. Unified Modeling Language (UML) diagram of the database scheme.

The “image” attributes table is a big table that records all valid information related to single HSR images. A unique ID, update time, and HDFS path for each image is automatically generated when data is uploaded. Supplementary metadata information including Global Positioning System (GPS) date and time, spatial information pertaining to latitude, longitude, and altitude, as well as shuttle (lag speed, pitch, roll, and yaw) and photographic (shutter speed and f-stop) information was collected from GPS devices during flight. Image parameters are extracted from raw image metadata, including image format, data size, width, height, resolution on x and y, band number, and processed output path for image snapshots and vector shapefiles. Extracted geophysical attributes based on the image including the concentration values for sea ice, open water, melt pond and shadow are also created in the image table. More information includes general attributes created for additional unstructured information for heterogeneous data sources.

The “image-collection” table stores all the essential attributes of one-time data uploads and transfer operations from users to the system. Each image collection contains images from the same collection mission with continuous timestamps. The attributes for image collection include ID number, related device and project ID, image capture time range, mission and campaign name, spatial extent in bounding box, description, tags, etc. Other data management information is kept in this table, including time of creation and last modification, data size, image number, and data source. Due to data license and usage policies, raw data can only be viewed, edited, analyzed or downloaded with permission from the data owner. Attributes edit permissions are created in the image collection table to store the privilege of a data editor based on the user’s ID.

The “device” table contains sensor information including manufacturer brands (e.g., Nikon and Canon) and models (such as EOS 5D Mark Ⅱ) utilized in the Operation IceBridge DMS dataset.

The “user_profile” and “organization_profile” tables are designed for data upload management, which means the original data owner might be different from the data upload user. Each organization may have multiple users while one user belongs to a specific organization. The user profile table records users’ email address as well as their unique IDs and other profile information such as full name, organization, profile creation and modification times, etc. Users' passwords are stored as an encrypted string for privacy and security. The organization profile table records ID, name, type, address and country information, and user and project lists associated with each organization.

The "project" table contains metadata for a research project with several image collection tasks based on flight missions. As an overview table for Arctic research, the attribute is designed for communities to review and cite related works and data. The attributes include information on project ID, name, metadata creation time, description, citation information, homepage link, publisher and maintainer information, and data permission information such as data license type and public access level. The project metadata can easily be utilized in a CKAN-based open science gateway.

3.2.2.2 Data Acquisition and ETL Process. In the ArcCI system, heterogeneous raw datasets from different sources are collected through three principal approaches, including File Transfer Protocol (FTP) server transfer from a current Arctic sea ice image portal as well as physical copy and browser uploading from data owners. These data transfer approaches depend on data volume and usage license/open-source policies. The acquired data can be classified into three different formats:

1. Packaged and georeferenced image products in Tagged Image File Format (TIFF) and Portable Document Format (PDF) file formats including raster imagery and all available metadata saved in the file header.

2. Raw image files in JPEG and PNG formats with supplementary metadata files related to each image in Comma Separated Values (CSV) and Text formats. Image files

only record raster-based information and image metadata, while other location and flight information is recorded by CSV and Text.

3. Raw image files with qualitative descriptions. For example, in early Arctic exploration surveys, few photos were taken in each mission and these photos generally have brief and simple records. Obviously, these images would not be available for Point of Interest (POI) based quantitative research.

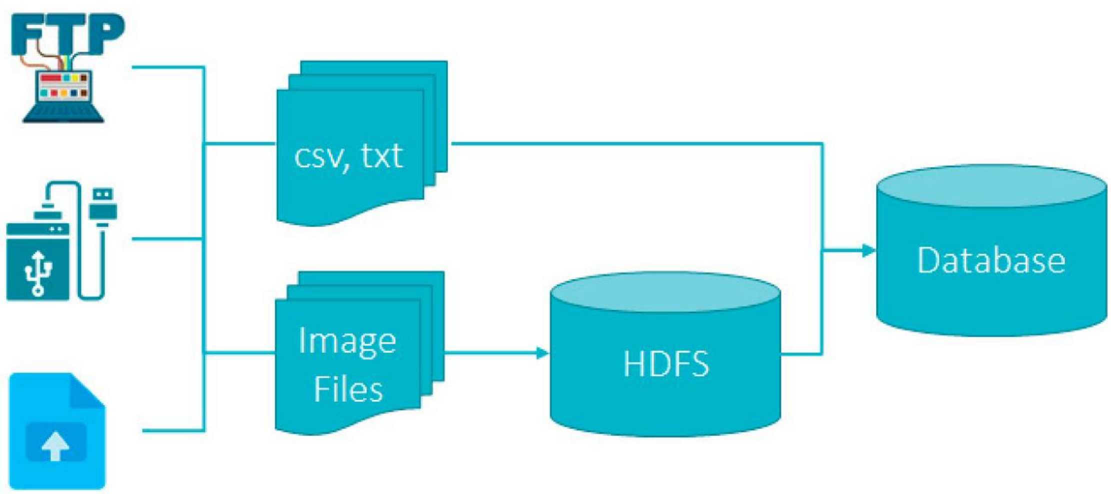


Figure 9. Extract-Transform-Load (ETL) workflow for ArcCI.

Once data is transferred into the system, an Extract-Transform-Load (ETL, Figure 9) process is automatically activated to process raw data into a data format for final client usage. In a traditional ETL workflow, data is extracted from online transaction processing databases and then transformed into a staging area. These transformations cover both data cleaning and optimization. Finally, the transformed data is loaded into an online analytical processing database. Figure 9 shows the data acquisition and ETL process, which is customized based on the application logic of HSR imagery in ArcCI.

1. Location and flight metadata are extracted from formatted CSV and TXT files into a relational database.

2. The image is stored in HDFS first as a binary file, then an image metadata extraction script is developed based on file format to read the file header and extract image metadata such as data size as well as image shape and resolution into the relational database.

3. Heterogeneous data from multiple sensors, sources, and formats is converted and transformed into the designed data structure and loaded into the image table.

3.2.2.3 Distributed Image Analysis Tool. The distributed image analysis tool is based on the Spark computing architecture. After the ETL process, each image file is stored in HDFS as a non-structured binary file. Binary image files are read into memory and represented in RDD format for transformations and operations. Through function transfer and integration into the Spark environment, the developed algorithm is packaged as an image processing API function to be utilized in the RDD transformation process. After the operation, the RDD instance will be processed on each work node based on cluster configuration and task allocation strategies. Then, each node will return the processed RDD into memory and write the result into HDFS or other databases.

Figure 10 shows the Jupyter-based data processing ecosystem setup within cloud computing virtual machines. For the bottom part, Python version 3.7.3 is selected as the basic programming language and the PySpark library is used as the distributed computing framework. The Anaconda platform is used to configure all Python-related components including the Jupyter Notebook for on-demand analysis and the Spyder scientific environment for development processes. GitHub is a code repository on the public cloud for real-time algorithm testing and deployment on clusters. Aside from fundamentally configuring Python, many third-party libraries are installed and imported including Geospatial Data Abstraction Library (GDAL) for raster format reading, NumPy for multi-dimensional array data structures, OpenCV for standard image preprocessing, the scikit-image package for the segmentation algorithm, the scikit-learn package for classification training and production, and other Python libraries for auxiliary tools in development workflow. This Jupyter Notebook engine plays the core role in image analysis which connects remote users, the data storage system, and data processing functions. All third-party libraries are configured on each of the compute nodes in cluster mode, and the developed image classification and parameter extraction software are packaged with user friendly GUIs. Users can easily call the function to process their data using simple scripting in the Jupyter Notebook.

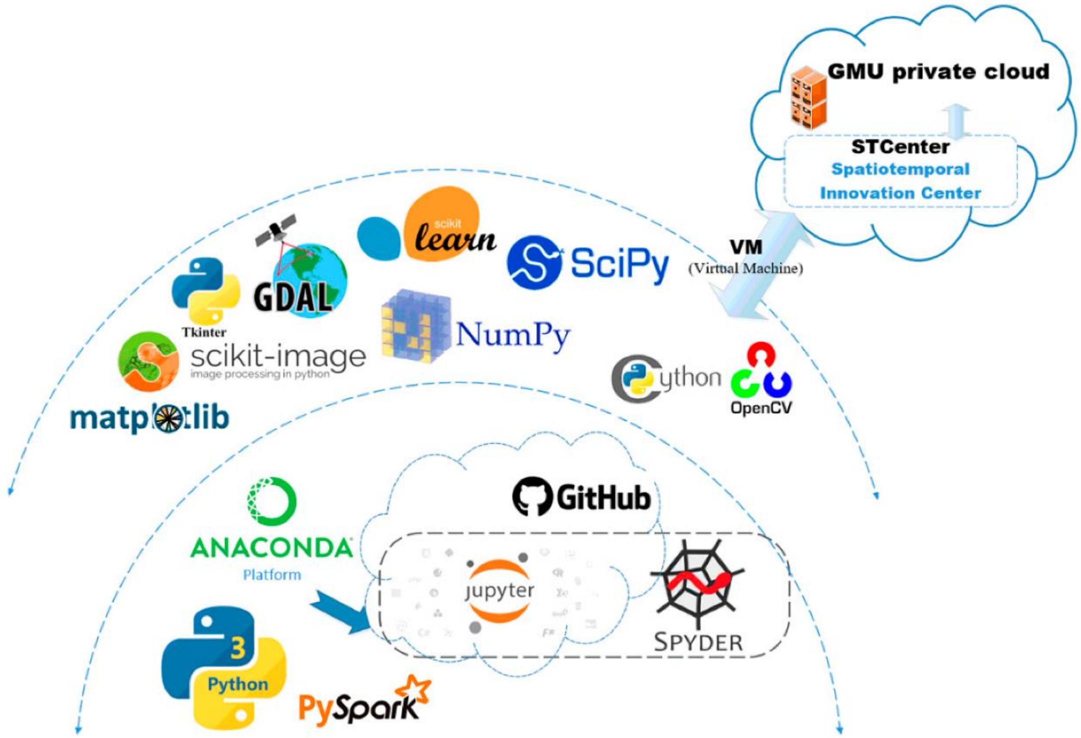


Figure 10. Jupyter Notebook ecosystem for image analysis.

3.2.3 3D Visualization Tool

The objective of 3D visualization is to use an effective way to visualize multidimensional geophysical data or features extracted from raw HSR imagery. Specifically, it selects and illustrates Arctic sea ice features in a 3D spatiotemporal space in an interactive manner. The module is developed using JavaScript front-end techniques and deployed on GeoServer publishing WFS Geo-JavaScript Object Notation (GeoJSON) format data.

The embedded 3D virtual globe is built upon Cesium, an open-source virtual globe made with Web Graphics Library (WebGL) technology. This technique utilizes graphic resources at the client side by using JavaScript based libraries and WebGL to accelerate client-side visualization. The virtual globe has the capability of representing many different views of geospatial features on the surface of the Earth, and can support the exploration of a variety of geospatial data. It can dynamically load and visualize different kinds of geospatial data including tiled maps, raster maps, vector data, high-resolution worldwide terrain data, and 3D models. By running on a web browser and integrating distributed geospatial services worldwide, the virtual globe provides an effective way to explore the 3D spatiotemporal correlations between heterogeneous datasets and discover evolution patterns in the 3D space-time domain.

The main functions supported by the 3D visualization module are listed as follows. (1) The base map of the virtual globe is formed by georeferenced and pre-rendered low spatial resolution imagery and related terrain data in the Arctic region. All available tiled map services such as the Web Map Tile Service (WMTS) developed by the OGC, the Tile Map Service developed by the Open-Source Geospatial Foundation, ESRI ArcGIS Map Server imagery service, OpenStreetMap, Mapbox, and Bing Maps, can be easily loaded into the virtual globe as a base map. (2) The virtual globe can support real-time rendered WMS map services, and WFS as geodata layers on top of the base map. Therefore, the added geometry data, such as GPS point and expedition route, can be layered in order and blended smoothly in the scene. Each layer’s brightness, contrast, gamma, hue, and saturation can be controlled by the end user and dynamically changed. (3) A plug-in filter tool allows users to select specific geoinformation to illustrate and filter data by metadata attributes such as time range, project ID, and owner information.

3.3 System Implementation

The ArcCI system leverages essential cloud computing resources including virtual machines (VM), storage/file systems, and networking. The system incorporates web-based geoscience information services and analysis programming tools to customize the user interface for Arctic sea ice studies. The Openstack private cloud at GMU with a 504-node computer cluster is used to support both the physical and cloud environments. 21 VM nodes of this cloud have been utilized to deploy a Spark cluster (v2.4.0 + Hadoop v2.6.0) with one master node and 20 worker nodes, and the cluster resource is managed by Yarn. Each VM is configured with 24 GPU cores, 4 TB of storage, and 64 GB of RAM on the Centos 7.7 operating system (OS). A public VM on AWS is utilized to allow the science gateway to integrate all web services on a private cloud. All components on the system can be extracted as cloud VM image resources which can be transferred to benefit other polar CI and polar science research.

On the Software as a Service (SaaS) level, the ArcCI portal Gateway has multiple loosely coupled functionalities, so as to provide a life cycle service for HSR images from data uploading, storage, management, analysis, visualization, and sharing.

3.3.1 ArcCI Science Gateway

We created the Arctic High Spatial Resolution (ArcHSR) Imagery Science Gateway (Figure 11) to provide metadata for the sea ice community (http://archsri.stcenter.net/). Both collected and processed public and longtail datasets have been prepared for querying, browsing, and sharing. The ArcHSR science gateway also enables data owners to register user accounts, organizational pages, and create dataset pages. Multiple data licenses are used for data reusing, coping, publishing, distributing, transmitting, and adapting. All datasets can be accessed and cited for non-commercial purposes. More importantly, a well-designed tagging and grouping system is designed based on toponymy and sensor types and can be used to filter out the most relevant dataset for researchers.

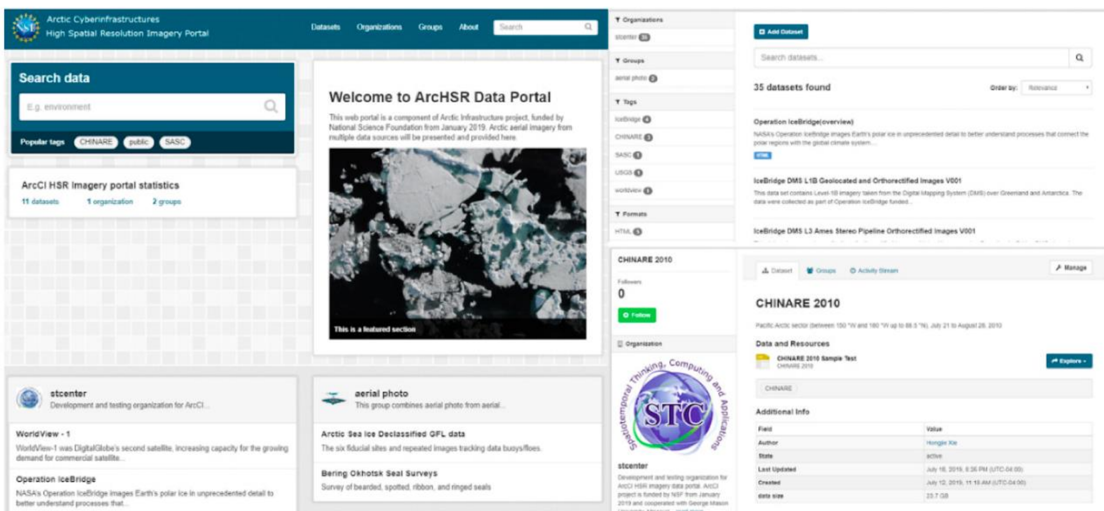


Figure 11. Screenshot of the ArcHSR imagery open portal.

So far, 35 sea ice image collections have been created and kept in the ArcHSR science gateway from multiple data sources including United States Geological Survey (USGS), National Aeronautics and Space Administration (NASA), National Snow and Ice Data Center (NSIDC), etc. Each collection is presented under an individual page with paragraph description and metadata pertaining to author name, contact method, creation time, data size and linkage, and samples of raw data. The raw data format related to sea ice metadata includes HyperText Markup Language (HTML) webpages, CSV tables with Point of Interest (POI) level records, unstructured text-based documents such PDF files and Word docs, and image examples in TIFF and JPEG format.

3.3.2 Data Workflow for Multiple users

The workflow for users with different demands for sea ice research is shown in Figure 12. We defined three typical users with different motivations for using this service. First, data owners have comprehensive control for uploading image data into the data storage server, managing datasets under permissions, and processing images based on provided services. Second, researchers can upload metadata or extract geophysical parameters through visual image interpretation. Third, users without data can still access sea ice geophysical parameters for climate model validation, simulation, and multi-platform data fusion. All visitors or users will be able to download extracted ice layers in a geospatial data format for further data analysis and fusion processes.

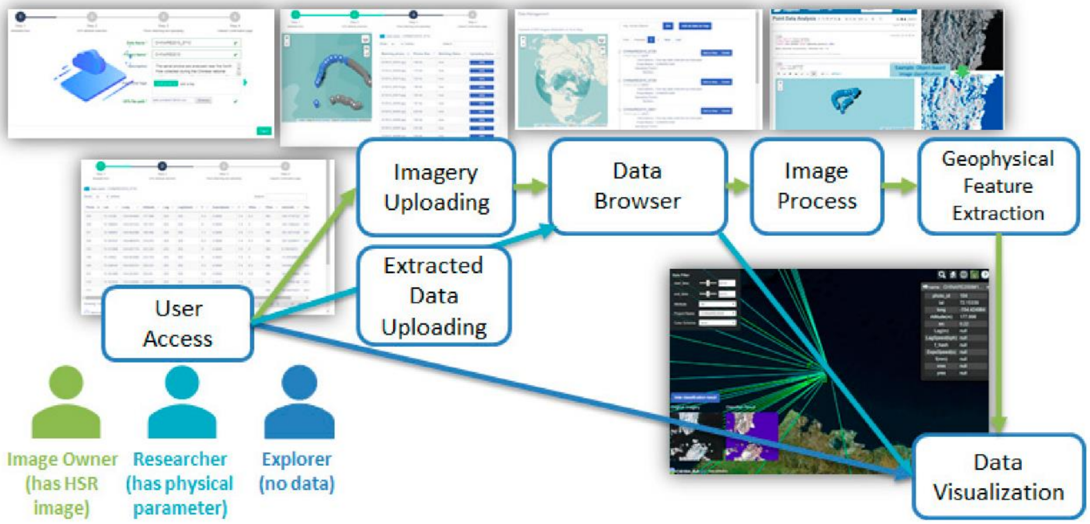


Figure 12. Users' views on functionalities.

3.3.3 3D Visualization Tool

The 3D spatiotemporal visualization tool (Figure 13) is designed to explore, visualize, and analyze sea ice evolution through an intuitive, interactive, and responsive GUI. The visualization module shows a 3D global map facing the North Pole from a slanted-top angle. The interactive interface allows users to move around and zoom in/out of the virtual global. In each scene, extracted attribute values are represented by self-adapting font size and classified colors while the location of the column-shape marker (in the central green square) refers to the coordinate of each processed HSR image. The top-left data filter tool provides a function to select sea ice parameters by time, attribute, and project ID.

By clicking each marker, detailed information for specific locations will pop up on the screen including (1) a top-right table which shows extracted attributes and metadata such as sea ice concentration, sea water concentration, melt pond concentration, latitude, longitude, and photo ID; 2) a bottom-left preview window which shows images before and after image classification; 3) a bottom-right chart figure which shows the proportion of four extracted geophysical parameters, i.e., sea ice, sea water, melt pond, and shadow.

Figure 13. 3D visualization module of extracted sea ice properties.

3.4 Summary

For better solving the big data challenge of HSR sea ice imagery in massive image amount, heterogenous data sources and quick update of new data, this research proposes and implements 1) a cloud computing-based cyberinfrastructure to collect, search, explore, visualize, organize, analyze and share the High Spatial Resolution images which are discrete in time and space; 2) a prototype of sea ice image online service for domain scientists to classify image and extract geophysical parameters. The developed ArcCI is a platform for integrating existing time-series images. Specifically, the functionalities of ArcCI web service include image data management, user management, batch image processing, results review and spatiotemporal visualization modules.

Specifically, the core database and the ETL data processing workflow is specifically designed for handling the various big data challenge to extract metadata of imagery collections from heterogenous data sources with different combination of data owner, organization, and scientific research projects. To address the big data volume and velocity challenge of processing demands with the discrete, high scale and resolution feature, a three-layer framework is designed and implemented as a practice. The framework includes cloud computing techniques, big data components and open source-based packages. The ArcCI web services 1) leverage the essential cloud computing resources, such as virtual machine (VM), distributed storage/file system, and networking; 2) incorporate common geoscience information service software and programming tools to enable user interface customization by the arctic science applications. The service software and tools will be integrated with operating system, delivered as cloud VM images for benefiting other polar CI and polar science projects; and 3) address the service integration and interoperability issues by hiding the underlying computing infrastructure across different organizations and platforms.

CHAPTER 4. SPATIOTEMPORAL ANALYSIS OF SEA ICE LEADS IN THE ARCTIC OCEAN RETRIEVED FROM ICEBRIDGE LAXON LINE DATA 2012-2018

4.1 Introduction

Arctic sea ice functions as a sensitive indicator of global warming, since it tends to be affected by a small temperature increase. On the other hand, Arctic sea ice is also an important driver of climate change, and it plays an important role in the Earth's solar radiation budget. This is due to the fact that sea ice has a significantly higher albedo compared to that of the water surface. Therefore, when the Arctic sea ice starts to melt, the oceans will absorb more solar radiation and heat up, thus accelerating the sea ice melting in a positive feedback way.

Among all types of sea ice features, leads are striking features with unique characteristics. A lead is an elongated crack in the sea ice developed by the diverging or shearing of floating ice floes when they move with currents and wind (Q. Wang et al., 2016). Leads vary in width from meters to hundreds of meters depending on their development and directions of surrounding pressure and tension. Since a lead is an open body of water within a sea ice floe, it allows for the direct interaction between the atmosphere and the ocean, which is important for Arctic sea ice ecology and local radiation energy budget. Specifically, it dominates the vertical exchange of energy, particularly in winter when turbulent heat fluxes over leads can be orders of magnitude larger than that over thick ice. The width of leads and their orientation markedly influence associated vertical sensible and latent heat fluxes as well as associated cloud formation (Hakkinen et al., 2008; Hirano et al., 2016). Recent studies suggest that these fluxes could influence the atmospheric properties tens to hundreds of kilometers downstream (Alam, 1997; Andreas & Murphy, 1986; Marcq & Weiss, 2012). Even a small fraction of thin ice and open water within the sea ice pack can significantly modify the total energy transfer between the ocean and the atmosphere (Worby & Allison, 1991). Furthermore, leads are elusive and inconsistent features. If the air temperature reduces, the water within a lead quickly refreezes and leads will be partly or entirely covered by a thin layer of new ice. Therefore, as an important component in the Arctic surface energy budget, a more quantitative study is necessary to explore and model the leads' impact on the Arctic climate system.

This study was motivated by the requirement of the spatiotemporal analysis of sea ice lead distribution through NASA's Operation IceBridge (OIB) mission, which used a systematic sampling scheme to collect high spatial resolution DMS aerial photos along critical flight lines in the Arctic region. We developed a practical workflow to classify the DMS images along the Laxon Line into four classes, i.e., thick ice, thin ice, water, and shadow, and extract sea ice lead and thin ice during 2012-2018. Finally, the spatiotemporal variations of leads along the Laxon Line are verified by ATM surface height data surface height data (freeboard) and correlated with sea ice motion as well as atmosphere temperature and wind data.

This chapter is organized as follows: Section 4.2 provides a background description of DMS imagery, the Laxon Line collection, and auxiliary sea ice data.

Section 4.3 describes the methodology and workflow. Section 4.4 presents and discusses the spatiotemporal variations of leads.

4.2 Dataset and Study Area

4.2.1 IceBridge DMS images and study area

In this study, the IceBridge DMS images were used to detect Arctic sea ice leads along the Laxon Line in 2012-2018. DMS images were collected during IceBridge sea-ice flights using an airborne digital camera. DMS has a high spatial resolution of 0.1- 2.5 m (Dominguez, 2010) depending on the aircraft flight height. It has three natural RGB (Red, Green, and Blue) bands and each image has a field of view of approximately 400 m by 600 m. The IceBridge campaigns have been designed to survey the Arctic region in March and April since 2009, so as to partially fill the temporal gap between ICESat (2003 - 2009) and ICESat-2 program starting in 2018.

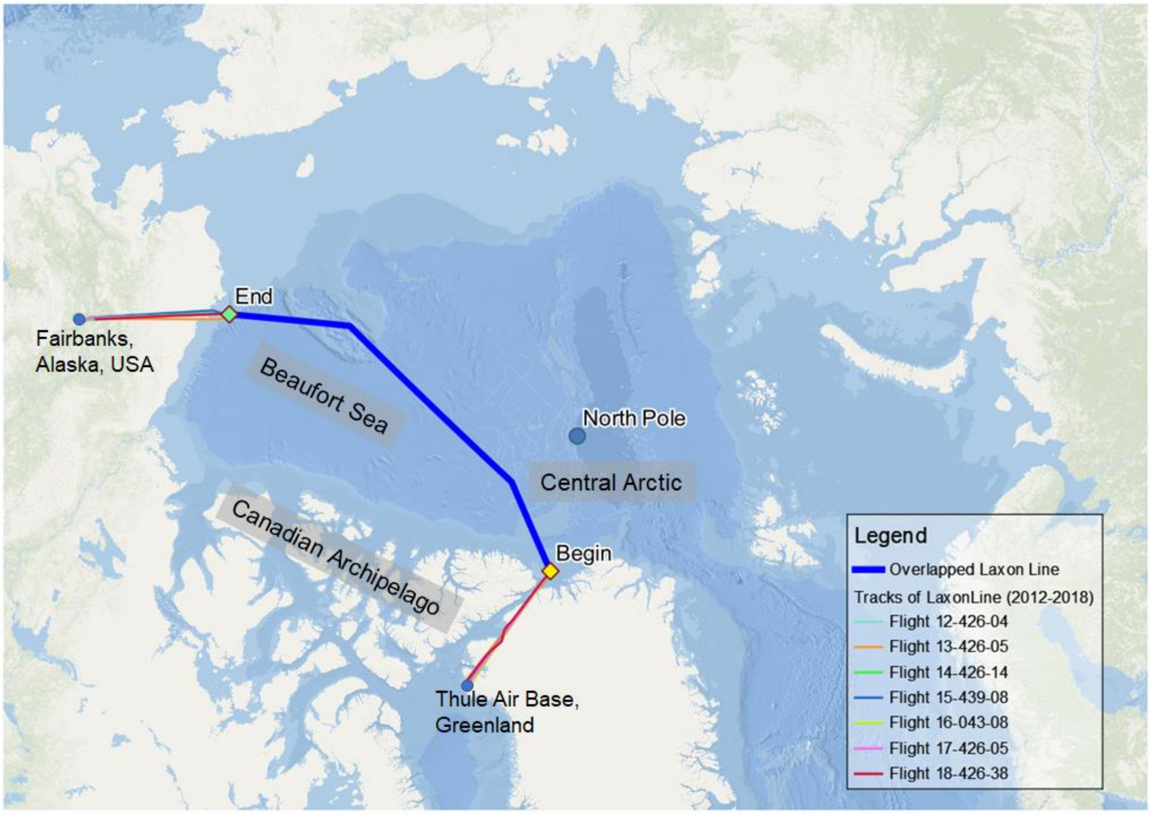


Figure 14. Spatial distribution of 7 tracks of the Laxon line from 2012-2018. The tracks are highly overlapped.

DMS images are collected, processed, and maintained by the Airborne Sensor Facility located at the NASA AMES Research Center. We downloaded the Level 1B geolocated and orthorectified images for the Arctic Laxon Line in the Spring from 2012 to 2018 from the NASA National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC). The Laxon line starts from Thule Air Base, Greenland and terminates at Fairbanks, Alaska, USA, transiting across the Arctic Ocean (Figure 14). It passes through both multi-year ice (MYI) regions in the north of the Canadian Archipelago and the first-year ice (FYI) regions in northern Alaska. Thus, sea ice data along this line provides useful insight on the transition of sea ice conditions over the

Central Arctic in Spring. Furthermore, the IceBridge mission collected data along this track repeatedly every year from 2012-2018, which is appropriate for spatiotemporal analysis of sea ice leads. The overall DMS image collection along the Laxon line is 106,674 aerial photos (1.54 TB) with an endlap of 60- 90%. The photo distribution from 2012-2018 is summarized in Table 4. The overall distance of the Laxon line is around 3398 km and the distance for the overlapped track through the years is around 2437.2 km.

Table 4. The DMS images selected for lead detection in the Laxon Line from 2012 to 2018.

Name	Date	Image Number	Sea ice leads including image Number	Selected / Original Image Size (GB)	Lighting Condition
Flight 12-426-04	3/14/2012	16544	1066	14.8/260	Cloudy
Flight 13-426-05	3/21/2013	18480	993	13.8/290	Normal
Flight 14-426-14	3/14/2014	14322	492	5.2/150	Cloudy
Flight 15-439-08	3/26/2015	20038	816	9.3/250	Normal
Flight 16-043-08	4/20/2016	15205	1069	18.4/270	Normal
Flight 17-426-05	3/10/2017	10939	659	8.67/93	Cloudy
Flight 18-426-38	4/6/2018	11146	1040	22.2/240	Normal

4.2.2 Auxiliary Sea Ice Data

4.2.2.1 AMSR data. AMSR (Advanced Microwave Scanning Radiometer) is a passive microwave satellite sensor launched by the Japan Aerospace Exploration Agency. Due to its low spatial resolution, AMSR data can only be used to examine sea ice concentration at a regional scale. We collected AMSR-E/AMSR-2 Unified Level 3 daily brightness temperature and sea ice concentration data with a spatial resolution of 12.5 km through NSIDC (W N Meier et al., 2018). The data contains vertical and horizontal brightness temperature (T_B) at four frequency channels: 18.7 GHz, 23.8 GHz, 36.5 GHz, and 89.0 GHz. The Arctic sea ice concentration (SIC) product is calculated by the NASA Team 2 (NT2) algorithm (W N Meier et al., 2018). The collected AMSR data coincide with days of the IceBridge missions from 2012 to 2018, so the SIC can be compared with that retrieved from the DMS images. Furthermore, the passive microwave data can be used to calculate thin ice concentration (TIC). Röhrs and Kaleschke (2012) used T_B at the vertically polarized 18.7 GHz and 89.0 GHz to identify water and thin ice (i.e., new ice, nilas, and pancake ice) from thick ice, and the sea ice leads and TIC agree with the MODIS, Envisat ASAR, and CryoSat-2 data. In this study, we also calculated TIC following Röhrs and Kaleschke's algorithm (Röhrs & Kaleschke, 2012). The coarser spatial resolution of 25 km of TIC were compared with the lead and thin ice fractions retrieved from the DMS images.

4.2.2.2 ATM surface height data (DMS level). Our DMS-based lead detection results can be used to cross-validate sea ice freeboard products derived from IceBridge Airborne Topographic Mapper (ATM) Level 1B data (Studinger, 2013). The ATM is an airborne conically-scanning laser altimeter with the wavelength of 532 nm. A laser pulse is emitted from the ATM at a rate of 5 kHz, and it has ~1 m of footprint size at a typical 500 m altitude above the surface. Since it covers exactly the same location and time with the DMS images with a smaller cross-track width (~400 m), DMS images are usually used as good reference data for the ATM-based lead detection studies (Kurtz et al., 2013; Tian et al., 2020; X. Wang et al., 2016). In this study, the ATM data are resampled in a 2 m grid and projected to the same projection system as DMS (NSIDC sea ice polar stereographic North) to match the geographical location. After retrieving thin ice and leads through DMS images, we geographically linked the leads with the ATM data to extract freeboard variations along the Laxon line, and cross verified with freeboard data derived from other algorithms.

4.2.3 Geophysical Parameter of Ocean and Atmosphere

NSIDC provides sea ice motion data with a spatial resolution of 25 km on the Equal-Area Scalable Earth grid (Tschudi et al., 2019). This sea ice motion vector is derived from multiple data sources including AVHRR, AMSR-E, Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave/Imager (SSMI), and Special Sensor Microwave Imager/Sounder (SSMI/S satellite sensors), International Arctic Buoy Program (IABP) buoys, and National Center for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) Reanalysis forecasts.

We also acquired a global sea ice type product provided by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI SAF, www.osi-saf.org). This product assigns different sea ice types such as multi-year ice (MYI), first-year ice (FYI), and open water from various satellite data sources. This is a daily product and has a spatial resolution of 10 km.

Another data we used include air temperature (2m above sea surface) and wind velocity data coincident with the DMS images acquired from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis model. The ERA5 product has 0.25° spatial resolution and consists of hourly variables and we integrated this hourly data into daily products, and resampled to 25 km resolution to match the ice motion data. This product was downloaded from the Climate Data Store (cds.climate.copernicus.eu) of the Copernicus Climate Change Service. In this study, the high spatial resolution lead fractions derived from DMS along the Laxon line are correlated with the coarse spatial resolution sea ice motion, air temperature, and wind velocity products.

4.3 Methodology

4.3.1 Object Based Image Analysis for Image Classification

Most of the high-resolution sea ice photos were analyzed through pixel-based methods (Lu et al., 2010b; Renner et al., 2013; Weissling et al., 2009). This method is based on pixel brightness values or spectral values, ignoring spatial autocorrelation and generating 'salt-and-pepper' noise in the classification (Liu & Xia, 2010; Xie et al., 2007). In contrast, object-based classification methodologies have been developed based on image segmentation, the process of partitioning an image into multiple objects or groups of pixels, thus making it more meaningful and easier to analyze (Hussain et al., 2013; Shapiro & Rosenfeld, 1992). This method not only considers spectral values but also spatial measurements that characterize the shape, texture, and contextual properties of the region so as to potentially improve classification accuracy (Liu & Xia, 2010). A flexible classification scheme is the key to multitasking polar applications. We have defined a suitable classification scheme for high spatial resolution multi-band photos which can be found in Table 5. Three major steps of this algorithm (Figure 15) are listed as follows.

Table 5. Classification scheme for object-based classification of sea ice photos.

#	Class Name	Class Description
1	Thick Ice (Ice/snow)	Bright white objects due to high reflectance of ice/snow.
2	Thin ice	Thinner ice above water along the edge, usually shown as a darker color compared to pure snow-covered thick ice due to the mixed reflection from ice surfaces and water. Thin ice and open water will be merged as sea ice lead definitions for calculation of ice lead fraction.
3	Shadow	Darker objects on the ice/snow caused by ridges and low solar elevation angles. Shadow is usually on ice/snow and can be combined into Ice/snow for calculation of ice concentration. In some cases, however, shadow could also be on ponds that often adjacent to ridges. Therefore, further treatment for shadow on ice and ponds are needed. Shadow will also be used for calculation of ridge height.
4	Water	Arctic ocean, objects are rather dark and smooth, formed by sea ice leads in the spring season of the Arctic.

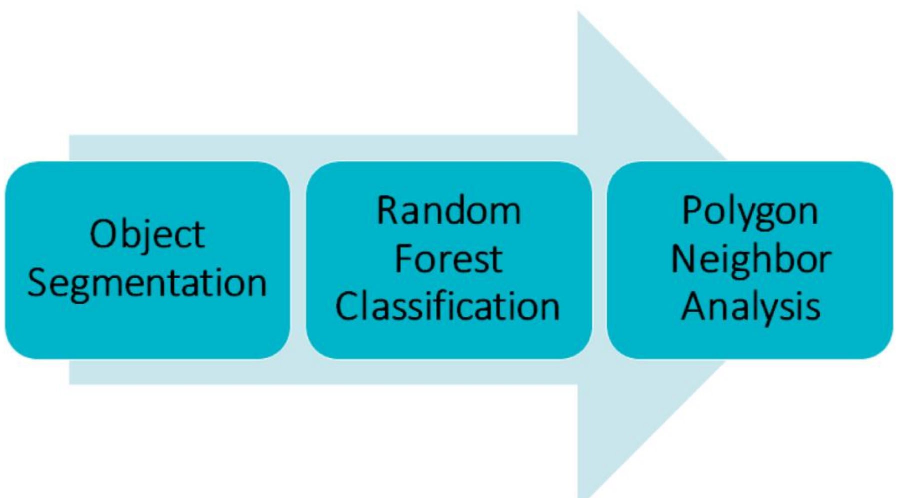


Figure 15. Object-based Image Analysis (OBIA) workflow for sea ice classification.

4.3.1.1 Object-based image segmentation. The watershed segmentation algorithm is chosen for sea ice HSR images, followed by object merging through Region Adjacency Graphs (RAG). The segmentation algorithm iteratively merges adjacent segments based on a combination of spectral and spatial information.

4.3.1.2 Random forest classification. The outputs from the image segmentation above are individual objects or polygons. Spectral, shape, and neighboring features of each object can then be derived (Table 6) for each object and imported into a random forest classifier for object-based classification. The random forest classifier is essentially a variant of the bagging tree ensemble classifier (Breiman, 2001; Breiman et al., 1984) and operates through randomly selecting a subset of input features for each decision split. This way, classification accuracy and feature importance can be evaluated by out-of-bag (OOB) estimations. This method is suitable for small sample problems such as object-based classification and cloud-based multi-core parallel computing.

Table 6. Definition of spectral and spatial features.

Spectral	Mean average BV (B_a)	$B_a = \sum BV / n$, where n is the number of pixels within an object boundary/polygon area, BV is the pixel brightness value in the object. Red, green and blue bands are calculated and utilized here.
	Standard Deviation BV (B_a)	
Spectral Index	Two band-based Index	$(B_{blue} - B_{red})/(B_{blue} + B_{red})$
		$(B_{blue} - B_{green})/(B_{blue} + B_{green})$
	Three band-based Index	$(B_{green} - B_{red})/(2 \times B_{blue} - B_{red} - B_{green})$
Shape	Segment Size	The overall pixel count in the segmented object.
	Entropy (T_e)	$T_e = -\sum p \cdot \log_2(p)$, where p is the frequency of histogram counts.

4.3.1.3 Polygon neighbor analysis. A major challenge for the sea ice leads application is that the spectral characteristics of thin ice is very similar compared to parts of thick ice-covered areas due to an illumination issue: shadows resulting from sunlight results in misclassification cases from confusion. We can use polygon neighbor analysis to separate shadows from thick ice types first. Then, in the final geophysical calculation, the shadow class can be merged into the thick ice class to derive total sea ice concentration. Additionally, in this classification scheme, thin ice and open water would be merged for sea ice lead fraction calculation in each classified image scene as needed.

4.3.2 Practical Batch Classification Processing Workflow

Since IceBridge DMS images are highly overlapped along the track (60-90%), we selected one image from every three consecutive images in the Laxon line to reduce computational burden. All images in continental land masses and poor-quality images due to overwhelming cloud coverage and low lighting conditions were manually removed (as shown in Figure 16). All images with sea ice leads were manually selected and saved as a collection of sea ice lead images (Table 4).

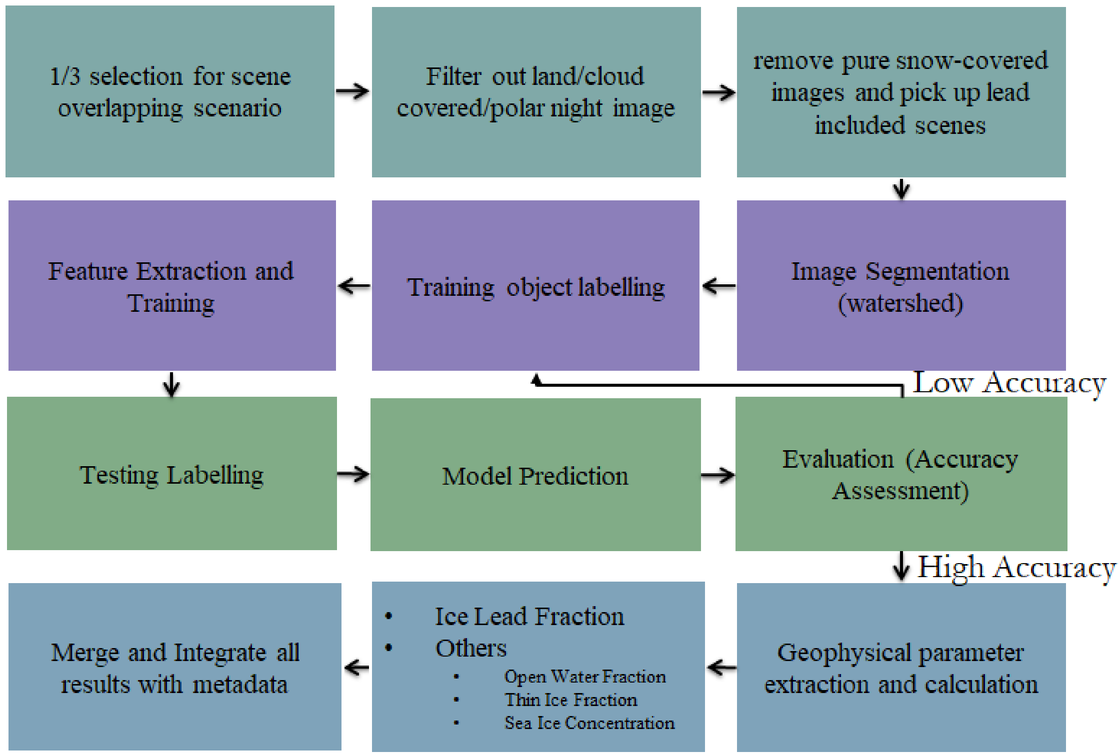


Figure 16. Practical sea ice lead feature extraction workflow.

The object-based classification scheme is designed based on the color and texture of sea ice features in DMS images. Four classes are defined: (1) *thick ice* is usually thick ice or snow-covered ice with a high albedo; (2) *thin ice* is usually fresh and newly-formed thin ice, which has a smooth surface with a low albedo due to solar radiation being partially absorbed by the water beneath it; (3) *open water* are bodies of dark water due to its strong absorbance of solar radiation; and (4) *shadow* is located within thick ice areas and are relatively dark features projected onto the ice surface by surrounding ridges or snow dunes. Each year of DMS images have different lighting conditions which affects image quality (Table 4). Furthermore, even in the same year, the quality of images is quite distinctive due to local cloud coverage and lighting conditions as shown in Figure 17. Normal images contain regular sea ice scenes with an appropriate exposure and contrast, and all sea ice classes are recognizable by color and texture. Gray images are partially cloudy images with poor lighting conditions, so they are relatively dark and shadows are difficult to detect. Dark images have extremely poor lighting conditions, and the boundaries between water, thick ice, and thin ice are blurred due to low contrast.

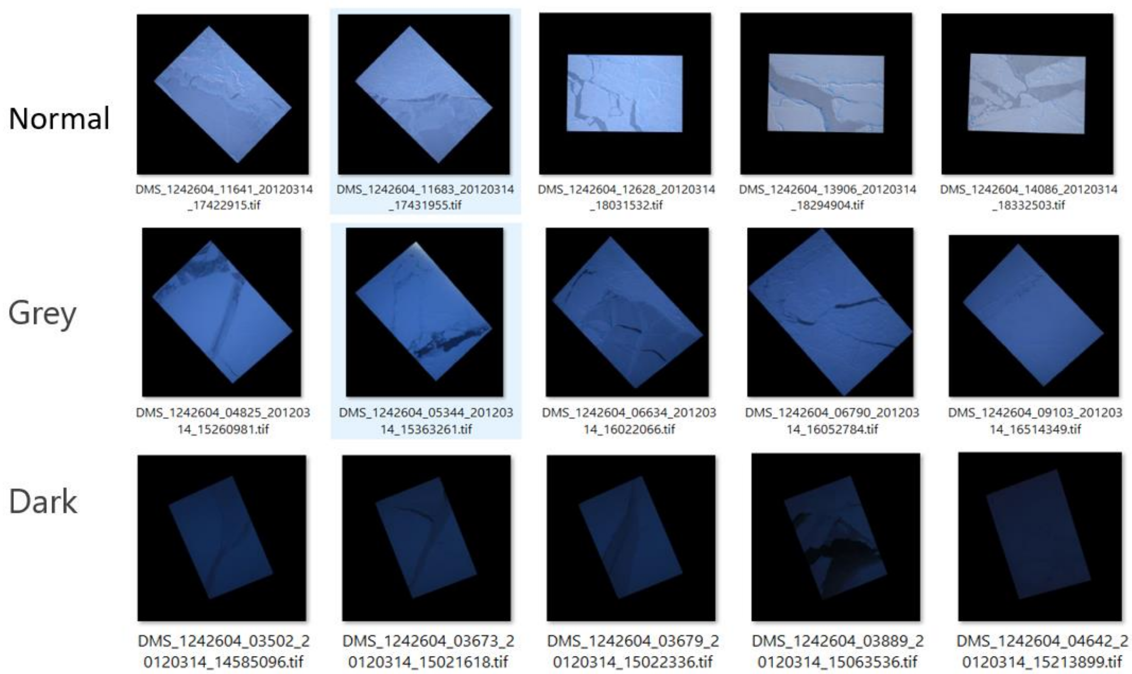


Figure 17. DMS sea ice sample images in 2012 are classified into three subgroups based on different lighting conditions.

Therefore, training samples were selected using a divide-and-conquer strategy based on image quality. All DMS images taken in 2013, 2015, 2016, and 2018 were taken under good lighting conditions and training samples were selected for all four sea ice features. On the other hand, the images taken in the other three years were processed in different ways. The training samples for all images taken in 2012, 2014, and 2017 were only selected for thin ice, open water, and thick ice without considering shadow due to low lighting conditions. Furthermore, the 2012 images were manually classified into three subgroups, i.e., normal, grey, and dark. The 2014 images were manually classified into two subgroups, i.e., normal and grey, and all dark images were abandoned due to serious vignetting caused by light hitting the lens aperture at a large angle resulting in brightness values on the four corners of the image being significantly reduced. The 2017 images were all classified into the grey subgroup. In summary, the independent training samples were collected for each subgroup and year for supervised classification.

The OSSP package uses an object-based classification scheme. For each image, the watershed segmentation method was used to convert pixels into objects. Therefore, training samples were labelled at the object level. Only distinctive and typical sea ice objects were selected across the whole scene, and each sea ice class has around 120-250 objects. The attributes of objects such as color values, band ratios, textures, and shape indexes were calculated and served as supervised classification features. Based on these training datasets, the OSSP package uses the random forest classification method to label all unknown objects in DMS images (Miao et al., 2015).

In order to evaluate the accuracy of the classification results, independent test object samples were also collected to verify the results. At least 10 test objects for each class were selected for each single image. Finally, the confusion matrix was generated at pixel level, so as to calculate the overall accuracy, user’s accuracy, producer’s accuracy, and Kappa Coefficient.

4.3.3 Sea Ice Leads Parameters Definitions

Based on the classified result in each surface type, we derive the sea ice leads by combining thin ice and open water. Then the sea ice lead fraction, open water fraction, thin ice fraction and sea ice concentration are calculated on a per-scene basis. The sea ice lead fraction for each DMS image can be calculated using the following equations:

Equation 1 Sea ice lead fraction

$$SILF = \frac{ThinIce + OpenWater}{\text{TotalPixel}} \times 100$$

where ThinIce, OpenWater and TotalPixel is pixel numbers of classified thin ice area, open water area and the total area for a DMS image.

4.3.4 Spatiotemporal Analysis with Environmental Factors

The auxiliary sea ice data sets can be used to assess and compare the DMS-based lead detection results, so as to deepen the understanding of the formation mechanism of leads. In this research, sea ice freeboard was derived from our lead detection algorithm, and compared to the NSIDC freeboard data at the scale of 400 m. Furthermore, the coincident AMSR TIC data, and the geophysical atmosphere and ocean data, such as temperature, wind velocity, and sea ice motion, were collected to compare to the lead fraction results.

Sea ice freeboard data from the ATM lidar data based on our DMS lead detection algorithm were retrieved, as suggested by Kurtz et al. (2013). We calculate the reference sea surface heights and calculate sea ice freeboards by using the ATM lidar data based on our DMS lead detection results. Specifically, we remove variations in the instantaneous sea surface height by subtracting geoid and ocean tide height. Then we calculate freeboard by subtracting locally determined leads surface height (Z_{ssh}) from the corrected height (H_{corr}).

Equation 2 Freeboard

$$Freeboard = H_{corr} - Z_{ssh}$$

where Z_{ssh} is determined from the sets of individual lead elevation estimates through ordinary kriging. We calculated the mean freeboard for each DMS image (around 400 m by 600 m) and resampled the value to 400m resolution. On the other hand, Kurtz et al. (2013) used an automated lead detection algorithm called SILDAMS (Sea Ice Lead Detection Algorithm) which employs a minimal signal transform (Kurtz et al., 2013; Onana et al., 2013), and then retrieved the freeboard at the resolution of 400 m. Therefore, the two products can be compared and cross-verified at this scale.

TIC can be calculated from AMSR as described in Röhrs and Kaleschke (2013). This AMSR-based TIC represents the existence of open water and thin ice on sea ice leads with a rather coarse spatial resolution of 25 km. This AMSR-based TIC represents the existence of open water and thin ice on sea ice leads, which is conceptually equivalent to SILF. Since the AMSR and DMS have different resolutions and geographical coverages, they cannot be compared directly. Therefore, we resample the DMS-based ice lead fractions for every 25 km to match the spatial resolution of AMSR data, as shown in Figure 18. Then the mean of sea ice lead fractions within the range of each 25 km block were calculated.

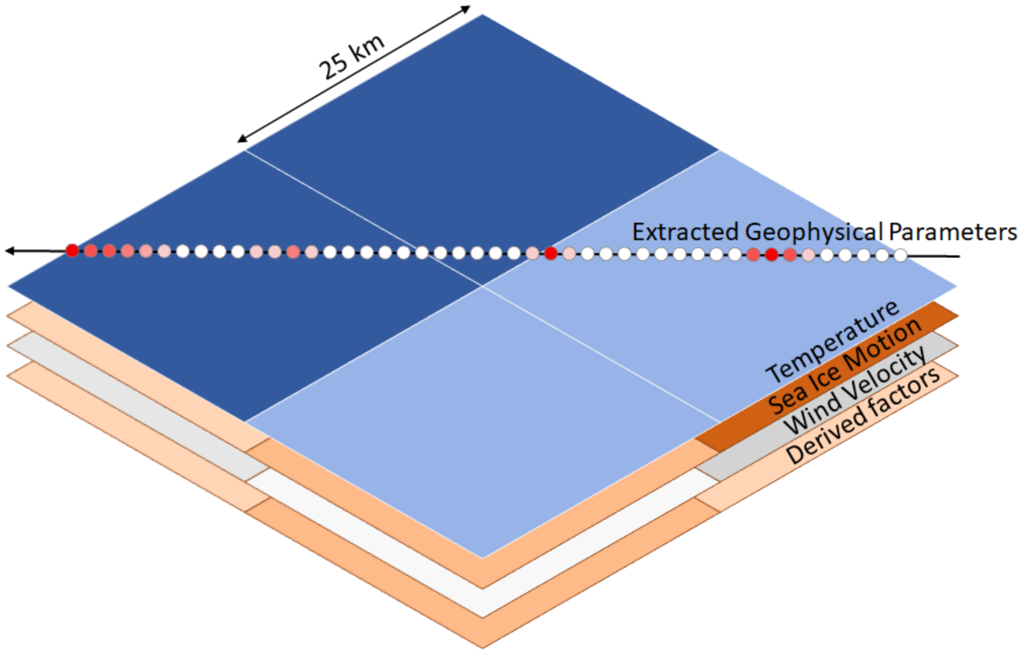


Figure 18. Data fusion diagram with derived geophysical parameters and DMS-based sea ice leads. (Each 25km AMSR pixel covers around 5-70 point of HSR image locations).

Furthermore, the 25-km resampled lead fractions were also correlated with other 25 km resolution sea ice and atmospheric data including NSIDC sea ice motion, ERA5 air temperature, and wind velocity. Since kinetic moments of sea ice movement can play an important role in formations of leads, four kinetic moments or tensions were calculated from the NSIDC sea ice motion data by the following equations (Molinari & Kirwan, 1975):

Equation 3 Divergency

$$divergency = \frac{dFx}{dx} + \frac{dFy}{dy}$$

Equation 4 Vorticity

$$vorticity = \frac{dFy}{dx} - \frac{dFx}{dy}$$

Equation 5 Shearing deformation

$$shearing\ deformation = \frac{dFy}{dx} + \frac{dFx}{dy}$$

Equation 6 Stretching deformation

$$stretching\ deformation = \frac{dFx}{dx} - \frac{dFy}{dy}$$

Finally, the average of these dynamic and thermodynamic variables up to 30 successive days before the DMS acquisition day were calculated, with the purpose of identifying the optimal temporal scale of the contribution of these external parameters to lead formation. All derived and aggregated variables are summarized in

Table 7.

Table 7. Variable description for the multiple linear regression.

Department	Factors	Description
Sea Ice Leads	mean_leads	Mean lead fraction for 25 km segment (only includes DMS images that contain leads)
Temperature	tmpXX	Averaged air temperature for last XX days (e.g. tmp03 means average temperature of last 1, 2, 3 days)
Wind	U10_XX	Averaged u-component of wind velocity for last XX days
	V10_XX	Averaged v-component of wind velocity for last XX days
	wind_XX	Averaged wind velocity for last XX days (e.g. wind_10 means wind velocity for last 10 days)
Sea Ice Motion	u_ice_XX	Averaged u-component of ice velocity for last XX days (e.g. u_ice_10 means u-velocity for last 10 days)
	v_ice_XX	Averaged v-component of ice velocity for last XX days (e.g. v_ice_10 means v-velocity for last 10 days)

	vel_ice_XX	Averaged ice velocity for last XX days (e.g. v_ice_10 means ice velocity for last 10 days)
	divXX	Averaged divergence of sea ice motion for last XX days (e.g. div10 means divergence for last 10 days)
	vorXX	Averaged vorticity of sea ice motion for last XX days (e.g. vor10 means vorticity for last 10 days)
	shrXX	Averaged shearing deformation of sea ice motion for last XX days (e.g. shr10 means shearing deformation for last 10 days)
	stcXX	Averaged stretching deformation of sea ice motion for last XX days (e.g. stc10 means stretching deformation for last 10 days)

The multiple linear regression (MLR) was used for modelling the mean lead fractions in terms of large-scale sea ice dynamic–thermodynamic variables, including the NSIDC sea ice motion data with four kinetic moments, ERA5 air temperature, and wind velocity data. The forward and backward stepwise regression method were used to identify the most important explanatory variables. This strategy refers to the process of building a regression model by adding or removing explanatory variables in a stepwise manner until the predicted variable will not change significantly (Wilkinson, 1979).

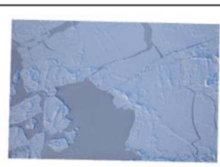
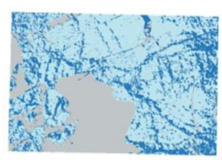
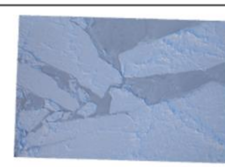
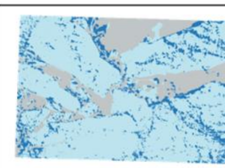
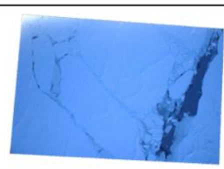
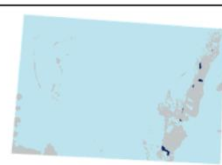
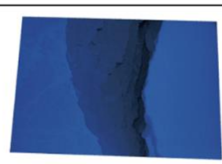
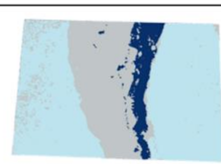
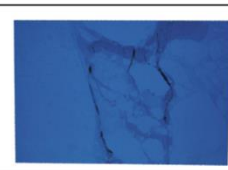
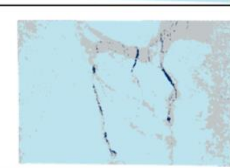
4.4 Results and Discussion

4.4.1 DMS Imagery Classification Result

A total of 106674 DSM images along the Laxon line from 2012-2018 were processed, and over 6135 images with sea ice leads were visually selected (Table 4). All images were classified through the OSSP package embedded in the ArcCI online service (Sha et al., 2020).

Some examples of the classified images taken in 2012 are shown in Table 8. The first row shows the classification results for the subgroup of normal images, the second row for the grey images, and the third row for the dark images. All 6 images were selected to show a variety of sea ice features under different lighting conditions. The classified results illustrate four sea ice classes: open water, shadow, thin ice, and thick ice.

Table 8. Comparison of original 2012 DMS images and classified results for three subgroups. Two samples are selected for each subgroup.

	Sample Result 1		Sample Result 2	
	Raw Image	Classified Result	Raw Image	Classified Result
Normal				
Grey				
Dark				

LEGEND Thick Ice Thin Ice Shadow Open Water

Table 9. Pixel-level Classification accuracy for each production group. All values except kappa are in percentages. User's accuracy and producer's accuracy for each classified ice type represented as UA_XX, and PA_XX, and XX could be thick ice, thin ice, shadow, or open water.

Testing Group	Overall Accuracy	Kappa Value	UA_Thick	UA_Thin	UA_Shadow	UA_Water	PA_Thick	PA_Thin	PA_Shadow	PA_Water
DMS2012_normal	88.9	0.83	88.0	91.7	83.8	nan	98.4	94.2	63.8	nan
DMS2012_grey	93.6	0.85	97.3	85.0	nan	95.5	93.8	93.1	nan	97.5
DMS2012_dark	93.8	0.86	95.0	96.0	nan	61.9	98.9	81.2	nan	94.9
DMS2013	96.4	0.95	92.2	100.0	99.4	95.5	99.7	96.5	88.3	99.9
DMS2014_normal	88.0	0.82	74.7	86.2	93.9	98.0	97.1	81.3	99.7	89.0
DMS2014_grey	93.7	0.89	91.7	96.3	nan	97.1	100.0	75.7	nan	97.1
DMS2015	86.4	0.78	86.6	83.5	98.6	93.4	99.8	80.9	82.2	57.9
DMS2016	87.9	0.83	82.1	89.3	95.0	95.7	99.4	68.8	89.7	90.2
DMS2017	86.7	0.75	87.4	82.8	nan	99.4	97.6	76.5	nan	60.7
DMS2018	93.5	0.88	91.9	96.5	95.2	97.9	98.5	79.1	89.4	98.4
Average Accuracy	90.9	0.84	88.7	90.7	94.3	92.7	98.3	82.7	85.5	87.3

*The nan value refers to this group not including the specific classification sea ice type.

The classification accuracies were evaluated at the pixel-level and all calculated accuracies are summarized in Table 9. The overall accuracy across the 10 test samples selected by year and illumination conditions is 90.9±3.5%, where the latter number is the standard deviation, and the kappa Coefficient is 0.85±0.05 across 10 sets of test samples selected by year and illumination conditions. Since the sea ice leads are defined as a combination of thin ice and open water, the classification accuracy is determined by these two classes. The user's accuracy for thin ice and water are 90.7±5.9% and 92.7±11.0%, respectively. The low accuracy of 61.9% for open water in the 2012 dark subgroup is due to the confusion between water and thin ice under extremely poor lighting conditions.

4.4.2 Overall Integrated Statistical Analysis and Variations Trend of Sea Ice Leads

4.4.2.1 Sea ice leads fraction, area, and frequency.

Figure 19(a) shows the averaged lead fraction for every 25 km along the Laxon Line. Relatively large lead fractions (> 15 %) are only observed near the Beaufort Sea area (track distance > 1200 km) in 2013, 2014, and 2016, where they are generally located in FYI region or transition region between FYI and MYI. On the other hand, the smaller lead fraction region in the central Arctic (track distance < 1,200 km) is primarily covered by MYI and thick ice.

Figure 19(b) portrays the averaged area of individual leads for the 25 km track segment, and Figure 19(c) portrays the ratio of the number of lead-included images to the total number of images for the 25 km segment. The lead fraction (Figure 19(a)) is determined by the individual lead area (Figure 19(b)) and the frequency of leads (Figure 19(c)). For example, although large leads are observed in 2013 for 0-500 km (Figure

19(b)), lead frequency for this part is low (Figure 19(c)) (i.e., small number of large leads). As a result, the averaged lead fraction for this segment is not high because of the low lead frequency. In addition, the lead frequency in 2018 for 1000-2500 km is relatively high, but their averaged lead fraction is not so high because of the small lead area (i.e., large number of small leads).

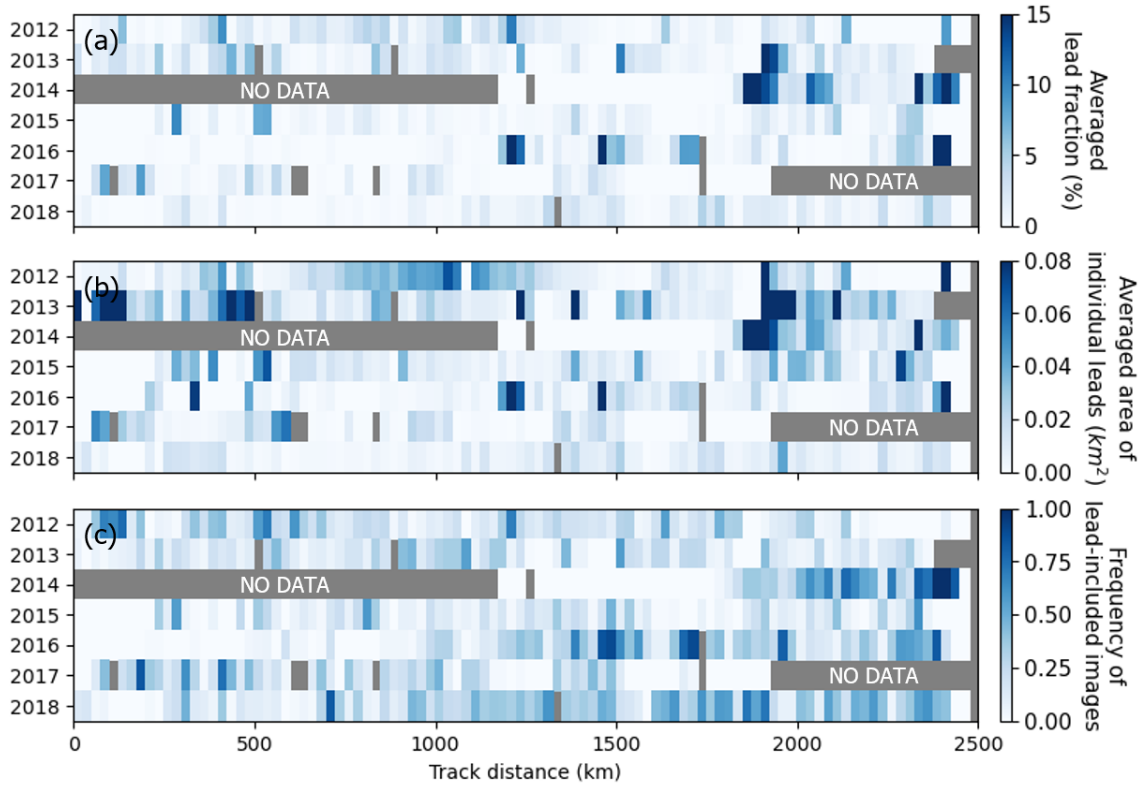


Figure 19. (a) Averaged lead fraction for every 25 km; (b) Area of detected leads for every 25 km; (c) Lead frequency for every 25 km. White parts indicate missing data.

4.4.2.2 Retrieval of freeboard.

Based on the DMS lead detection result, we calculate 400-m mean sea ice freeboard from the ATM surface height data (Figure 20). The MYI area (near central

Arctic Ocean) at track distance < 1,200 km shows higher freeboard (i.e. thicker ice) compared to the FYI area (near the Beaufort Sea with a track distance larger than 1,200 km). As shown in Table 10, the FYI area always shows lower freeboard than the MYI area. In addition, the freeboard retrieved from our lead detection shows a good correlation with the ATM freeboard product provided by NSIDC (Kurtz et al., 2013): ~0.832 of correlation coefficient (R) and ~0.105 m of root mean square difference (RMSD) (Table 11). It is also noted that 2015, 2016, and 2017 show relatively lower R and higher RMSE than the other years (Table 11 and Figure 21), which might be associated with the lower classification accuracy of these years (Table 9). Some misclassified leads can make substantial differences in estimation of sea surface height, eventually leading to the differences between our freeboard estimation and the NSIDC freeboard products. Nevertheless, the freeboard differences between MYI and FYI and the cross-validation with the NSIDC freeboard product show that our lead detection result is reasonable and compatible with other lead detection products.

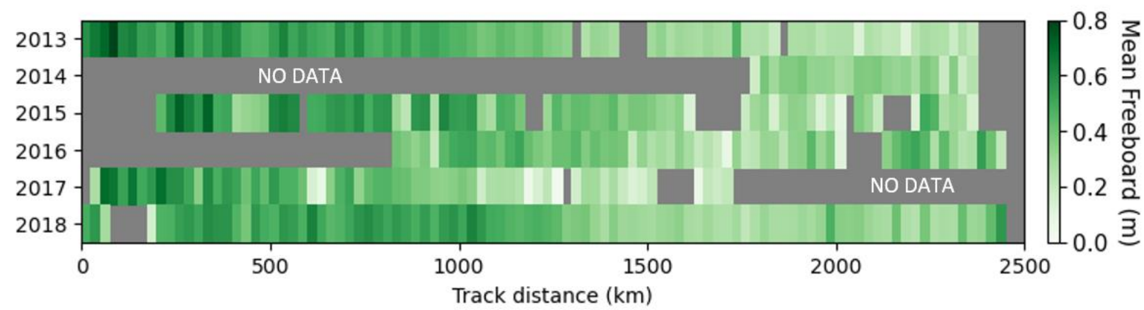


Figure 20. Averaged ATM freeboard for every 25 km for each year.

Table 10. ATM sea ice freeboard retrieved from the DMS lead detection.

Year	FYI	MYI	Total
2013	0.263 m	0.519 m	0.409 m
2014	0.277 m	0.339 m	0.320 m
2015	0.275 m	0.470 m	0.407 m
2016	0.335 m	0.398 m	0.354 m
2017	0.211 m	0.467 m	0.366 m
2018	0.320 m	0.505 m	0.414 m

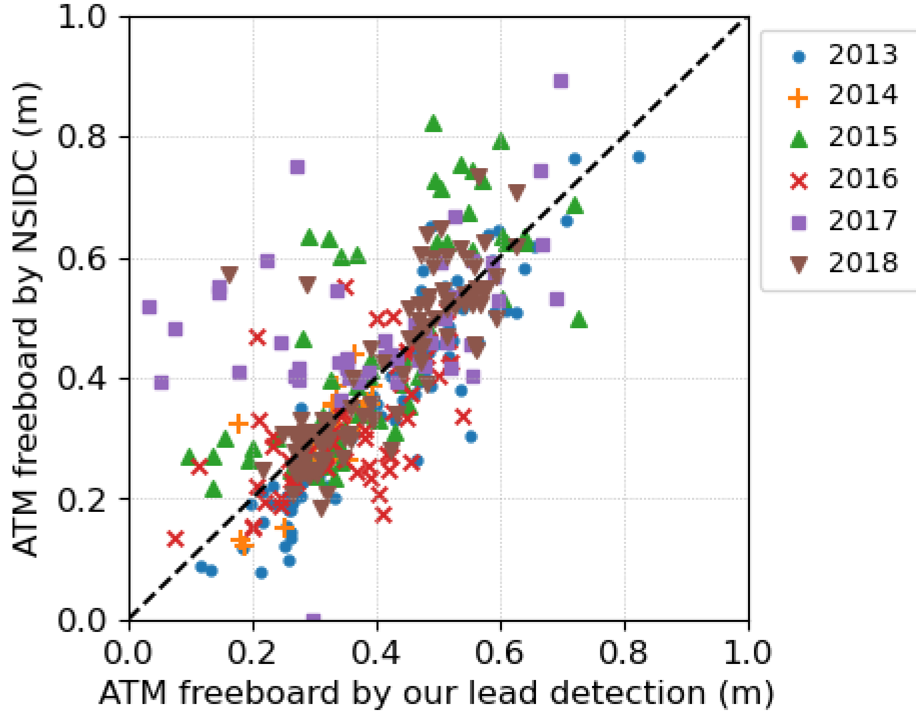


Figure 21. Scatter plot between ATM freeboard derived by our lead detection and NSIDC freeboard product for every 400 meters (2% random selection of the total points).

Table 11. R and RMSE between our freeboard estimation and NSIDC freeboard estimation.

Year	R	RMSD (m)
2013	0.928	0.089
2014	0.907	0.063
2015	0.755	0.140
2016	0.784	0.114
2017	0.742	0.119
2018	0.869	0.082
Total	0.832	0.105

4.4.3 Comparison with Auxiliary Sea Ice Products

Since the OIB missions were conducted during the end of the freezing season (March to April), the widths of individual leads were usually less than 1 km. Indeed, as shown in Figure 6b, most leads have less than 0.1 km2 of area, which account for a tiny portion of the entire 25 km x 25 km grid cells. Hence, it is reasonable that the DMS-based lead detection and AMSR-based TIC are not highly correlated (R ~ 0.21) because narrow leads are hardly detected by the coarse resolution satellite data. For example, we find that most of AMSR-based TIC along the track is zero but AMSR-based SIC is 100 % even though the DMS images have leads in that area.

Figure 22 shows the lead fractions and possible dynamic and thermodynamic variables at the scale of 25 km on the days that DMS images were taken from 2012-2018: (a) DMS-based lead fraction and nearby ice type; (b) ERA5 air temperature; (c) ERA5 wind velocity; (d) sea ice motion for each year. In general, the lead fractions do not show significantly correlation with any single auxiliary variable or kinetic properties from sea ice motion data. It is reasonable because (1) these ancillary data have 25-km spatial resolution, which is much coarser than the spatial resolution of DMS image; (2) the DMS images have only ~500 m of width, only representing a small portion along the Laxon Line; and (3) the accumulative effects of these dynamic and thermodynamic variables on the forming of ice leads are not considered.

In order to explore the relationship between the lead fractions and large scale sea ice dynamic–thermodynamic variables, we constructed a series of multiple variables linear regression models. The lead fraction variable is the mean of all DMS image-based lead fractions within a 25 km-block. On the other hand, all dynamic-thermodynamic variables, including four kinetic moments from the NSIDC sea ice motion data, and ERA5 air temperature, and wind velocity data were averaged by 1, 2, 5, 10, 20, and 30 days prior to the date when the DMS image was taken, considering the accumulative effects of these explanatory variables.

After exploring all possible multiple linear regression models, we found that dynamic-thermodynamic variables integrated by 10 days show the highest correlation coefficient. Therefore, we use these explanatory variables to reconstruct the linear regression models using the forward and backward stepwise regression approach. The results are illustrated in Table 12. All variables are normalized in the models.

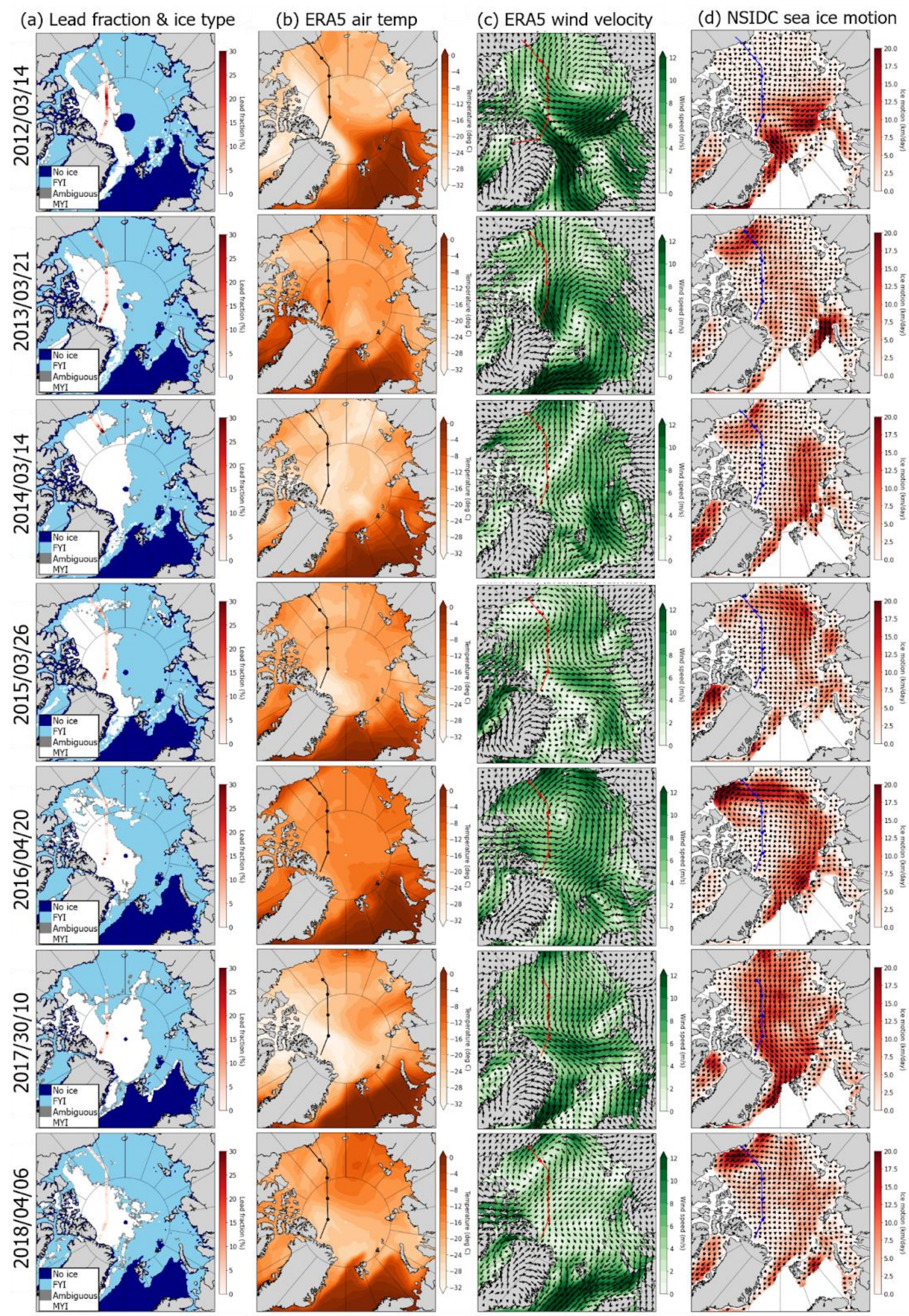


Figure 22. (a) DMS-based lead fraction and nearby ice type; (b) ERA5 air temperature; (c) ERA5 wind velocity (vectors) and speed (shaded); (d) sea ice motion for each year.

Table 12. Stepwise linear regression results.

Year	Approach	R^2	tmp10	U10 _10	V10 _10	wind_10	u_ice_10	v_ice_10	vel_ice_10	div10	vor10	shr10	stc10	constant
2012	forward	0.26	/	/	/	/	-0.39	-0.38	0.16	-0.10	-0.08	/	/	0.41
	backward	0.26	0.10	/	/	/	-0.34	-0.19	/	-0.12	/	/	/	0.31
2013	forward	0.48	-1.19	/	/	0.35	-6.46	-2.78	9.51	/	-0.01	-0.14	/	0.60
	backward	0.48	-1.18	/	/	0.35	-6.44	-2.75	9.45	/	/	-0.15	/	0.08
2014	forward	0.87	4.61	-5.60	-0.97	1.09	1.24	15.31	-12.98	/	0.89	-0.55	/	-2.08
	backward	0.87	4.64	-5.37	/	/	1.16	13.34	-11.25	-0.16	0.87	-0.59	/	-1.94
2015	forward	0.34	/	/	-0.53	/	-1.35	/	1.19	0.15	0.14	0.28	-0.33	0.40
	backward	0.34	/	/	-0.53	/	-1.35	/	1.19	0.15	0.14	0.28	-0.33	0.40
2016	forward	0.29	/	-0.79	/	/	/	/	0.29	0.30	-0.39	0.57	0.15	0.21
	backward	0.34	0.67	-4.62	-0.53	4.09	/	/	/	/	-0.36	0.46	/	0.22
2017	forward	0.66	-1.17	-6.54	-3.08	6.77	2.98	-0.09	-2.01	-0.19	/	/	/	1.50
	backward	0.66	-1.15	-6.57	-3.11	6.86	3.02	/	-2.09	-0.19	/	/	/	1.45
2018	forward	0.30	0.34	-1.40	-1.40	1.83	/	/	/	/	-0.03	/	-0.31	0.45
	backward	0.30	0.34	-1.31	-1.33	1.72	/	/	/	/	/	/	-0.32	0.42

There are 11 thermodynamic-dynamic variables, including one thermodynamic variable (temperature), six dynamic variables (velocity of wind and ice motion), and four kinetic moments caused by ice motion. The forward and backward stepwise regression models for each year identify different sets of explanatory variables (Table 12). Both 2012 models identify the ice motion velocity and divergence as the significant explanatory variables. The 2013 models mainly identify the ice motion velocity and temperature variables. Besides ice motion velocity and temperature, the 2014 models include wind velocity at u-direction, and the correlation coefficient is significantly higher than other models. The 2015 models emphasize the functions of wind and ice motion velocity. The 2016 forward model identifies more kinetic moments, but the backward model emphasizes more on wind velocity, which represents the possible correlation among these variables. And the 2017 and 2018 models show significant influence of wind velocity and temperature.

Except for 2012, all other models have only moderate correlation. It is reasonable because (1) the sea ice fractions were derived from high spatial resolution DMS images, and the dynamic-thermodynamic variables have a much coarser resolution of 25 km; (2) the atmospheric and oceanic dynamics that contribute to lead formation can occur in a much smaller scale (< 25 km scale), which cannot be captured by coarse resolution products; and (3) the uncertainty of the DMS-based lead detection (90% of accuracy) can be carried and exaggerated in the data fusion and resampling process.

Based on all stepwise regression results, the relative explanatory variable importance can be ranked based on their frequencies, as summarized in Figure 23. It shows that temperature and ice motion vorticity are the leading factors of the formation of sea ice leads, followed by wind vorticity and kinetic moments or tensions of ice motion.

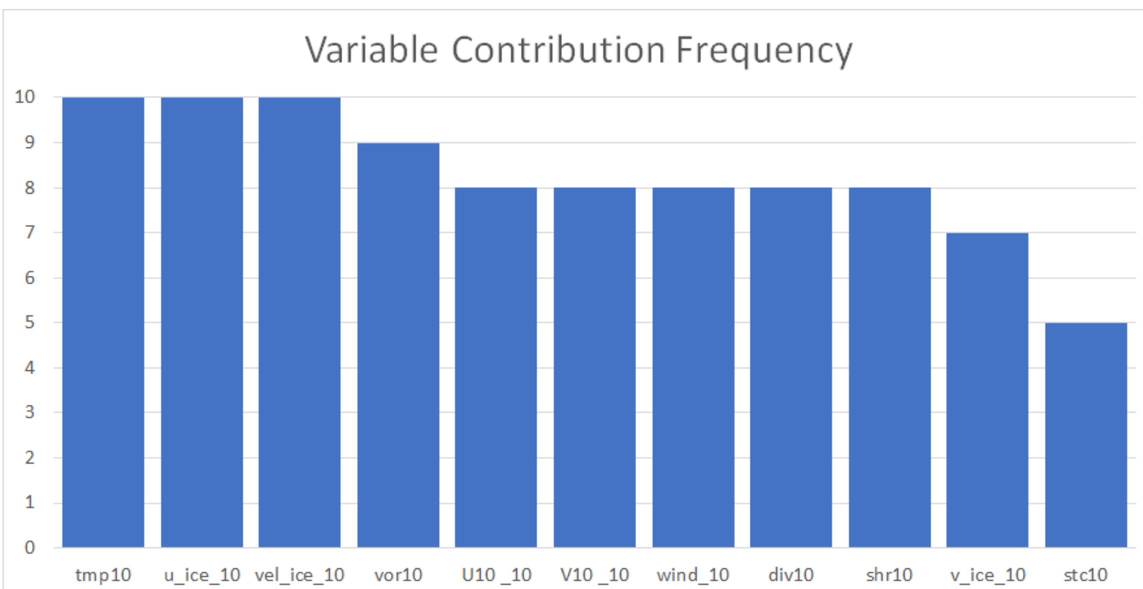


Figure 23. Relative importance of dynamic-thermodynamic explanatory variables.

4.4 Summary

This research demonstrates a scientific case study for sea ice lead detection during 2012-2018 along the IceBridge Laxon Line through using ArcCI cyberinfrastructure. To address the lack of standard image processing workflow for sea ice parameter extraction from massive and long-term HSR imagery, a practical object-based image classification workflow is implemented based on the OSSP package to extract multi-scale and different sea ice features, calculate the sea ice leads fractions and freeboard parameters. These sea ice products can be directly used to validate other coarse resolution remote sensing images/products. Furthermore, the above derived sea ice properties can be analyzed to address the scientific objectives, and specifically the sea ice fractions are modeled by large scale dynamic-thermodynamic variables.

In the future, the collaborators and other scientists would be engaged to use and test the algorithms and parameters derived from this study, to improve and modify the algorithms and parameters to derive.

CHAPTER 5. CONCLUSIONS AND FUTURE RESEARCH

5.1 Conclusion

Sea ice plays an important role in climate change. HSR sea ice images captured by satellites or airplanes provide detailed observational data for extracting geophysical attributes of sea ice features such as floe or melt pond shape, distribution, and coverage. HSR images, however, pose the serious challenge of discovering spatiotemporal patterns of sea ice from heterogeneous big data in a timely manner (M. Yu et al., 2020). This dissertation proposes and implements the ArcCI system and service based on cloud computing to handle the big data challenge in sea ice HSR image feature extraction from HSR imagery and demonstrates a geophysical extraction and spatiotemporal analysis case study of sea ice leads and relevant climate and environmental factors. Such capabilities are achieved by the following theoretical and technological contributions:

1. The ArcCI web service provides a one-stop platform for HSR image management (storage, archival retrieval/access, and backup), analysis (image processing, classification, and statistics), and visualization.
2. HSR sea ice observation imagery from the past 20 years was reviewed, collected, and integrated into an operational science gateway providing valuable metadata collection on ground-reference level observations in polar research. Potentially available sea ice geophysical parameters are reviewed for future functionality improvement and enhancement.

3. A practical object-based image classification workflow is established to process 10,000+ HSR images for geophysical parameter extraction of sea ice leads.
4. A scientific study is conducted by analyzing the correlation relationship between sea ice leads and relative climate and environmental factors from a spatiotemporal perspective.

5.2 Future Works

In the future, to improve the proposed Arctic Cyberinfrastructure, several directions for future research are proposed. The first one is to enhance the ArcCI system (Yang et al., 2020) by (1) including more scalable computing resources for the dynamic on-demand Web service, which would enable users to process and analyze HSR images using pixel-based or object-based methods; for example the integration of the lambda framework and elastic container techniques could convert the image processing usage in a pay-as-you-go mode and the load balance/autoscaling modules could adjust the operation style according to real-time demands of users' requests; (2) integrating more data visualization functionalities for data exploratory analysis; for example, the study methods employed in polar sciences can be easily applied to research in the Antarctic regions, research on land cover area with glaciers, and other ecosystem studies; and (3) optimizing high performance computing for big data processing by taking advantage of Spark in distributed memory or other advantage processing frameworks such as the elastic AWS components (Vance et al., 2016) including Simple Cloud Storage (S3) and

Relational Database Service (RDS) while taking into account the original distributed file storage and database settings.

The second direction is to improve the accuracy of classification and release the manually labelled work based on traditional supervised machine learning approaches. Several newer deep learning methodologies have been used to improve upon sea ice classification and detection accuracy (Dowden et al., 2020; Han et al., 2018, 2020; W. Li et al., 2021). While not a lot of literature on sea ice classification and detection based on optical imagery has been published in recent years, it is important to see how new developments in deep learning methodologies can help better tackle misclassification caused by different lighting and weather conditions as well as shadows caused by overhangs and hills. DL requires a large amount of labelled training data, which can be collected or generated by the comprehensive sea ice training dataset from the ArcCI platform with multiple cooperative strategies in an operational mode.

The third direction is to use the developed web-service to answer more critical scientific questions in polar studies by (1) supporting more sea ice HSR image processing methods from new polar sea ice observation sensors (Bhardwaj et al., 2016; Leary, 2017; M. Wang et al., 2018); (2) integrating data fusion analysis by combining low spatial resolution satellite images to extract geophysical properties at different scales (Bogdanov et al., 2007; W. Li et al., 2021); and (3) verifying and evaluating the derived parameters with other relevant factors of climate, environment, and other geophysical simulation models with a larger spatial and temporal coverage.

REFERENCES

Alam, A. (1997). Determination of surface turbulent fluxes over leads in Arctic sea ice. *Journal of Geophysical Research C: Oceans*, *102*(C2), 3331–3343. https://doi.org/10.1029/96JC03606

Andreas, E. L., & Murphy, B. (1986). Bulk Transfer Coefficients for Heat and Momentum over Leads and Polynyas. *Journal of Physical Oceanography*, *16*(11), 1875–1883. https://doi.org/10.1175/1520-0485(1986)016<1875:BTCFHA>2.0.CO;2

Bhardwaj, A., Sam, L., Akanksha, Martín-Torres, F. J., & Kumar, R. (2016). UAVs as remote sensing platform in glaciology: Present applications and future prospects. *Remote Sensing of Environment*, *175*, 196–204. https://doi.org/10.1016/j.rse.2015.12.029

Bogdanov, A. V., Sandven, S., Johannessen, O. M., Alexandrov, V. Y., & Bobylev, L. P. (2007). Multi-sensor approach to automated classification of sea ice image data. *Image Processing for Remote Sensing*, *43*(7), 293–324.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Cavalieri, D. J., Parkinson, C. L., & Vinnikov, K. Y. (2003). 30-Year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability. *Geophysical Research Letters*, *30*(18). https://doi.org/10.1029/2003GL018031

Dominguez, R. (2010). IceBridge DMS L1B geolocated and orthorectified images. *NASA National Snow and Ice Data Center Distributed Active Archive Center: Boulder, CO, USA*.

Dowden, B., De Silva, O., Huang, W., & Oldford, D. (2020). Sea Ice Classification via Deep Neural Network Semantic Segmentation. *IEEE Sensors Journal*, 1–1. https://doi.org/10.1109/JSEN.2020.3031475

Duncan, K, Farrell, S. L., Connor, L. N., Richter-Menge, J., Hutchings, J. K., & Dominguez, R. (2018). High-resolution airborne observations of sea-ice pressure ridge sail height. *Annals of Glaciology*, *59*(76pt2), 137–147. https://doi.org/10.1017/aog.2018.2

Duncan, Kyle, Farrell, S. L., Hutchings, J., & Richter-Menge, J. (2020). Late Winter Observations of Sea Ice Pressure Ridge Sail Height. *IEEE Geoscience and Remote Sensing Letters*, 1–5. https://doi.org/10.1109/lgrs.2020.3004724

Farrell, S. L., Brunt, K. M., Ruth, J. M., Kuhn, J. M., Connor, L. N., & Walsh, K. M. (2015). Sea-ice freeboard retrieval using digital photon-counting laser altimetry.

Annals of Glaciology, *56*(69), 167–174. https://doi.org/10.3189/2015AoG69A686

Flocco, D., & Feltham, D. L. (2007). A continuum model of melt pond evolution on Arctic sea ice. *Journal of Geophysical Research: Oceans*, *112*(8), 8016. https://doi.org/10.1029/2006JC003836

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, Issue 2). MIT press Cambridge.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gui, Z., Yang, C., Xia, J., Li, J., Rezgui, A., Sun, M., Xu, Y., & Fay, D. (2013). A visualization-enhanced graphical user interface for geospatial resource discovery. *Annals of GIS*, *19*(2), 109–121. https://doi.org/10.1080/19475683.2013.782467

Hakkinen, S., Proshutinsky, A., & Ashik, I. (2008). Sea ice drift in the Arctic since the 1950s. *Geophysical Research Letters*, *35*(19), 19704. https://doi.org/10.1029/2008GL034791

Han, Y., Li, P., Zhang, Y., Hong, Z., Liu, K., & Wang, J. (2018). Combining active learning and transductive support vector machines for sea ice detection. *Journal of Applied Remote Sensing*, *12*(02), 1. https://doi.org/10.1117/1.jrs.12.026016

Han, Y., Wei, C., Zhou, R., Hong, Z., Zhang, Y., & Yang, S. (2020). Combining 3D-CNN and Squeeze-and-Excitation Networks for Remote Sensing Sea Ice Image Classification. *Mathematical Problems in Engineering*, *2020*, 1–15. https://doi.org/10.1155/2020/8065396

Hirano, D., Fukamachi, Y., Watanabe, E., Ohshima, K. I., Iwamoto, K., Mahoney, A. R., Eicken, H., Simizu, D., & Tamura, T. (2016). A wind-driven, hybrid latent and sensible heat coastal polynya off Barrow, Alaska. *Journal of Geophysical Research: Oceans*, *121*(1), 980–997. https://doi.org/10.1002/2015JC011318

Howell, S. E. L., Small, D., Rohner, C., Mahmud, M. S., Yackel, J. J., & Brady, M. (2019). Estimating melt onset over Arctic sea ice from time series multi-sensor Sentinel-1 and RADARSAT-2 backscatter. *Remote Sensing of Environment*, *229*, 48–59. https://doi.org/10.1016/j.rse.2019.04.031

Huang, Q., Yang, C., Benedict, K., Chen, S., Rezgui, A., & Xie, J. (2013). Utilize cloud computing to support dust storm forecasting. *International Journal of Digital Earth*, *6*(4), 338–355. https://doi.org/10.1080/17538947.2012.749949

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 91–106. https://doi.org/10.1016/j.isprsjprs.2013.03.006

Hutchings, J. K., Hughes, N., Orlich, A. R., MacFarlane, S., Cowen, A., Farmer, L., & Faber, M. K. (2016). Ice Watch: Standardizing and expanding Arctic ship based sea ice observations. *White Paper Presented at Arctic Observing Summit, Fairbanks, Alaska, USA.*

Jeffries, M. O., Overland, J. E., & Perovich, D. K. (2013). The Arctic shifts to a new normal. *Physics Today*, *66*(10), 35–40. https://doi.org/10.1063/PT.3.2147

Jodha Khalsa, S., Parsons, M., Yarmey, L., Truslove, I., Pearlman, J., & Boldrini, E. (2013). The Advanced Cooperative Arctic Data and Information Service (ACADIS). In *Geophysical Research Abstracts* (Vol. 15).

Kraaijenbrink, P. D. A., Shea, J. M., Pellicciotti, F., Jong, S. M. d., & Immerzeel, W. W. (2016). Object-based analysis of unmanned aerial vehicle imagery to map and characterise surface features on a debris-covered glacier. *Remote Sensing of Environment*, *186*, 581–595. https://doi.org/10.1016/j.rse.2016.09.013

Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., Onana, V. D., Panzer, B., & Sonntag, J. G. (2013). Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data. *Cryosphere*, *7*(4), 1035–1056. https://doi.org/10.5194/tc-7-1035-2013

Kurtz, N. T., & Markus, T. (2012). Satellite observations of Antarctic sea ice thickness and volume. *Journal of Geophysical Research: Oceans*, *117*(C8), n/a-n/a. https://doi.org/10.1029/2012JC008141

Kwok, R. (2010). Satellite remote sensing of sea-ice thickness and kinematics: a review. *Journal of Glaciology*, *56*(200), 1129–1140. https://doi.org/10.3189/002214311796406167

Kwok, Ronald, & Untersteiner, N. (2011). New High-Resolution Images of Summer Arctic Sea Ice. *Eos, Transactions American Geophysical Union*, *92*(7), 53–54. https://doi.org/10.1029/2011EO070002

Leary, D. (2017). Drones on ice: an assessment of the legal implications of the use of unmanned aerial vehicles in scientific research and by the tourist industry in Antarctica. *Polar Record*, *53*(271), 343–357. https://doi.org/10.1017/S0032247417000262

Lei, RuiBo, LI, Z., CHENG, B., YANG, Q., & LI, N. (2016). Investigation of the thermodynamic processes of a floe-lead system in the central Arctic during later summer. *Advances in Polar Science*, 10–16. https://doi.org/10.3724/SP.J.1085.2011.00010

Lei, Ruibo, Zhang, Z., Matero, I., Cheng, B., Li, Q., & Huang, W. (2012). Reflection and transmission of irradiance by snow and sea ice in the central Arctic Ocean in summer 2010. *Polar Research*, *31*(1), 17325. https://doi.org/10.3402/polar.v31i0.17325

Li, W., Liu, L., & Zhang, J. (2021). Fusion of SAR and Optical Image for Sea Ice Extraction. *Journal of Ocean University of China*. https://doi.org/10.1007/s11802-021-4824-y

Li, Z., Yang, C., Huang, Q., Liu, K., Sun, M., & Xia, J. (2017). Building Model as a Service to support geosciences. *Computers, Environment and Urban Systems*, *61*, 141–152. https://doi.org/10.1016/j.compenvurbsys.2014.06.004

Li, Z., Yang, C. P., Wu, H., Li, W., & Miao, L. (2011). An optimized framework for seamlessly integrating OGC Web Services to support geospatial sciences. *International Journal of Geographical Information Science*, *25*(4), 595–613. https://doi.org/10.1080/13658816.2010.484811

Liu, D., & Xia, F. (2010). Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, *1*(4), 187–194. https://doi.org/10.1080/01431161003743173

Lu, P., Li, Z., Cheng, B., Lei, R., & Zhang, R. (2010a). Sea ice surface features in Arctic summer 2008: Aerial observations. *Remote Sensing of Environment*, *114*(4), 693–699. https://doi.org/10.1016/j.rse.2009.11.009

Lu, P., Li, Z., Cheng, B., Lei, R., & Zhang, R. (2010b). Sea ice surface features in Arctic summer 2008: Aerial observations. *Remote Sensing of Environment*, *114*(4), 693–699. https://doi.org/10.1016/j.rse.2009.11.009

Marcq, S., & Weiss, J. (2012). Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere. *Cryosphere*, *6*(1), 143–156. https://doi.org/10.5194/tc-6-143-2012

Markus, T., Cavalieri, D. J., Tschudi, M. A., & Ivanoff, A. (2003). Comparison of aerial video and Landsat 7 data over ponded sea ice. *Remote Sensing of Environment*, *86*(4), 458–469. https://doi.org/10.1016/S0034-4257(03)00124-X

Marshall, M. (2013). Arctic ice low kicks off a cascade of tipping points. *New Scientist*, *217*(2906), 6–7. https://doi.org/10.1016/S0262-4079(13)60528-X

McLaren, A. S., Walsh, J. E., Bourke, R. H., Weaver, R. L., & Wrttmann, W. (1992). Variability in sea-ice thickness over the North Pole from 1977 to 1990. *Nature*, *358*(6383), 224–226. https://doi.org/10.1038/358224a0

Meier, W N, Markus, T., & Comiso, J. C. (2018). AMSR-E/AMSR2 Unified L3 Daily 12.5 km Brightness Temperatures, Sea Ice Concentration, Motion \& Snow Depth Polar Grids, Version 1, November 2018. *Arctic Grid, Https://Doi. Org/10.5067/RA1MIJOYPK3P*. https://doi.org/https://doi.org/10.5067/RA1MIJOYPK3P

Meier, Walter N. (2016). Losing Arctic sea ice: observations of the recent decline and the long-term context. In *Sea Ice* (pp. 290–303). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118778371.ch11

Mell, P., Grance, T., & others. (2011). *The NIST definition of cloud computing*.

Miao, X., Xie, H., Ackley, S. F., Perovich, D. K., & Ke, C. (2015). Object-based detection of Arctic sea ice and melt ponds using high spatial resolution aerial photographs. *Cold Regions Science and Technology*, *119*, 211–222.

https://doi.org/10.1016/j.coldregions.2015.06.014

Miao, X., Xie, H., Ackley, S. F., & Zheng, S. (2016). Object-Based Arctic Sea Ice Ridge Detection from High-Spatial-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters*, *13*(6), 787–791. https://doi.org/10.1109/LGRS.2016.2544861

Molinari, R., & Kirwan, A. D. (1975). Calculations of Differential Kinematic Properties from Lagrangian Observations in the Western Caribbean Sea. *Journal of Physical Oceanography*, *5*(3), 483–491. https://doi.org/10.1175/1520-0485(1975)005<0483:CODKPF>2.0.CO;2

National Research Council. (2007). *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. National Academies Press. https://doi.org/10.17226/11820

Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015). Deep learning for emotion recognition on small datasets using transfer learning. *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, 443–449. https://doi.org/10.1145/2818346.2830593

Norwegian Polar Institute. (n.d.). *Norwegian Polar Data Centre*. Retrieved June 13, 2021, from https://data.npolar.no/home/

Onana, V.-D.-P., Kurtz, N. T., Farrell, S. L., Koenig, L. S., Studinger, M., & Harbeck, J. P. (2013). A Sea-Ice Lead Detection Algorithm for Use With High-Resolution Airborne Visible Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 38–56. https://doi.org/10.1109/TGRS.2012.2202666

Parkinson, C. L. (2019). A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(29), 14414–14423. https://doi.org/10.1073/pnas.1906556116

Parkinson, C. L., & Comiso, J. C. (2013). On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophysical Research Letters*, *40*(7), 1356–1361. https://doi.org/10.1002/grl.50349

Peng, G., Matthews, J. L., & Yu, J. T. (2018). Sensitivity Analysis of Arctic Sea Ice Extent Trends and Statistical Projections Using Satellite Data. *Remote Sensing 2018, Vol. 10, Page 230*, *10*(2), 230. https://doi.org/10.3390/RS10020230

Perovich, D. K. (2003). Thin and thinner: Sea ice mass balance measurements during SHEBA. *Journal of Geophysical Research*, *108*(C3), 8050. https://doi.org/10.1029/2001JC001079

Perovich, D. K., Grenfell, T. C., Light, B., Elder, B. C., Harbeck, J., Polashenski, C., Tucker, W. B., & Stelmach, C. (2009). Transpolar observations of the morphological properties of Arctic sea ice. *Journal of Geophysical Research: Oceans*, *114*(1), 0–04. https://doi.org/10.1029/2008JC004892

Renner, A. H. H., Dumont, M., Beckers, J., Gerland, S., & Haas, C. (2013). Improved characterisation of sea ice using simultaneous aerial photography and sea ice thickness measurements. *Cold Regions Science and Technology*, *92*, 37–47. https://doi.org/10.1016/j.coldregions.2013.03.009

Röhrs, J., & Kaleschke, L. (2012). An algorithm to detect sea ice leads by using AMSR-E passive microwave imagery. *The Cryosphere*, *6*(2), 343–352. https://doi.org/10.5194/tc-6-343-2012

Rösel, A., & Kaleschke, L. (2011). Comparison of different retrieval techniques for melt ponds on Arctic sea ice from Landsat and MODIS satellite data. *Annals of Glaciology*, *52*(57), 185–191. https://doi.org/10.3189/172756411795931606

Sandven, S., Johannessen, O. M., & Kloster, K. (2006). Sea Ice Monitoring by Remote Sensing. In *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470027318.a2320

Scambos, T. A., Haran, T. M., & Massom, R. (2006). Validation of AVHRR and MODIS ice surface temperature products using in situ radiometers. *Annals of Glaciology*, *44*, 345–351. https://doi.org/10.3189/172756406781811457

Scharfen, G. R., Hall, D. K., & Riggs, G. A. (1997). MODIS snow and ice products from the NSIDC DAAC. In W. L. Barnes (Ed.), *Earth Observing Systems II* (Vol. 3117, Issue September 1997, pp. 143–147). https://doi.org/10.1117/12.283803

Sha, D., Miao, X., Xu, M., Yang, C., Xie, H., Mestas-Nuñez, A. M., Li, Y., Liu, Q., & Yang, J. (2020). An On-Demand Service for Managing and Analyzing Arctic Sea Ice High Spatial Resolution Imagery. *Data*, *5*(2), 39. https://doi.org/10.3390/data5020039

Shapiro, L., & Rosenfeld, A. (1992). *Computer vision and image processing*.

Studinger, M. (2013). IceBridge ATM L1B elevation and return strength, Version 2. *Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center. DOI*, *10*.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era*.

Teleti, P. R., & Luis, A. J. (2013). Sea Ice Observations in Polar Regions: Evolution of Technologies in Remote Sensing. *International Journal of Geosciences*, *04*(07), 1031–1050. https://doi.org/10.4236/ijg.2013.47097

Thomas, M., Geiger, C. A., & Kambhamettu, C. (2008). High resolution (400 m) motion characterization of sea ice using ERS-1 SAR imagery. *Cold Regions Science and Technology*, *52*(2), 207–223. https://doi.org/10.1016/j.coldregions.2007.06.006

Tian, L., Xie, H., Ackley, S. F., Tinto, K. J., Bell, R. E., Zappa, C. J., Gao, Y., & Mestas-Nuñez, A. M. (2020). Sea ice freeboard in the Ross Sea from airborne altimetry IcePod 2016-2017 and a comparison with IceBridge 2013 and ICESat 2003-2008. *Remote Sensing*, *12*(14). https://doi.org/10.3390/rs12142226

Tschudi, M., Meier, W. N., Stewart, J. S., Fowler, C., & Maslanik, J. (2019). Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 4. *Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center.*, *10*. https://doi.org/https://doi.org/10.5067/INAWUWO7QH7B

Vance, T. C., Merati, N., Yang, C., & Yuan, M. (2016). Cloud computing for ocean and atmospheric science. *OCEANS 2016 MTS/IEEE Monterey*, 1–4. https://doi.org/10.1109/OCEANS.2016.7761270

Vecchiola, C., Pandey, S., & Buyya, R. (2009). High-performance cloud computing: A view of scientific applications. *I-SPAN 2009 - The 10th International Symposium on Pervasive Systems, Algorithms, and Networks*, 4–16. https://doi.org/10.1109/I-SPAN.2009.150

Walker Johnson, G., Gaylord, A. G., Franco, J. C., Cody, R. P., Brady, J. J., Manley, W., Dover, M., Garcia-Lavigne, D., Score, R., & Tweedie, C. E. (2011). Development of the Arctic Research Mapping Application (ARMAP): Interoperability challenges and solutions. *Computers & Geosciences*, *37*(11), 1735–1742. https://doi.org/10.1016/j.cageo.2011.04.004

Wang, M., Su, J., Mingfeng, W., Jie, S., Tao, L., Xiaoyu, W., Qing, J., Yong, C., Long, L., & Yilin, L. (2018). Determination of Arctic melt pond fraction and sea ice roughness from Unmanned Aerial Vehicle (UAV) imagery. *Adv Polar Sci*, *29*(3), 181–189. https://doi.org/10.13679/j.advps.2018.3.00181

Wang, Q., Danilov, S., Jung, T., Kaleschke, L., & Wernecke, A. (2016). Sea ice leads in the Arctic Ocean: Model assessment, interannual variability and trends. *Geophysical Research Letters*, *43*(13), 7019–7027. https://doi.org/10.1002/2016GL068696

Wang, X., Guan, F., Liu, J., Xie, H., & Ackley, S. (2016). An improved approach of total freeboard retrieval with IceBridge Airborne Topographic Mapper (ATM) elevation and Digital Mapping System (DMS) images. *Remote Sensing of Environment*, *184*, 582–594. https://doi.org/10.1016/j.rse.2016.08.002

Wang, X., Xie, H., Ke, Y., Ackley, S. F., & Liu, L. (2013). A method to automatically determine sea level for referencing snow freeboards and computing sea ice thicknesses from NASA IceBridge airborne LIDAR. *Remote Sensing of Environment*, *131*, 160–172. https://doi.org/10.1016/j.rse.2012.12.022

Watanabe, E., Wang, J., Sumi, A., & Hasumi, H. (2006). Arctic dipole anomaly and its contribution to sea ice export from the Arctic Ocean in the 20th century. *Geophysical Research Letters*, *33*(23), L23703. https://doi.org/10.1029/2006GL028112

Weissling, B., Ackley, S., Wagner, P., & Xie, H. (2009). EISCAM — Digital image acquisition and processing for sea ice parameters from ships. *Cold Regions Science and Technology*, *57*(1), 49–60. https://doi.org/10.1016/j.coldregions.2009.01.001

Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin*, *86*(1), 168–174. https://doi.org/10.1037/0033-2909.86.1.168

Worby, A. P., & Allison, I. (1991). Ocean-atmosphere energy exchange over thin, variable concentration Antarctic pack ice. *Annals of Glaciology*, *15*, 184–190. https://doi.org/10.3189/1991aog15-1-184-190

Wright, N. C., & Polashenski, C. M. (2018). Open-source algorithm for detecting sea ice surface features in high-resolution optical imagery. *Cryosphere*, *12*(4), 1307–1329. https://doi.org/10.5194/tc-12-1307-2018

Wright, N. C., Polashenski, C. M., Mcmichael, S. T., & Beyer, R. A. (2020). Observations of sea ice melt from Operation IceBridge imagery. *The Cryosphere, 14,* 3523–3536. https://doi.org/10.5194/tc-14-3523-2020

Xie, H., Lei, R., Ke, C., Wang, H., Li, Z., Zhao, J., & Ackley, S. F. (2013). Summer sea ice characteristics and morphology in the Pacific Arctic sector as observed during the CHINARE 2010 cruise. *Cryosphere*, *7*(4), 1057–1072. https://doi.org/10.5194/tc-7-1057-2013

Xie, H., Tian, Y. Q., Granillo, J. A., & Keller, G. R. (2007). Suitable remote sensing method and data for mapping and measuring active crop fields. *International Journal of Remote Sensing*, *28*(2), 395–411. https://doi.org/10.1080/01431160600702673

Xindong Wu, Xingquan Zhu, Gong-Qing Wu, & Wei Ding. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Yang, C., Clarke, K., Shekhar, S., & Tao, C. V. (2020). Big Spatiotemporal Data Analytics: a research and innovation frontier. *International Journal of Geographical Information Science*, *34*(6), 1075–1088. https://doi.org/10.1080/13658816.2019.1698743

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M., & Fay, D. (2011). Spatial cloud computing: How can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, *4*(4), 305–329. https://doi.org/10.1080/17538947.2011.587547

Yang, C., & Huang, Q. (2013). *Spatial Cloud Computing: A Practical Approach*. CRC Press, Inc.

Yang, C., Sun, M., Liu, K., Huang, Q., Li, Z., Gui, Z., Jiang, Y., Xia, J., Yu, M., Xu, C., Lostritto, P., & Zhou, N. (2015). Contemporary Computing Technologies for Processing Big Spatiotemporal Data. In *Space-Time Integration in Geography and GIScience* (pp. 327–351). Springer Netherlands. https://doi.org/10.1007/978-94-017-9205-9_18

Yu, M., Bambacus, M., Cervone, G., Clarke, K., Duffy, D., Huang, Q., Li, J., Li, W., Li, Z., Liu, Q., Resch, B., Yang, J., & Yang, C. (2020). Spatiotemporal event detection: a review. *International Journal of Digital Earth, 13*(12), 1339–1365. https://doi.org/10.1080/17538947.2020.1738569

Yu, Y., Maykut, G. A., & Rothrock, D. A. (2004). Changes in the thickness distribution of Arctic sea ice between 1958-1970 and 1993-1997. *Journal of Geophysical Research: Oceans*, *109*(C8), n/a-n/a. https://doi.org/10.1029/2003JC001982

BIOGRAPHY

Dexuan Sha graduated from Xian Middle School, Xi'an, China, in 2010. He received his Bachelor of Science from Hainan University in 2014 and completed his Master of Sciences study in Geospatial Science from Missouri State University in 2016.