Technische Universität München
Methodik der Fernerkundung
Fakultät für Luftfahrt, Raumfahrt und Geödäsie
Univ.-Prof. Dr.techn. Mag.rere.nat. Roland Pail

Mapping the grounding line of Antarctica in SAR interferograms with machine learning techniques

Sindhu Ramanath Tarekere

Master's Thesis

Master's Course in Earth Oriented Space Science and Technology

Supervisor(s):

1. Prof. Dr. Michael Eineder
   Deutsches Zentrum für Luft- und Raumfahrt (DLR)

2. Dr. Dana Floricioiu
   Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Date of Submission: May, 2022

Statement of Authorship

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

Munich, 04.05.2022

Author

Acknowledgement

I have met and learnt from several exceptional people during the course of my Masters, which means I have several people to thank. Although there is no ranking order for my gratitudes, I would first like to thank my mother. She is the steady foundation of my family and also the only person who makes family photos look somewhat pleasant. Her patience is something I have unfortunately not inherited, but I still try to lead my life the way she brought me up. She and my brother have really held the fort in my absence. My father has supported me in every way, it was only because of his persistence that I made the leap to study a Master's course 7000 km away from home. I am not sure how wise this decision was, but 2.5 years later, here I am, writing the acknowledgment for my Master thesis.

I would like to express my gratitude to Professor Eineder for employing me as a working student at MF-SAR. In terms of sculpting the thesis, a massive thank you to Dr. Dana Floricioiu. I will try to apply her meticulousness in my future academic writings.

A special thanks to Dr. Homa Ansari, without whom I might not have had a thesis topic. I will always remember her professional advice and our mutual dislike of German winters. In fact, it was through her that I was introduced to Dr. Lukas Krieger, to whom I atleast owe a huge bag of Indian coffee powder, if not my gratitude. His support in every aspect of the thesis is a major reason for its successful completion. I am especially grateful to him for patiently listening to my incoherent ramblings, pointing out a potentially disastrous mistake and helping me make plots with PyGMT (a hideously difficult task). A huge thanks to Konrad Heidler for helping me with everything machine learning and for answering my repeated questions.

This acknowledgment would be incomplete without mentioning my friend and ESPACE colleague Dharani, who very cleverly suggested that I go to Antarctica and train penguins to stand at the grounding line. Also a big thanks to Anagha and Akshatha for the cathartic zoom

vent-out sessions. Thank you to all my other ESPACE peers, it was very nice getting to know you even though we spent a very short time interacting in person.

Abstract

The grounding line marks the transition between ice grounded at the bedrock and the floating ice shelf. Its location is required for estimating ice sheet mass balance, modelling of ice sheet dynamics and glaciers and for evaluating ice shelf stability, which merits its long-term monitoring. The line migrates both due to short term influences such as ocean tides and atmospheric pressure, and long-term effects such as changes of ice thickness, slope of bedrock and variations in sea level.

Of the numerous in-situ and remote sensing methods currently in use to map the grounding line, Differential Interferometric Synthetic Aperture Radar (DInSAR) is, by far, the most accurate technique which produces spatially dense delineations. Tidal deformation at the ice sheet-ice shelf boundary is visible as a dense fringe belt in DInSAR interferograms and its landward limit is taken as a good approximation of the grounding line location (GLL). The GLL is usually manually digitized on the interferograms by human operators. This is both time consuming and introduces inconsistencies due to subjective interpretation especially in low coherence interferograms. On a large scale and with increasing data availability a key challenge is the automation of the delineation procedure.

So far, a limited amount of studies were published regarding the delineation processes of typical features on the ice sheets using deep neural networks (DNNs). The objectives of this thesis were to further explore the feasibility of using machine learning for mapping the interferometric grounding line, as well as exploring the contributions of complementary features such as coherence estimated from phase, Digital Elevation Model, ice velocity, tidal displacement and atmospheric pressure, in addition to DInSAR interferograms. A dataset composed of manually delineated GLLs generated within ESA's Antarctic Ice Sheet Climate Change Initiative project and corresponding DInSAR interferograms from ERS-1/2, Sentinel-1 and TerraSAR-X missions over Antarctica together with the above mentioned features was compiled and used for training two DNNs: Holistically-Nested Edge Detection (HED) and

UNet. The developed processing chain handles creation of the training feature stack, training the DNNs and performing post processing functions on the resulting predictions.

HED outperformed UNet and was able to achieve a median deviation (from manual delineations) of 209.23 m with a median absolute deviation of 152.91 m. Analysis of the individual feature contributions revealed that only the phase and derived features (real and imaginary interferogram components and coherence estimates) substantially influence the predicted GLLs. This finding is advantageous in terms of saving time, computational effort and memory in creating and storing the above mentioned feature stack.

Although the delineations generated from HED do not perfectly follow the true GLL in all locations, the gains in efficiency and consistency are considerable, compared to the time and effort spent for manual digitizations. This study shows the potential of DNNs for automating the interferometric GLL delineation process.

Contents

List of Tables

List of Figures

1. Introduction

1.1. Motivation

The Antarctic Ice Sheet (AIS) and Greenland Ice Sheet (GrIS) constitute the majority (99.5%) of land ice and hold 70% of the Earth's freshwater (Bamber et al. 2018). The volume of water contained in the AIS converts to 58.3m sea level equivalent (SLE) (Vaughan et al. 2013), making it one of the largest potential sources of sea level rise. The release of freshwater affects circulation, seawater salinity and biogeochemistry of oceans. Sea level rise poses an imminent danger to coastal communities via increased incidences of flooding, land submergence, coastal erosion and salinization of soil, ground and surface water (Oppenheimer et al. 2019).

Recent studies summarised in the sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC) (Fox-Kemper et al. 2021) indicate an increase in the rate of ice mass loss from the past decades, with major contributions from the West Antarctic Ice Sheet (WAIS) (Fig. 1.1)

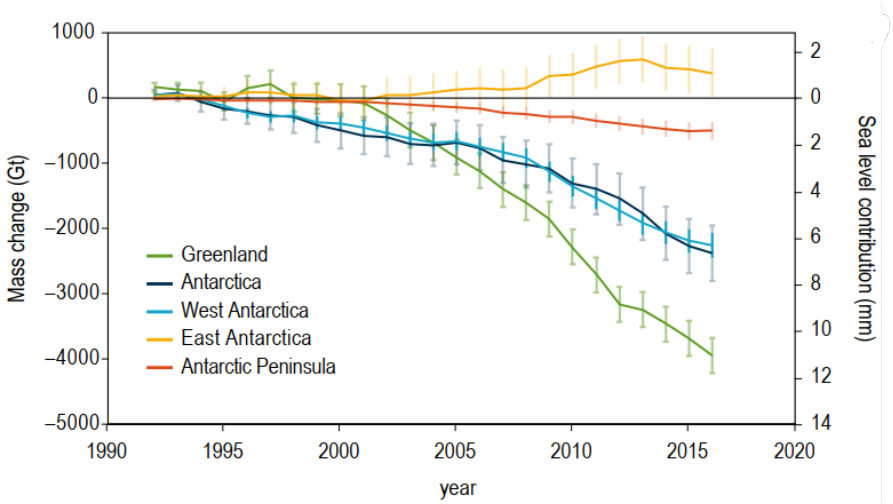


Figure 1.1: Cumulative Ice Sheet mass change (results derived from several studies), 1992–2016 (Meredith et al. 2019). Uncertainties are 1 standard deviation.

While ice sheet mass is gained via snow accumulation, mass loss occurs, roughly, through surface melt runoff and iceberg calving at outlet glaciers (Smith et al. 2020). The balance between these competing processes is governed by complex interactions. Numerical models aim to understand these dynamics and predict the evolution of ice sheets. An important parameter of these models is the grounding line location.

The grounding line defines the boundary between the grounded ice sheet and floating ice shelf. This transition zone, described in Section 2.2.1, plays an important role in controlling ice sheet dynamics as it is the point where the ice sheet is first in contact with ocean water and experiences oceanic forcing. The rate at which ice flows out from the grounded part of the ice sheet is assessed as the ice flux through the grounding line and is dependent on the ice thickness and velocity at this position. Besides its role in determining the rate of ice discharge, the grounding line marks the geometrical bounds of an ice sheet, aiding in the calculation of ice sheet area.

Grounding lines migrate back and forth on short timescales as a consequence of the ice bending due to ocean tides (Paterson 2016), as well as in the long term due to changes in ice thickness, which can be caused by the accumulation of snowfall, topography of the bed, basal melting from warm ocean water intrusion and basal shear stress (Schoof 2007).

Most of Antarctica‘s coastline (75%) is fringed by floating ice shelves and its grounding line marks the landward limit of the grounding zone that is several kilometres wide, which can be difficult to directly observe on the featureless ice surface. Fortunately, there are various remote sensing techniques to detect the grounding line at large scales. Numerous satellite campaigns and the availability of high resolution datasets containing large temporal and spatial extents is a great advantage in favour of mapping the grounding line with satellite-based remote sensing methods compared to in-situ methods.

The technique of interest for this thesis is the Differential Interferometric Synthetic Aperture Radar (DInSAR) (explained in detail in Section 2.3.3). The required interferograms are

calculated from spaceborne SAR data with short repeat cycles, such as those acquired by the ERS-1/2 tandem, TerraSAR-X or the Sentinel-1 missions. The grounding zone appears in the differential interferograms as a distinct pattern - a dense fringe belt which can be used to map the grounding line (Rignot 1996).

While there are several methods for extracting the grounding line position from the double difference interferogram (Section 3.1.1), most often, human experts perform manual digitizations. This is not just costly in terms of time, but also introduces inconsistencies due to variable interpretations among operators. A further challenge is dealing with low coherence interferograms, resulting in incomplete lines.

This thesis aims at automating the digitization process using machine learning techniques. Being a relatively nascent area of research, existing literature does not fully cover the feasibility of using machine learning for this problem. The results from the experiments carried out within the study would shed some light on this in terms of features and architectures explored for this task.

1.2. Thesis Objectives

This research aims to address the following questions:

1. **Feasibility of using machine learning for delineating grounding lines**

   Traditional grounding line delineation methods found in literature (Section 3.1.1) often involve unwrapping the phase of the double difference interferograms, whose quality in turn is dependent on the interferometric phase coherence. This makes the digitization process tedious. While phase unwrapping is a non-trivial problem, the heuristics employed by human operators are considered to be within an acceptable range of error (Rignot et al. 2011). In regards to automating this process with machine learning, an attempt was made by (Mohajerani et al. 2021), wherein delineation is viewed as a semantic segmentation problem. This thesis aims at tackling grounding line delineation as an edge detection problem instead, and further tests the feasibility of detection by using

non-image data (vectorized grounding lines) with classic deep learning architectures such as Holistically-Nested Edge Detection (HED) and UNet (Section 4.2).

2. **Feature exploration**

   Grounding line locations do not just manifest as fringes in double difference interferograms. They are sometimes also apparent as a break in slope on the surface or in regions where there is a noticeable change in ice velocity and tidal displacements. Apart from interferometric phase, the grounding line location could also be gleaned by tracking changes in the coherence of the corresponding interferograms. Areas of low coherence could be an indicator of a fast changing phase, implying the presence of the grounding zone. Investigating the contribution of these features towards grounding line detection is the second objective of the thesis. In this regard, nine features were investigated (Section 4.1.1).

2. Theoretical Overview

2.1. Antarctic Ice Sheet (AIS)

The Antarctic continent is covered by a continental ice sheet and is nearly completely surrounded by ice shelves, with smaller topographic features such as crevasses, ice rises and ice rumples scattered throughout. Ice rises and ice rumples are regions of grounded ice that are either partially or completely surrounded by ice shelves. Ice rises have an elevation several hundred metres above the enclosing ice shelf. Ice rumples on the other hand, are only tens of metres high (Matsuoka et al. 2015).

The continental ice, referred to as the Antarctic Ice Sheet, has a 43,500 km long coastline (Liu and Jezek 2004), spans an area of $14 \times 10^6 km^2$ (Fox et al. 1994) and has a grounding line of 67,000 km length (estimated from the MEaSUREs dataset described in 3.1.2 (Rignot et al. 2016)).

AIS is separated into the East Antarctic Ice Sheet (EAIS) and the West Antarctic Ice Sheet (WAIS) by the Transantarctic Mountain Belt (Fig 2.1). Glaciers in the WAIS have experienced drastic ice mass loss in the past four decades (Gardner et al. 2018), (Shepherd et al. 2018), (Rignot et al. 2019). The ice mass loss is thought (arguably) to be exacerbated due to the retrograde (landward sloping) bedrock of the WAIS (Pattyn 2018). Large uncertainties in the mass balance computation for EAIS (Meredith et al. 2019) makes it difficult to comment on its stability. However, several glaciers in the EAIS (Totten, Denman and Lambert) are thought to be in a state of dynamic imbalance due to intermittent periods of ice mass loss(Meredith et al. 2019).

2.2. Grounding line and grounding zone

2.2.1 Grounding zone configurations

Different forces act upon grounded ice and ice shelves. The grounding zone is an area of varying width (2 km - 10 km) (Brunt et al. 2010) at the ice sheet - ice shelf boundary which,

Figure 2.1: The Antarctic Ice Sheet and its surrounding ice shelves, plotted on basemap from Quantarctica 3 (Matsuoka et al. 2021), MEaSUREs (Rignot et al. 2016) grounding line in black.

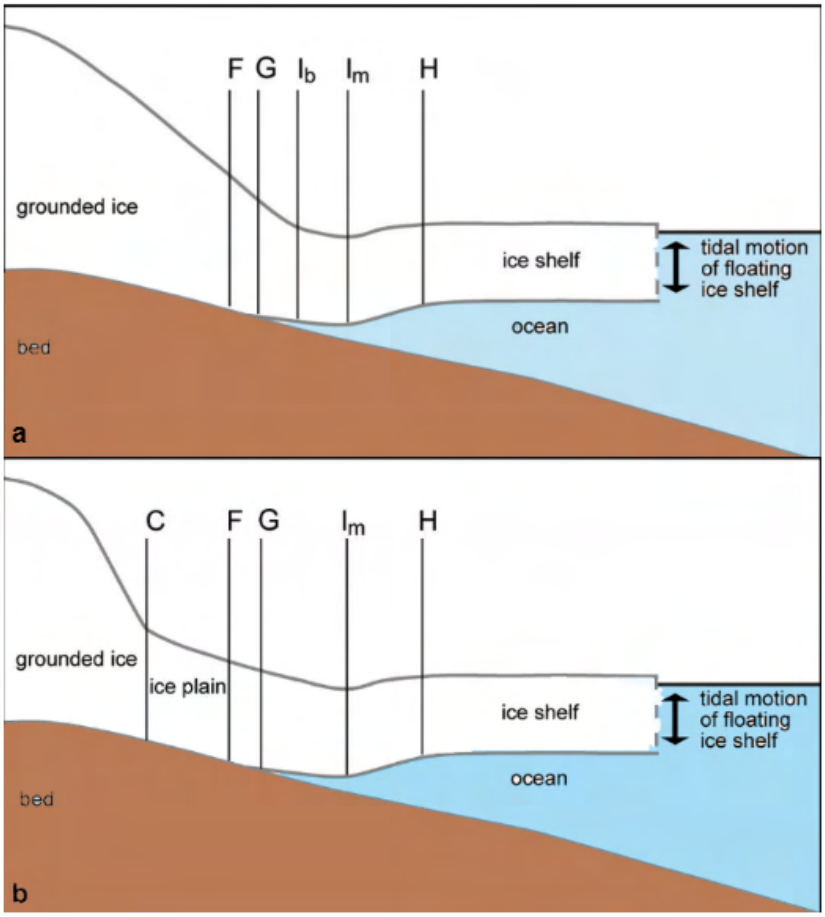


Figure 2.2: Schematic showing grounding zone features, adapted from (Brunt et al. 2011). (a) no ice plain and (b) with ice plain. Point F is the landward limit of ice flexure due to tides, G is the true grounding line location, point I_b is the break-in-slope, point I_m is the local minimum in topography, and point H is the hydrostatic point where the ice first reaches approximate hydrostatic equilibrium. The coupling line (C) is the first break in slope seen in ice plains.

in addition to containing the grounding line, also consists of several other features. Fig. 2.2 a shows the features of a typical grounding zone configuration. The grounding line G, is the point where the ice shelf just lifts off the bedrock. Although only the ice shelf is directly affected by the tides, elastic properties also cause the ice sheet to experience vertical displacement extending from a few metres up to hundreds of metres inland from the grounding point. This point (F) is termed as the hinge line (Rignot 1996), (Rignot et al. 2011).

In an idealised configuration, a break in the slope (I_b) is seen in between G and elevation minimum (I_m). However, in the presence of ice plains (grounded ice adjacent to the grounding line with low surface slope (Brunt et al. 2011), (Corr et al. 2001)), ice rises and ice rumples (Section 2.1), there can be several breaks in the slope either landward or far seaward of the

grounding line (Fricker et al. 2006), (Fricker et al. 2009), (Brunt et al. 2011) (Fig 2.2 b). In such cases, the first break in slope occurs inland of the grounding line at the coupling point C. H marks the hydrostatic point where the ice shelf is first in hydrostatic equilibrium with the ocean. The grounding zone spans the F - H extent. The width is affected spatially by the slope of bedrock, ice thickness [Schoof, 2007] and temporally by tides and the Inverse Barometric Effect (IBE), which causes a change in the sea level due to variation in atmospheric pressure.

2.2.2 Significance of grounding line location

The stability of the grounding line location (GLL) is an indicator of the balance between ice accumulation and ice discharge. Moreover, the GLL is related to the resistive force provided by ice shelves against ice sheets (Haseloff and Sergienko 2018), (Pegler 2018). While the line is stable against periodic, small-term perturbations due to seasonal effects and tidal forcing, basal melting of ice shelves can trigger a positive feedback effect, causing further retreat (Pattyn 2018). It is important to know the exact position of the grounding line because an evaluation of ice fluxes at a mislocated grounding line could wrongly suggest either a reduced melting rate of ice shelves or reduced ice flow from the interiors to the ocean [Rignot et al., 2011].

2.3. Synthetic Aperture Radar (SAR)

2.3.1 Working principle

The main components of a satellite based SAR instrument are a transmitting and a receiving antenna. Complementary to a conventional Radar system, microwave pulses are transmitted and the corresponding echoes scattered by objects on the ground are received. The distance between the antenna and the scatterers is calculated by measuring the time taken for the echoes to return. A frequency modulated carrier wave called the chirp signal is transmitted with a frequency defined by the Pulse Repetition Frequency (PRF). The received echo is a delayed and attenuated version of the transmitted pulse that is correlated with a replica of the transmitted signal (Rosen et al. 2000), (Bamler 2000).

Due to its side looking geometry, the energy of the echoes maps to a two dimensional image of complex numbers (Fig. 2.4). The dimensions are the along-track (flight direction or 'azimuth') and the perpendicular to flight direction ('slant range'). The area on the ground covered by each pulse is referred to as the antenna footprint (shown in bright orange in Fig. 2.3). The range resolution is given by:

$$\delta_R = \frac{c}{2B} \tag{2.1}$$

where c is the speed of light and B is the bandwidth of the transmitted chirp. The azimuth resolution is a function of the antenna beamwidth, which, with the assumption of far-field radiation is $\delta_x = R\beta = \frac{R\lambda}{L}$, where λ is the wavelength of the transmitted wavelength, L the antenna length and R the range distance. In order to achieve a reasonable azimuth resolution (few metres), for a fixed R (typical Low Earth Orbit (LEO) altitudes: 550 km - 850 km) and λ (3 cm - 60 cm), the antenna would have be impractically large. The SAR system makes use of 'coherent' processing, wherein the complex numbers in the generated data matrix are combined in manner so as to build a 'virtual' aperture, having a length much greater than the physical antenna length. This results in the azimuth resolution being equal to $\frac{L}{2}$ (accounting for two-way travel) (Bamler 2000). This is referred to as synthetic aperture, indicated in Fig. 2.3 and Fig. 2.4.

The scattered energy captured by the SAR sensor is a summation of individual scatterers that are present in a single resolution cell. This yields a noisy looking SAR image and is termed as speckle. For a large number of scatterers, the response is described as a circular Gaussian process (Bamler 2000).

SAR has several advantages over optical sensors. Operating in the microwave region of the spectrum, SAR sensors are active sources (self illuminating) and can penetrate clouds and (partially) soil, snow and canopy. The viewing geometry of SAR introduces distortions, distinguished as foreshortening, layover and shadow effects (Bamler 2000). Foreshortening causes steep terrain that is sloped towards the sensor to appear squeezed and slopes tilted away

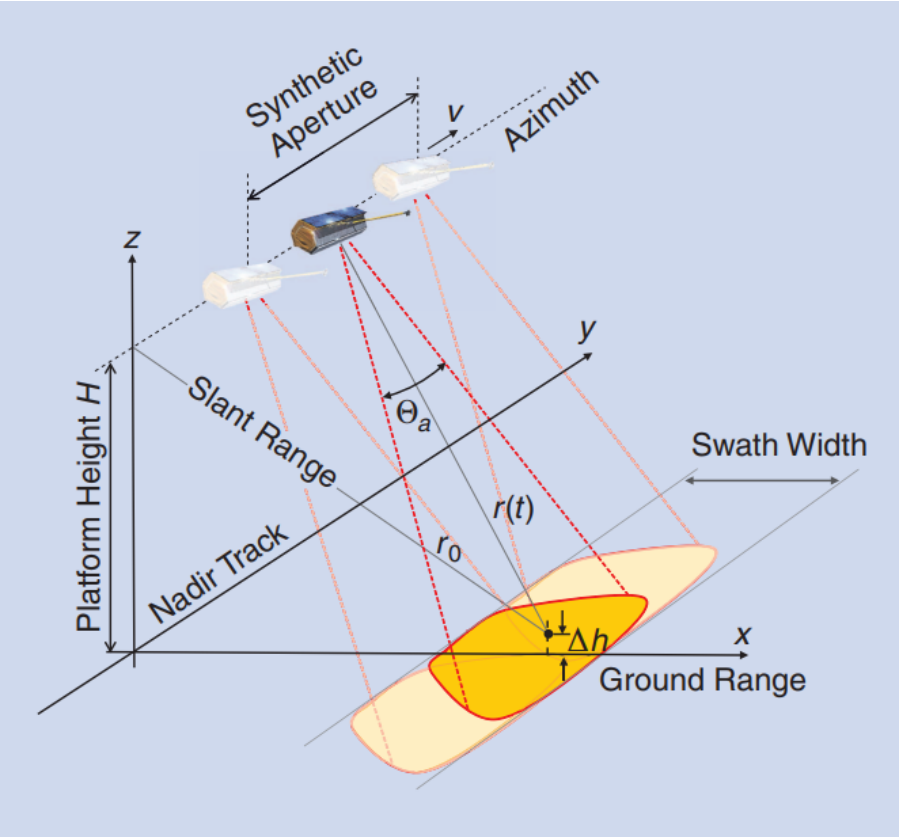


Figure 2.3: Illustration of SAR geometry (Moreira et al. 2013)

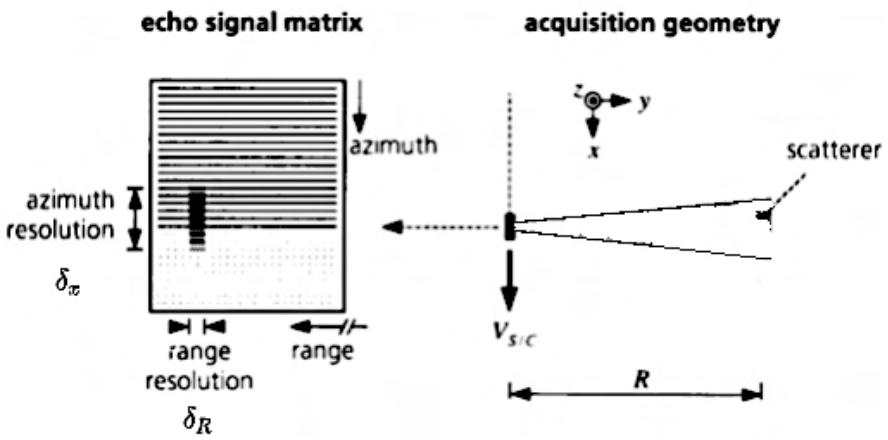


Figure 2.4: Filling the data matrix (adapted from (Bamler 2000))

from the sensor to appear stretched. Layover occurs when the inclination of the observed terrain is greater than the incidence angle of the SAR antenna, multiple regions get mapped into the same pixel. A shadow is seen when steep terrain is imaged at shallow incidence angles. Fig. 2.5 illustrates these effects.

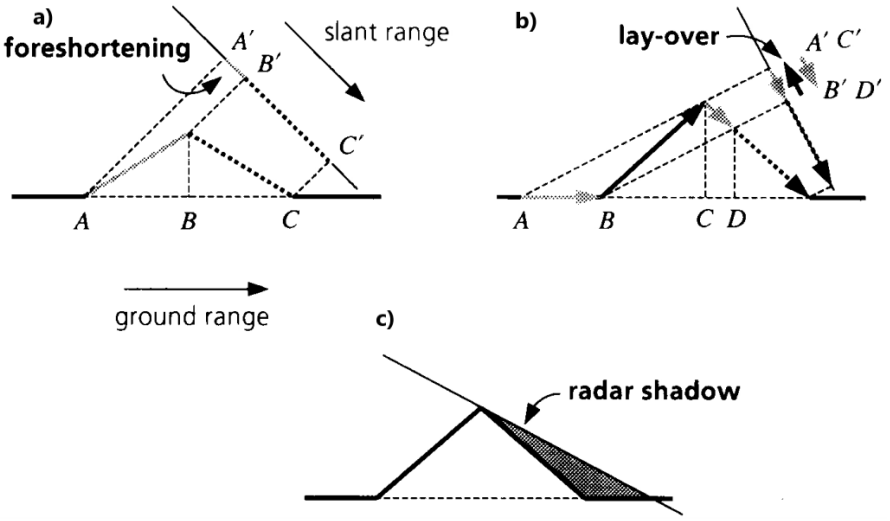


Figure 2.5: Geometric distortions in SAR (Bamler 2000). All letters with ‘ in the superscript show the mappings of the corresponding points on the ground in the SAR image a) shows squeezing of a slope facing the sensor b) illustrates layover, wherein the slope tilted away from the sensor overlaps the slope facing the sensor c) shadow effect as a result of imaging a steep slope from a low incidence angle

While the above mentioned effects are general limitations, they can be partially corrected by using Digital Elevation Models (DEMs) (Shimada 2010). Furthermore, they do not impede the application of SAR in grounding line detection which are usually situated in relatively flat areas. Some specific issues can affect SAR data acquisitions over ice and snow. Dielectric properties of ice and snow along with the wavelength of the SAR signal and the surface roughness influence the extent of signal penetration (Mätzler 1987). However, this does not adversely affect grounding line detection to a large extent either.

2.3.2 Interferometric SAR (InSAR)

SAR images can be interpreted as magnitude (or square of magnitude, intensity) and phase. However, the phase of a single SAR image does not convey much information. In SAR interferometry, the differential phase of two or more SAR images acquired with at least one differing parameter (acquisition times, imaging positions, wavelength) provides additional information to derive topography, subsidence and velocities of moving objects, ocean currents, glaciers and lava flow (Bamler and Hartl 1998), (Moreira et al. 2013). The InSAR configuration

which measures displacement in the slant range of Line of Sight (LOS) is the focus of this section.

Repeat-pass interferometry involves using two SAR images of the same region acquired at different times to extract surface displacement. Ideally repeat tracks would overlap each other exactly, however, there usually exists a separation in the flight paths termed as the baseline (B) and its component perpendicular to the look (range) direction called the perpendicular baseline ($B_\perp$) (Bamler and Hartl 1998) (Fig. 2.6).

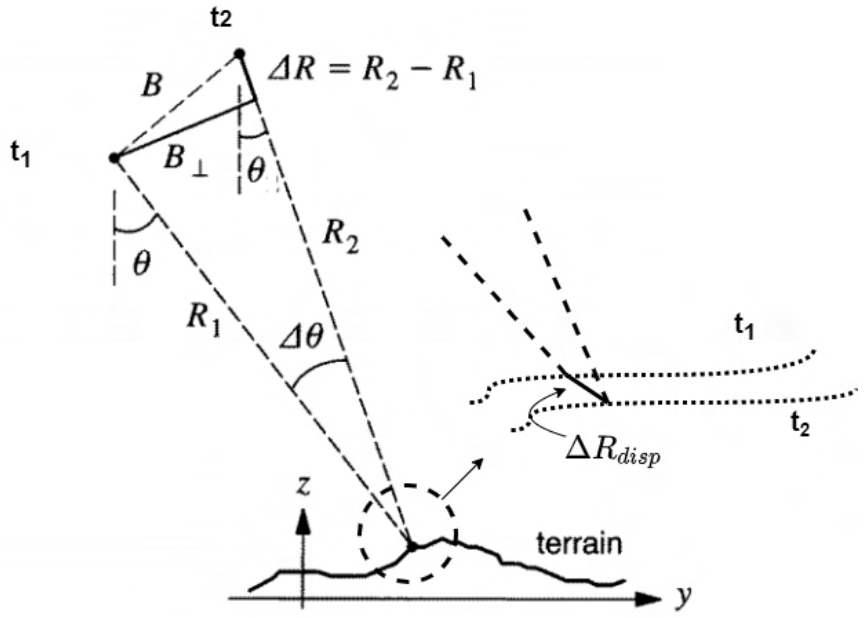


Figure 2.6: Repeat pass interferometry, satellite flight direction is into the plane of the figure (adapted from (Bamler and Hartl 1998))

The LOS deformation is visible in the phase difference which is computed by a complex-conjugate multiplication of the SAR image pair, called the interferogram. Before an interferogram can be formed, the image pair needs to be coregistered. Shift vectors are found by performing patchwise cross-correlations between one of the images (secondary) and the reference (primary) image. Correlation peaks are oversampled to achieve sub-pixel matching. The secondary is then interpolated over the primary image.

The interferometric phase contains the following components:

$$\phi_2 - \phi_1 = \Delta\phi = \Delta\phi_{\Delta R_{disp}} + \Delta\phi_{topo} + \Delta\phi_{atmos} + \Delta\phi_{flat\,earth} \tag{2.2}$$

where ϕ_1 and ϕ_2 are the measured phases at t_1 and t_2 respectively. $\Delta\phi_{\Delta R_{disp}}$ is the change in phase due to displacement of the observed surface element in LOS. In terms of range difference, this is expressed as:

$$\Delta\phi_{\Delta R_{disp}} = \frac{4\pi}{\lambda}\Delta R_{disp} \tag{2.3}$$

λ is the wavelength of the transmitted pulse. The additional phase contributions are due to the topography ($\Delta\phi_{topo}$), side looking geometry ($\Delta\phi_{flat\,earth}$) and atmospheric effects ($\Delta\phi_{atmos}$). $\Delta\phi_{topo}$ and $\Delta\phi_{flatearth}$ can be removed using a DEM. $\Delta\phi_{atmos}$ introduces both phase delays and advances and is a source of error (for differential InSAR applications). Methods of mitigation are discussed in a later part of this section.

$\Delta\phi$ results in 'wrapped' phases because it is ambiguous within integer multiples of 2π. Visually, this forms contour-like patterns called fringes. Absolute phase is determined using phase unwrapping algorithms, detailed in (Rosen et al. 2000).

The quality of an interferogram is expressed in terms of its complex coherence:

$$|\hat{\gamma}| = \frac{|(u_1 u_2^*)|}{\sqrt{\langle|u_1^2|\rangle\langle|u_2^2|\rangle}} \tag{2.4}$$

u_1 and u_2 represent the complex matrices of two SAR images and the angular brackets indicate averaging over windows of specified size. $\hat{\gamma}$ varies from 0 to 1, values close to 1 indicating a temporal stable phase of the target and a visible fringe pattern can be expected in the interferogram.

Phase decorrelation occurs due to several factors. Atmospheric pressure, temperature and water vapour content variation in the troposphere cause phase delays which could lead to measurement errors in the range of decimeters. The total electron content (TEC) in the ionosphere is another influencing factor. These quantities vary at spatial scales (in the order of kilometres) (Zebker 2021) that are larger than the width of the grounding zone and therefore do not play a significant role in grounding line detection.

A major reason for low coherence interferograms is due to the changing position of scatterers between acquisitions, known as temporal decorrelation. Snowfall, surface melt, snow drift (due to wind erosion and deposition), and deformation (shear processes and large ice movements) in regions of fast flowing ice can result in noisy fringes, making it difficult to unwrap phases or even leading to complete decorrelation in some cases.

2.3.3 Grounding line detection with DInSAR

Differential Interferometric SAR detects tidal deformation which occurs at the hinge line (Fig. 2.2 a). Two pairs of interferograms are subtracted to remove phase contributions from topography and horizontal ice motion, assuming steady flow. The double difference can either be formed using three or four SAR images of different epochs. Following Eqn. 2.2 (not accounting for atmospheric effects, IBE and sensor biases), two interferograms between images 1, 2, 3 acquired at times $t_1 < t_2 < t_3$ are:

$$\phi_{12} = \phi_2 - \phi_1 = \phi_{12,topo} + \phi_{12,ice\,velocity} + \phi_{12,tidal\,deformation} \tag{2.5}$$

$$\phi_{32} = \phi_2 - \phi_3 = \phi_{32,topo} + \phi_{32,ice\,velocity} + \phi_{32,tidal\,deformation} \tag{2.6}$$

The ice velocity and tidal deformation are the two components of LOS displacement. The velocity and tidal components are expressed as (Rignot 1996):

$$\phi_{ab,ice\,velocity} = \frac{4\pi}{\lambda}(-V_x sin\theta + V_z cos\theta)(t_b - t_a) \tag{2.7}$$

$$\phi_{ab,tidal\ deformation} = \frac{4\pi}{\lambda}(z_b - z_a)cos\theta \tag{2.8}$$

where V_x and V_z are velocity components resolved along LOS. V_y is the azimuthal component and therefore not captured. θ is the incidence angle with respect to the vertical. t and z refer to acquisition time and elevation respectively. Removal of the horizontal deformation component is dependent on the order in which the SAR images are acquired and the temporal baseline. Fig. 2.7 a and Fig. 2.7 b show the possible combinations for images with the constant temporal separation, i.e. $t_2 - t_1 = t_3 - t_2$.

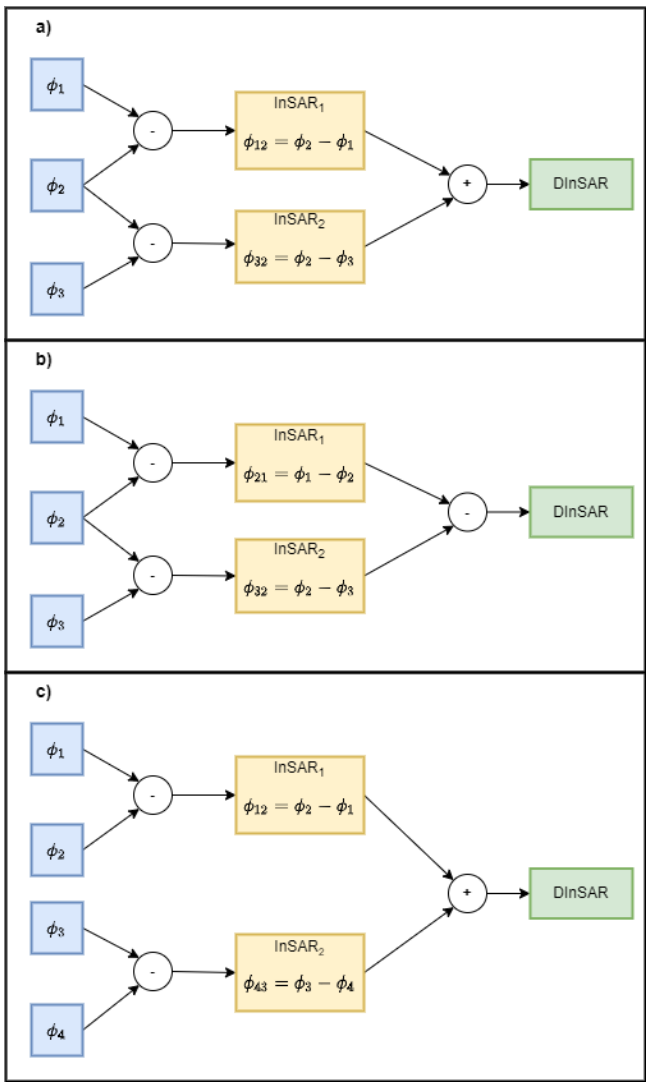


Figure 2.7: Formation of double difference interferograms from a), b) three and c) four SAR acquisitions. The data are acquired in sequence, $t_1 < t_2 < t_3 < t_4$

After removing topographic phase with a DEM, the double difference contains phase changes due to tidal displacement (Rignot 1996):

$$\phi_{12,\,tidal\,deformation} + \phi_{32,\,tidal\,deformation} = \frac{4\pi}{\lambda}(2z_2 - z_3 - z_1)cos\theta \tag{2.9}$$

Tidal deformation is seen as dense fringes in the double difference interferogram (Fig 2.8). The most inland interferometric fringe of this belt marks the landward extent of the ice deformation (bending) due to tidal motion and therefore the location of the hinge line. Many DInSAR based GLL datasets (Section 3.1.2) employ this visual feature and manually digitise the GLL on the double difference interferogram. In certain circumstances InSAR can also be used to map the GLL (Rignot 1996), (Rignot 1998), (Rignot et al. 2011), (Rignot et al. 2014), (Milillo et al. 2022). The elevation profile obtained after phase unwrapping is fitted to modelled tidal displacements based on elastic beam theory (Holdsworth 1969). However, this process is computationally intensive and requires high coherence interferograms.

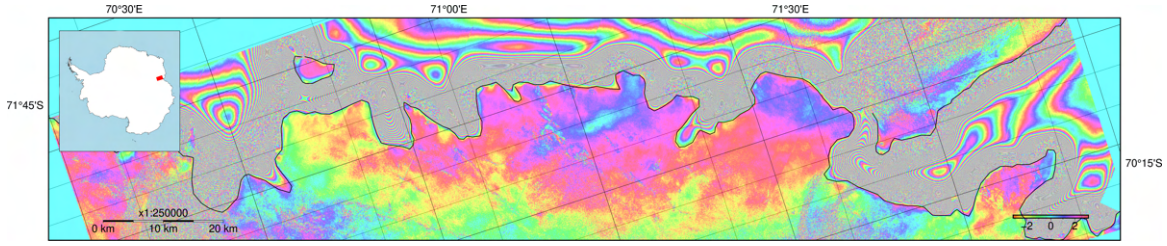


Figure 2.8: Double difference interferogram formed from 3 consecutive Sentinel-1 SAR acquisitions, with temporal baseline of six days. The red rectangle in the inset shows the location of the Amery Ice Shelf. The grounding line (black) is from the AIS_cci project (Horwath and Groh 2018)

2.4. Machine Learning

Machine learning (ML) algorithms are a set of automated methods (with a basis in statistics) that find patterns in data and use this information to make further predictions or decisions (Murphy 2012, 1.1). Relative to non-ML techniques, ML methods find a representation for large amounts of data without the need for programmed instructions or an explicit framework.

The task of extracting information from data by a ML algorithm is termed as 'learning' and the type of data used for this task defines the learning problem. While there are several categorizations of learning problems (specified in Murphy 2012, 1.1.1), supervised learning (learning the mapping from input data samples to their corresponding labels) is pertinent to this study and is the basis for further content in this section.

The ability of ML algorithms to automate everyday tasks have led to them being integrated into modern day technologies. They are also increasingly being used for processing voluminous remote sensing data (Zhu et al. 2017). Deep learning (DL) networks (a subset of ML algorithms) in particular have been gaining popularity since the last decade. These methods have been applied to scene classification (Minetto et al. 2019), (He et al. 2019), object detection (Deng et al. 2018) and semantic segmentation (Tasar et al. 2019) tasks and have been shown to outperform conventional techniques.

Section 2.4.1 provides an overview of deep learning. The specific problem of semantic segmentation or pixel wise classification and the popular solution to it, Convolutional Neural Networks (CNN) are discussed in Section 2.4.2.

2.4.1 Deep Learning

Deep learning methods (or deep neural networks (DNN)) are characterised by several layers of representations of the input data. Each data representation is obtained by mathematical transformation of the input by a collection of basic units called neurons (Fig. 2.9). The concept of the neuron was derived from a mathematical interpretation of human brain cells (McCulloch and Pitts 1943). The neuron performs a nonlinear transformation (via activation function) on a weighted linear combination of the inputs and a constant bias:

$$a = \sigma(\sum_{i=1}^{n} w_i x_i + b) \tag{2.10}$$

The activation function constrains the output to a desired range of values. Fig. 2.12 illustrates commonly used activation functions. (Lederer 2021) provides an overview of

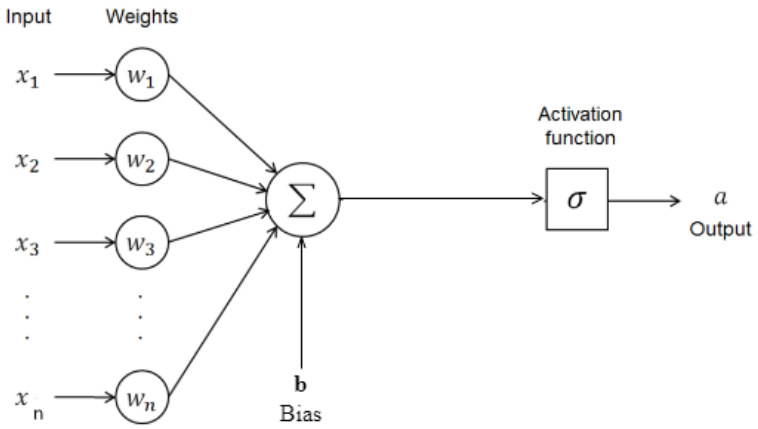


Figure 2.9: Artificial neuron, the basic unit of every DNN. Inputs x_i to an ANN are weighted (with weight w_i), summed and nonlinearly transformed (via activation function σ), following Eqn. (2.10)

different activation functions and their applications. A DNN typically contains two or more collections of stacked neurons (in 'hidden' layers) between the inputs and outputs (Fig. 2.11). The weights and biases of all the neurons are adjusted simultaneously during training. Training involves allowing the network to inspect the data multiple times, with small adjustments being made to the network parameters by an optimization algorithm after every iteration. These adjustments are made with the aim to reduce the error or loss (quantified by means of a 'loss function') between network predictions and expected outcomes (Fig. 2.10).

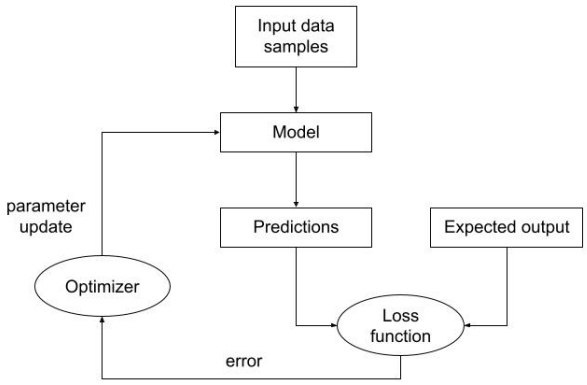


Figure 2.10: Training workflow for supervised learning tasks aiming at reducing errors by minimizing the loss function (adapted from (Chollet 2021))

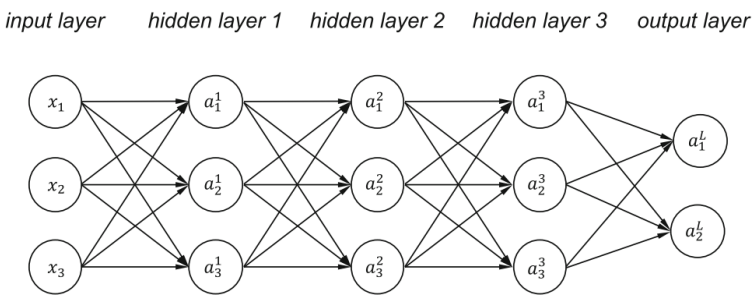


Figure 2.11: Structure of a simple deep neural network with three hidden layers. a_l^i represents one neuron with l referring to the layer index and i the neuron index (Dong et al. 2020)

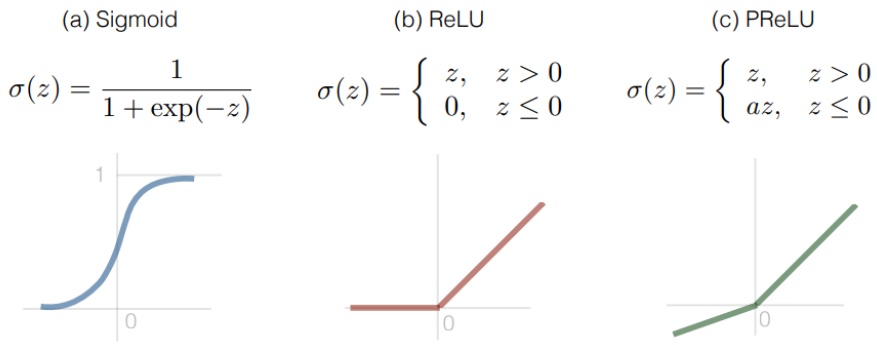


Figure 2.12: Popular activation functions a) Sigmoid b) Rectified Linear Unit (ReLU) c) Parametric or leaky ReLU (PReLU) ((Pang et al. 2018))

The optimizer essentially computed the gradient of the loss function with respect to the weights and uses them to adjust the weights vector in a direction opposite to the gradient vector (LeCun et al. 2015). The gradients are calculated using a method called backpropagation (Rumelhart et al. 1986). There are several optimization techniques used for loss minimization (Choi et al. 2019).

Unlike traditional ML techniques which contain relatively few representation layers, tens or hundreds of successive data transformations in multilayer DNNs yields a variety of features, enhancing their ability to learn complex functions. Additionally, these networks are end-to-end models, i.e., they do not require a handcrafted feature extractor (LeCun et al. 2015).

2.4.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a variant of DNNs that are primarily used to process multi dimensional data (LeCun et al. 2015). The architecture is characterised by convolutional layers. Convolution here refers to the dot product between a convolution kernel and inputs, which occurs patchwise as the kernel moves across the input tensor with a defined stride (Fig. 2.13). This provides the advantages of having fewer trainable weights (in comparison to standard fully connected layers of a DNN) and exploitation of local, translation invariant patterns (Chollet 2021 8.1.1).

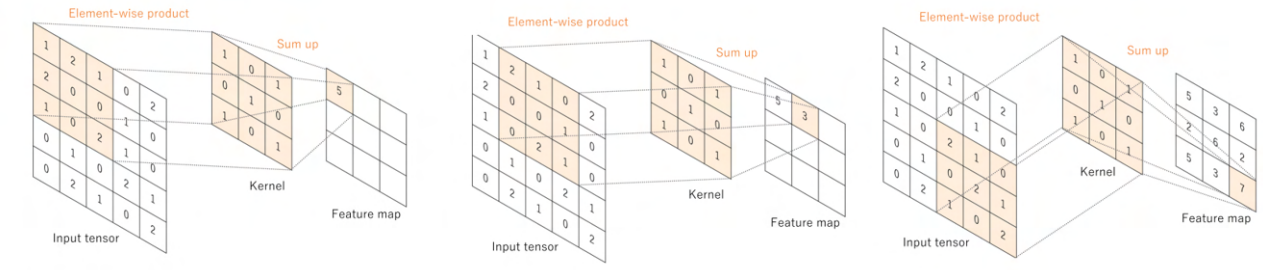


Figure 2.13: An example of convolution operation with a 3x3 kernel, no padding and stride of 1 (adapted from (Yamashita et al. 2018))

Convolutional layers are typically followed by a dimensionality reduction operation known as pooling (Fig. 2.14). The aim of pooling is to merge semantically similar features (LeCun et al. 2015). A typical CNN architecture consists of several blocks of convolutions, nonlinearity and pooling or subsampling followed by a final fully connected layer (Fig. 2.15). Features learnt by the convolutional layers essentially form a feature hierarchy, enabling the network to learn complex and abstract patterns (Chollet 2021, 8.1.1), (LeCun et al. 2015). These aspects have led to extensive and successful usage of CNNs for several tasks.

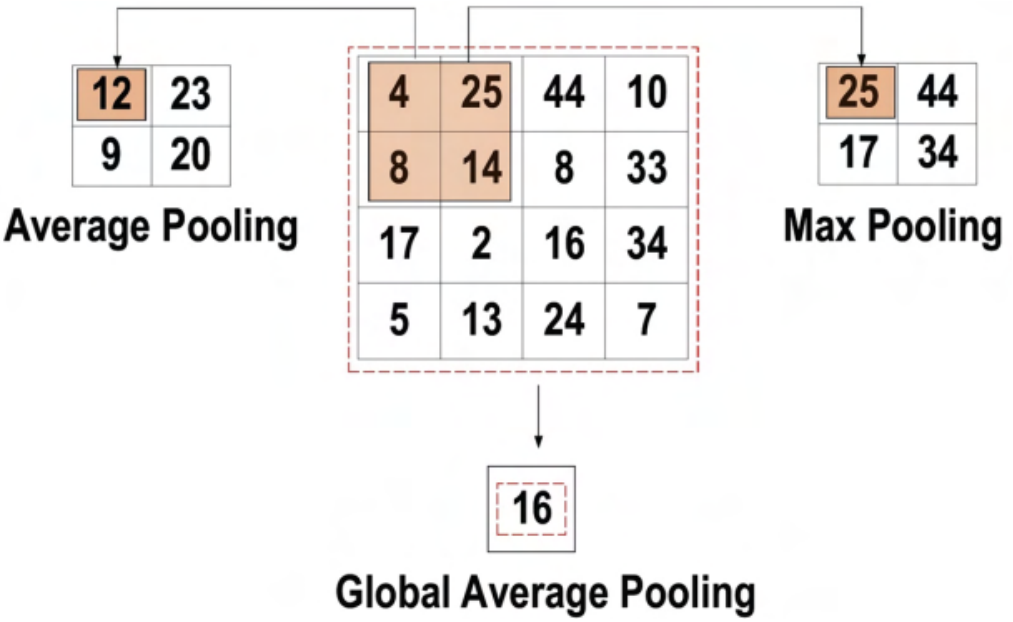


Figure 2.14: Schematic illustrating 3 different pooling operations which are typically used after convolution operation (Figure 2.13) in order to reduce the dimensionality.

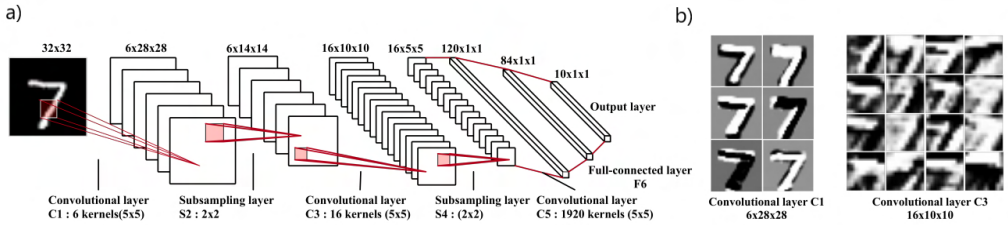


Figure 2.15: Example of a typical CNN architecture a) LeNet-5 (LeCun et al. 1989) used for handwritten digit recognition b) visualisation of features extracted by intermediate convolutional layers. Figure adapted from (Gu et al. 2018).

3. State of the art

This chapter outlines state of the art techniques in detecting grounding lines and existing grounding line datasets (Section 3.1) and machine learning architectures typically used both for grounding line delineation and similar tasks such as calving front detection (Section 3.2).

3.1. Overview of grounding line detection techniques

3.1.1 Detection methods

The techniques mentioned here follow the classification specified in the review paper from (Friedl et al. 2020). The purpose of this section is to provide an overview of existing methods and hence does not discuss equations, minute details or all available literature connected to them. Fig. 3.1 shows an overview of the methods. Only in-situ and remote sensing techniques are discussed in this section.

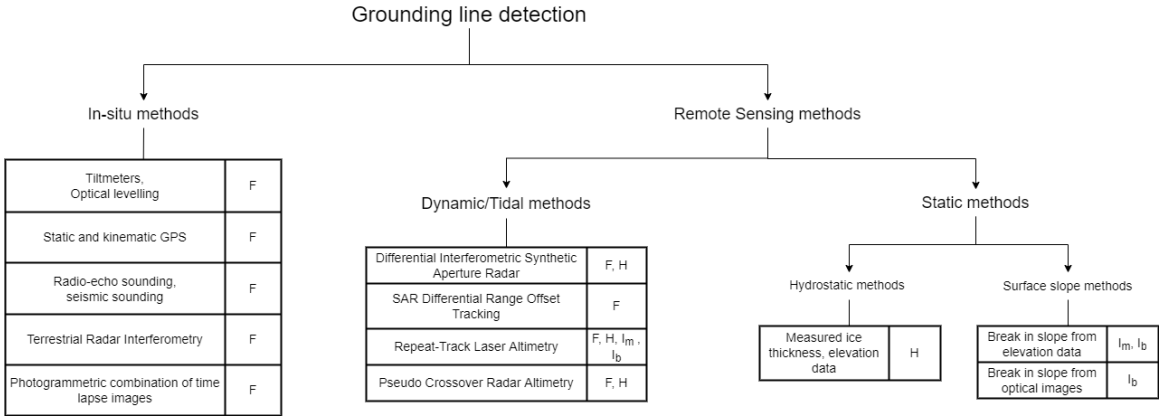


Figure 3.1: Classification of grounding line detection techniques (adapted from (Friedl et al. 2020))

1. **In-situ techniques**

   Field measurements involve using a variety of instruments to derive geophysical parameters such as ice velocity, tidal displacement, surface and bedrock elevation. (Smith 1991) used tiltmeters to determine change in surface slope due to tidal displacement. Along with ice velocity information, surface and bedrock elevation derived from optical levelling and seismic sounding equipment, respectively, was used to model a part of the Ronne Ice Shelf's bending as a two dimensional elastic beam. Tidal components were extracted from tilt measurements using Fourier analysis. The landward limit of ice sheet flexure (F in Fig. 2.2) was obtained from the model.

In addition to tiltmeters, (Riedel et al. 1999) used gravity metres to measure vertical deflection due to ocean and earth tides. Kinematic Global Positioning Systems (GPS) were used to measure overall ice motion, from which ice velocities were derived. Similar to (Smith 1991), Fast Fourier Transform (FFT) was applied to obtain phases, amplitudes and frequencies of tides. While the above mentioned instruments are accurate and are able to locate F within 500m or better (Smith 1991), data gaps occurred due to power failure and downloading errors (Riedel et al. 1999). Model assumptions do not hold in all locations due to different grounding zone configurations.

Using a processing method similar to InSAR (Section 2.3.2), LOS velocities and an elevation time series were generated using Terrestrial RADAR Interferometry (TRI) by (Xie et al. 2018). These parameters were used to find tidal components and hence find the boundary between grounded and floating ice. TRI is sensitive to displacements of 1 mm in LOS (Ku band) and is independent of weather and lighting conditions. Setting up the radar is, however, a difficult task and could introduce errors due to mistakes in positioning the equipment.

Ice trajectory vectors in the study from (Rosenau et al. 2013) were computed by photogrammetrically combining terrestrial imagery. Surface features such crevasses were tracked in the images and coregistered using cross correlation. The vertical components of the trajectory were cross correlated with tidal amplitudes derived from tide models. Free float values (ratio between tidal amplitude and vertical component of the trajectory at an epoch) were used to delineate the grounding line, with values close to 1 indicating a good coincidence of vertical ice movement and tides. The delineation had an accuracy of 100m.

Apart from the expense, logistical difficulties and risks involved in collecting measurements in polar environments, in-situ techniques are also limited in terms of spatial and

temporal resolution. Still, such field campaigns are very useful when focused on validation of ice deformation models because the temporal resolution of such measurements is comparable to that of the tide models and thus they can capture entire tidal cycles (Wild et al. 2019).

2. **Remote Sensing techniques**

   a) **Tidal/Dynamic methods**

   Tidal methods detect the vertical deformation that occurs at the hinge line due to tidal forcing.

   Extraction of grounding line location from one single interferogram is limited to regions where ice movement is sufficiently slow, i.e., contribution from horizontal ice movement is negligible compared to vertical deformation from tides. This also implies that the fringe pattern characteristic to vertical deformation at the flexure point, similar to the one visible in Fig. 2.8, is distinguishable in high coherence interferograms. This was demonstrated in the study from (Goldstein et al. 1993) using ERS-1 interferograms with a 6 day temporal baseline. Horizontal velocity components were accounted for by using velocity estimated from ground based observations.

   As explained in Section 2.3.3, F is detectable in double difference interferograms formed with Differential Interferometric SAR (DInSAR) under the assumption of constant horizontal ice velocity. While DInSAR is advantageous with respect to spatial coverage, accuracy (<1cm) (Goldstein et al. 1993), (Rignot et al. 2011), capturing short term grounding line variation and independence to weather and illumination conditions, acquiring interferograms with high coherence is impeded by decorrelating factors such as snow accumulation, snow drift, melting events and fast glacier flow. The assumption of steady ice flow does not hold true for outlet glaciers where changes in ice velocity induced by tides results in an additional range

component. The grounding lines manually delineated on double difference interferograms have a reported positioning accuracy of 100 m in the study of (Rignot et al. 2011) and 250 m in dataset compiled by (Horwath and Groh 2018).

The SAR Differential Range Offset Tracking (DROT) method overcomes the need for high coherence interferograms. Sub-pixel offsets in azimuth and range directions are estimated in a pair of coregistered SAR intensity images by cross correlating with speckle patterns, also known as speckle tracking (Joughin 2002). Similar to DInSAR, a double difference of (triple or quadruple) offsets is generated. Range displacements contain real horizontal motion and a bias from tidal displacement which is introduced due to the side looking geometry of SAR. The horizontal components (assuming steady flow) and equal vertical biases are cancelled in the double differenced fields. The remaining displacement associated with tidal flexure is a function of the incidence angle (Marsh et al. 2013). Although independent of coherence, DROT is less sensitive to vertical displacement and hinge lines derived using DROT are 2 km inland of those derived using DInSAR (Joughin et al. 2016). Since the tidal bias is only present in the range offsets, glaciers that flow in the azimuth direction cannot be considered. Additionally, the technique cannot differentiate between displacement caused by tidal and non tidal ocean processes such as storm surges, ocean eddies and Inverse Barometric Effect (Rignot et al. 2000).

Another method that detects the hinge line is repeat track laser-altimetry (RTLA). A mean elevation profile is computed from several repeat pass measurements. Temporal changes in elevation ('elevation anomalies') are tracked by differencing each elevation profile from a mean elevation. F, H and I_b are determined from visual inspection of the profiles. The elevation changes are validated with tidal models (Brunt et al. 2010), (Fricker et al. 2009), (Brunt et al. 2011). The standard deviation between RTLA based grounding line delineation from Ice, Cloud and

land Elevation Satellite (ICESat) and DInSAR measurements was measured to be 0.9 km (Dawson and Bamber 2017). However, the spatial resolution is poor due to large distances (5-20 km) between ground tracks. Noise introduced by surface roughness and attenuation from cloud cover limit the number of valid measurements (Brunt et al. 2010).

The study from (Dawson and Bamber 2017) adapts the Pseudo Crossover Radar Altimetry (PCRA) technique from (Wouters et al. 2015) to find F and H. The method uses radar altimetry measurements and modelled tidal amplitudes to fit a bilinear equation that describes the elevation at each point. The equations are solved simultaneously for topography and tidal contribution, which is represented by a dimensionless quantity called T_d, which varies from 0 (no correlation) to 1 (complete match to modelled tides). Given the spatial coverage of CryoSat-2 measurements, the PCRA method could potentially estimate nearly all the grounding lines in Antarctica. The standard deviation between grounding lines derived from CryoSat-2 and DInSAR (derived from AIS-cci and MEaSUREs datasets 3.1.2) is 1.1 km. While computationally effective, PCRA is less sensitive to small tidal amplitudes. Measurements are limited in areas of steep slope and lower latitudes due to large track spacing.

b) **Static methods**

Static methods are based on a single gridded field, either a satellite image product or a DEM (Brunt et al. 2010).

i. Hydrostatic methods

Hydrostatic methods estimate the position of H. The technique determines the ice thickness required for an ice shelf to be in hydrostatic equilibrium with the ocean. Ice thickness is estimated from ice surface and ice bed elevations above mean sea level (orthometric elevation or freeboard height) (Fricker et al. 2002). The observed ice thickness and assumed densities of ice, water and firn are used to determine the theoretical freeboard height of ice in hydrostatic equilibrium.

Where the difference between the two heights is close to zero, ice is considered to be floating. The first location where there is a deviation between the values would be slightly landward of the hydrostatic point. The elevation measurements are obtained from DEMs or interpolated from ice thickness maps. Some studies also used airborne radio-echo sounding and LiDAR measurements for specific regions in Antarctica and Greenland, enabling the capture of temporal variation of the grounding line (Friedl et al. 2020).

This method is computationally efficient. However, the accuracy of locating H is dependent on the variables in the buoyancy equation which vary locally (e.g. ice and firn density). Since H can be kilometres away from the grounding line, a small error in these input variables can greatly affect the estimated location. The spatial resolution is limited to that of the available datasets.

ii. Surface slope methods

One way to identify the break in slope is to derive it from elevation data measured using satellite or airborne radar or laser altimetry. While the elevation profiles provide location estimates for single timestamps, a more continuous coverage can be obtained using DEMs. If DEMs are a compilation of data from different years they cannot be used to detect fast moving variations. A major advantage of this technique is the speed and relative ease of computation. Using the break in slope as a proxy for grounding line location is not always viable, especially in fast flowing regions, where the grounding line positions can vary by more than 100 km from DInSAR estimates (Friedl et al. 2020). Where ice plains are present, the coupling line C (Fig. 2.2 b) could be misinterpreted as a break in slope.

Break in slope is also identifiable from optical imagery by a technique known as photoclinometry. The photoclinometric equation relates imaging geometry, surface illumination, radiance captured by the sensor and reflectance properties

of the surface to surface slopes (Bindschadler and Vornberger 1994). This relation shows the break in slope as a change in shading in optical images. The derived grounding line locations agree well with DinSAR locations where the slope is sufficiently steep and ice is slow moving. Similar to the break in slope method based on elevation data, for gentle slopes or fast moving glaciers, the optical data based estimates differ up to 150 km from DInSAR measurements (Rignot et al. 2011). Cloud cover and illumination conditions limit the use of optical imagery in general.

3.1.2 Grounding Line Datasets over Antarctica

This section discusses existing datasets, primarily those that cover a major part of the Antarctic continent.

The Making Earth System Data Records for Use in Research Environments (MEaSUREs) Antarctic grounding line (Rignot et al. 2016) is a dataset based on DInSAR measurements, distributed by NASA National Snow and Ice Data Center (NSIDC). It consists of mapped upper limit of tidal flexure (F) over a long period (1992-2014) from several SAR missions: ERS-1/2 (1992, 1994-1996, 1999, 2000), RADARSAT-1 (2000), ALOS Palsar (2007, 2008), RADARSAT-2 (2009), COSMO SkyMed (2013) and Sentinel-1 (2014). Different operating frequencies, interferometric baselines and revisit frequencies of the missions affect the quality and accuracy of the generated lines. The lines were manually digitized on the interferograms, with an overall standard error of 100 m when validated against multiple mappings, epochs and datasets such as MODIS Mosaic of Antarctica (MOA) (Rignot et al. 2011).

The Antarctic Ice Sheets Climate Change Initiative (AIS_cci) project from ESA develops several datasets related to measuring different climate relevant parameters of Antarctica. Similar to the MEaSUREs product the AIS_cci Grounding Line Location (GLL) product contains DInSAR derived upper limit of tidal flexure (F) derived from ERS-1/2, Sentinel-1 A/B and TerraSAR-X data. The dataset has a temporal coverage from 1992-2021. While

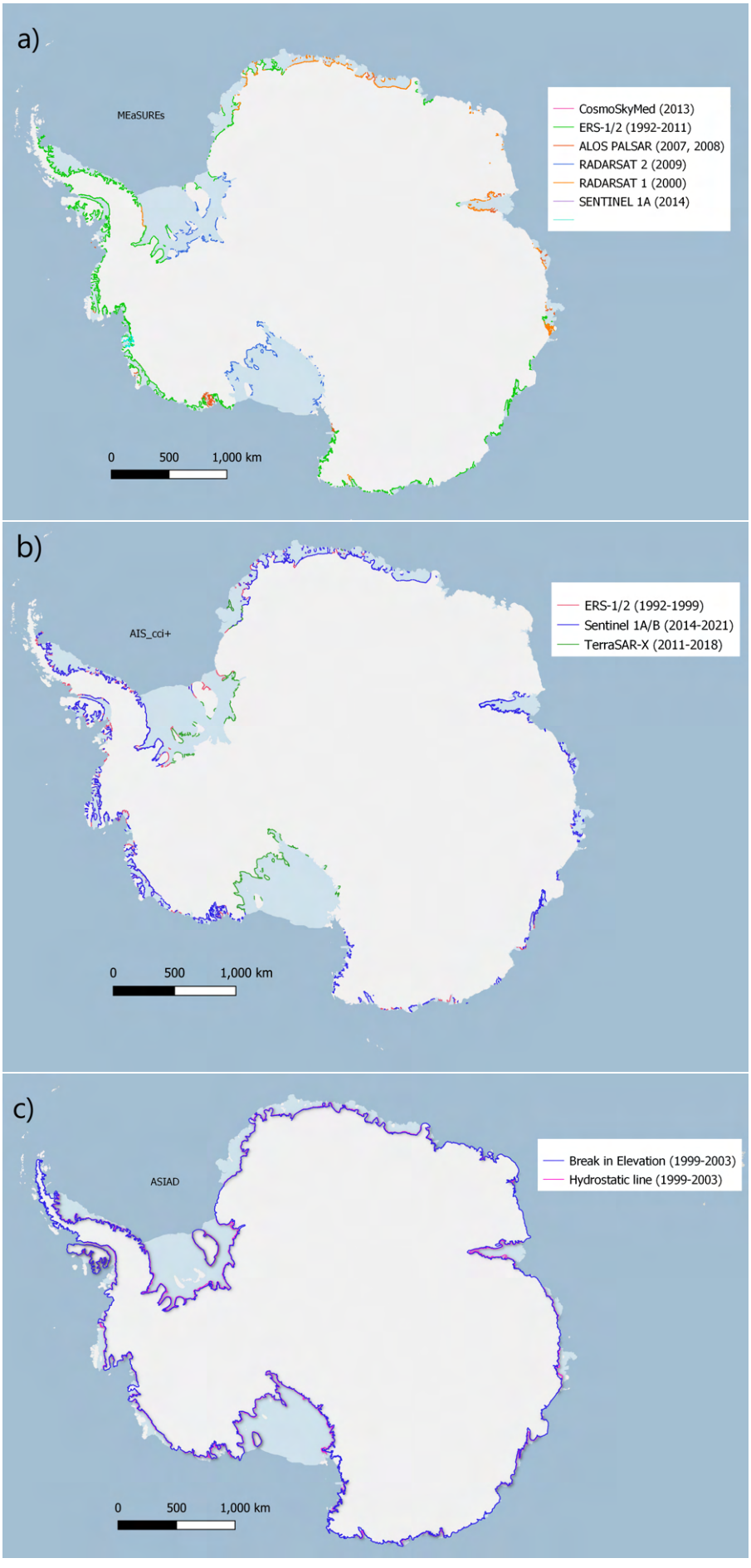
a data gap exists between 2000-2011 (due to lack of missions with appropriate temporal baseline), spatially about 75% of the continent is covered. The AIS_cci GLLs were validated against MEaSUREs, Antarctic Surface Accumulation and Ice Discharge (ASAID) and MOA datasets on selected areas and revealed an average deviation of 0.57 km , 1.2 km and 1.8 km respectively (Horwath and Groh 2018).

MOA (Scambos et al. 2007) and ASAID (Bindschadler et al. 2011) are datasets derived from identifying break in slope (I_b) from optical images. MOA lines were generated by using surface morphology from 260 Moderate-resolution Imaging Spectroradiometer (MODIS) images acquired between November 2003 and February 2004. The grounding lines were manually digitized by following the seaward-most continuous slope break on the permanent coastline in images of the surface morphologies. The spatial resolution is 250 m and the standard error is 250 m. The MOA dataset covers the whole continent and all islands.

Generated using a similar procedure, the ASAID product contains both I_b and hydrostatic point locations (H). Landsat 7 imagery acquired between 1999 and 2003 and ICESat/GLAS laser altimetry from 2003 and 2008 were used. The spatial resolution is 15 m. The positional accuracies of the lines vary from ±52 m to ±502 m. The hydrostatic line is positioned with errors of over 2 km. The ASAID dataset covers the whole continent and three islands.

The ICESat-Derived Grounding Zone for Antarctic Ice shelves contains F, H and I_b, measurements obtained for a period between 2003 and 2009 (Brunt et al. 2010). The dataset contains 1497 Point F, 1470 Point H, and 1493 Point I_b locations. Recently, a newer version of the dataset from ICESat2 measurements with a much higher point density (21346 Point F, 18149 Point H , and 36765 Point I_b) and almost complete coverage was made freely available by NSIDC (Li et al. 2021). Repeat track measurements between 30 March 2019 and 30 September 2020 were used. The processing chain is automated and uses tidal methods (elevation anomalies) to estimate the positions of F and H and break in slope technique to estimate the location of I_b. The F points agree well with DInSAR derived lines (Sentinel-1

data from 2018), with a mean absolute separation and standard deviation of 0.02 and 0.02 km, respectively (Li et al. 2022). Fig. 3.2 shows the spatial distribution of the datasets mentioned above.


a)
MEaSUREs
CosmoSkyMed (2013)
ERS-1/2 (1992-2011)
ALOS PALSAR (2007, 2008)
RADARSAT 2 (2009)
RADARSAT 1 (2000)
SENTINEL 1A (2014)
0
500
1,000 km
b)
AIS_cci+
ERS-1/2 (1992-1999)
Sentinel 1A/B (2014-2021)
TerraSAR-X (2011-2018)
0
500
1,000 km
c)
ASIAD
Break in Elevation (1999-2003)
Hydrostatic line (1999-2003)
0
500
1,000 km

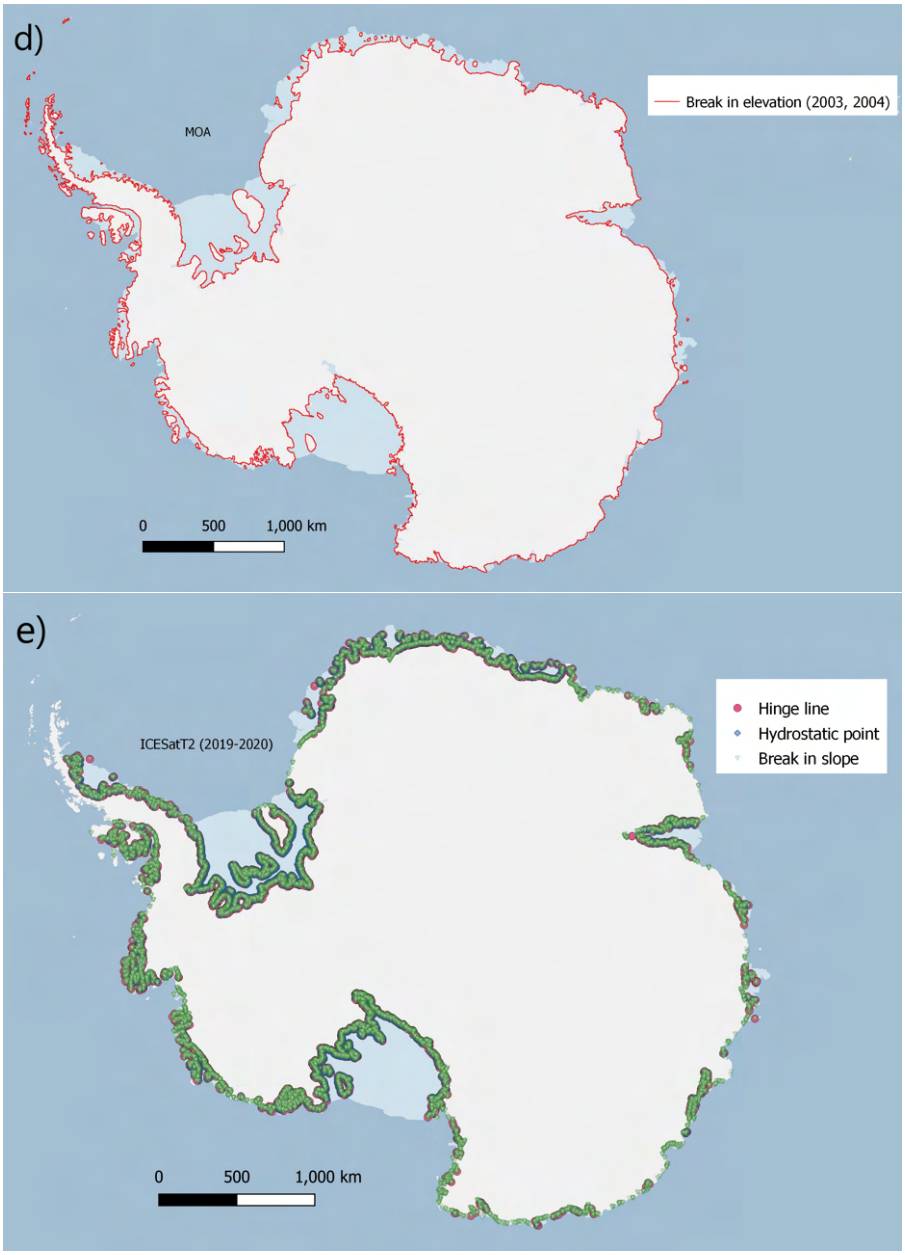


Figure 3.2: Temporal and spatial distributions of grounding line datasets a) MEaSUREs V2 (Rignot et al. 2016) b) Antarctic Ice Sheets climate change initiative (AIS_cci) (Horwath and Groh 2018) c) Antarctic Surface Accumulation and Ice Discharge (ASAID) (Bindschadler et al. 2011) d) Mosaic Of Antarctica (MOA) (Scambos et al. 2007) e) ICESat 2 - Derived Grounding Zone for Antarctic Ice shelves (Li et al. 2021). Plot was made using Quantarctica 3 (Matsuoka et al. 2021)

3.2. Automatic delineation methods

The spatial and temporal extent, number of satellite missions and variety of sensors and volume of data measuring geophysical parameters of polar regions has urged the automation of processing chains. The simultaneous advancements in machine learning algorithms, deep

learning in particular, makes it possible to analyse and process such large datasets. Section 3.2.2 discusses some of the architectures that have been used for grounding line detection.

The automation of coastline and calving front (CF) detection is better explored compared to grounding line delineation due to their visibility in optical as well as SAR images. There exists both classical image processing and, more recently, machine learning techniques for the fast changing calving front. Section 3.2.1 describes the identification issues in brief, remote sensing based detection techniques and enumerates automatic delineation methods.

Coastline and calving front detection could be seen as analogous tasks to grounding line delineation due to their appearance as a thin edge at the borders between large areas of land or grounded ice and open sea or floating ice. In satellite images very few pixels belong to the coastline or calving front or grounding line in comparison to the rest and therefore can be solved using similar solutions.

3.2.1 Calving front (CF) detection

The calving front refers to the boundary between a mass of ice that is either grounded or physically bound to an ice shelf and its environment (open water, sea ice, calved icebergs). In addition to changing the shape of the coastline, ice shelf calving also directly impacts its mass, thereby changing buttressing forces and affecting glacier dynamics terminating into the ice shelf in a significant way (Baumhoer et al. 2018), (Wuite et al. 2019).

Calving fronts are detectable in optical imagery, SAR imagery and DEMs. Image processing based methods exploit the reflective properties of ice and water, identifiable either in multispectral images or backscattering characteristics. (Sohn and Jezek 1999) applied edge enhancement and local thresholding for optical images and found regions of drastic contrast difference in SAR images (texture features), followed by edge detection. The Canny edge detector was applied to SAR intensity images in (Krieger and Floricioiu 2017) to estimate the calving front location between a start and end point. Djikstra's algorithm is then used to find the optimal delineation. Both methods showed good performance (mean difference between

automatic delineation and manual delineation of 200 m and 246 m for Sentinel-1 and 159m for TerraSAR-X respectively), but both were tested on small study areas only.

Similarly, (Seale et al. 2011) used the Sobel operator and a brightness gradient to find edges on cloud free MODIS images. Although the method was applicable to large scale datasets, a manual orientation of images in the direction of glacier flow was required.

Active contours method was used in (Klinger et al. 2011) to iteratively find the best line, given initial calving front locations. This is also one of the drawbacks of the method and required post processing corrections. (Baumhoer et al. 2018) reviews several other semi-automatic and automatic delineation methods.

(Dong et al. 2022) developed an algorithm to automatically delineate calving fronts from DEMs through elevation and roughness based features which enhances the contrast between the ice sheet and the ocean. The method performed well with the TanDEM-X DEM at two posting sizes and with the optical photogrammetry based Reference Elevation Model of Antarctica (REMA), but is limited by the availability and quality of DEMs. The mean deviation from the manually delineated CFs was 14 m with REMA, 20 m with the 12 m and 70 m with the 90 m TanDEM-X DEMs respectively.

Machine learning algorithms for delineation are generally seen as a segmentation problem. Convolutional Neural Networks are widely used due to their ability to glean contextual information from images at various scales. The UNet network was used by (Mohajerani et al. 2019) with preprocessed images from Landsat 5, Landsat 7 and Landsat 8 to delineate glacier fronts in Greenland. The resultant deviations were very close to the manual lines (1.97 pixels on average).

A modified version of the UNet was developed by (Baumhoer et al. 2019) to delineate CFs in Sentinel-1 intensity images that were stacked as four channels: HH and HV polarizations, ratio of HH/HV polarizations and TanDEM-X DEM at 90 m posting. The training images were

manually preprocessed, followed by segmentation into land ice and ocean classes, followed by rasterization of the outputs. They were able to achieve a mean difference of 154 m.

The study by (Cheng et al. 2021) incorporates multi sensor (Landsat 7, Sentinel-1 and TerraSAR-X) data covering glaciers both in Greenland and Antarctica. Called the Calving Front Machine neural network (CALFIN-NN), the network segments pixels into ice and ocean. The delineations deviate on average by 86.76 ± 1.43 m from the measured front.

(Heidler et al. 2021) aimed to reduce misclassification at the boundary between land and ocean by merging an edge detection architecture, HED with UNet (see Section 4.2). Deep supervision and hierarchical attention mechanisms are implemented to merge intermediate, multiscale outputs into a final prediction. Sentinel-1 images of the Antarctic coast were used for training. A mean deviation of 345 ± 24 m was achieved for a test site in the Antarctic Peninsula. In general, the model outperformed several edge detection methods and other deep learning architectures.

3.2.2 Machine learning for grounding line detection

Automatic grounding line delineation is impeded by several obstacles which are absent from calving front delineation. The grounding line is not a surface feature and is instead detected by measuring proxies (Section 3.1). Edge detection algorithms have, so far, not been explored for delineation on either interferograms or optical images.

To date, only one study (Mohajerani et al. 2021) explores extracting grounding line locations from Sentinel-1 double difference interferograms. The architecture was adopted from (Cheng et al. 2021), with real and imaginary components of the interferograms forming the two channels of each input. The network differentiates between grounding line and non grounding line pixels. The network was trained on DInSAR interferograms (with temporal baselines of six and twelve days) of Sentinel-1 scenes for the period of 2018, covering the Getz Ice Shelf. DInSAR interferograms covering the AIS were used to validate the model.

The model was able to achieve a mean deviation of 232 m and a median absolute deviation of 101 m in comparison to manual delineations. They also concluded that the predicted lines are consistent and reproducible, unlike human drawn lines. Overall, the predictions follow the ground truth well, although spurious branches were generated in some regions.

4. Data and methods

This chapter describes the grounding line product, associated features that form the dataset and chosen machine learning architectures.

4.1. Dataset

4.1.1 The AIS_cci Grounding Line Location

The Grounding Line Location (GLL) product of the AIS_cci dataset was used as ground truth in this study. The spatial and temporal extent are mentioned in Section 3.1.2 and pictorially represented in Fig. 3.2. The grounding lines were delineated on double difference interferograms formed from Sentinel-1A/B, ERS-1/2 and TerraSAR-X images.

Sentinel-1A and Sentinel-1B are polar orbiting satellites with SAR payloads which operate in the C-band. For the GLL product, SAR images acquired in the Interferometric Wide Swath (IW) mode with Terrain Observation by Progressive Scans (TOPS) were used. TOPS involves scanning along elevation while also rotating the antenna forward in azimuth. This achieves a wide swath coverage at nearly uniform Signal-to-Noise Ratio (SNR). The burst-wise processing of the images is described in (Yagüe- Martınez et al. 2016). Three consecutive repeat pass acquisitions with a temporal baseline of both 6 days and 12 days were used to generate interferograms.

The European Remote Sensing satellites (ERS) 1 and 2 were also operating at C-band and shared the same polar orbit. Two and three image interferograms with temporal baselines of 1 day and 3 days respectively acquired during the tandem mission were used. X-band data from TerraSAR-X with 11 days temporal baseline were also used. Both ERS-1/2 and TerraSAR-X images were acquired in Stripmap mode. Using interferograms from different satellite missions provides the advantages of a long time series and filling in spatial gaps, as in the case of TerraSAR-X which covers the polar hole not covered by Sentinel-1.

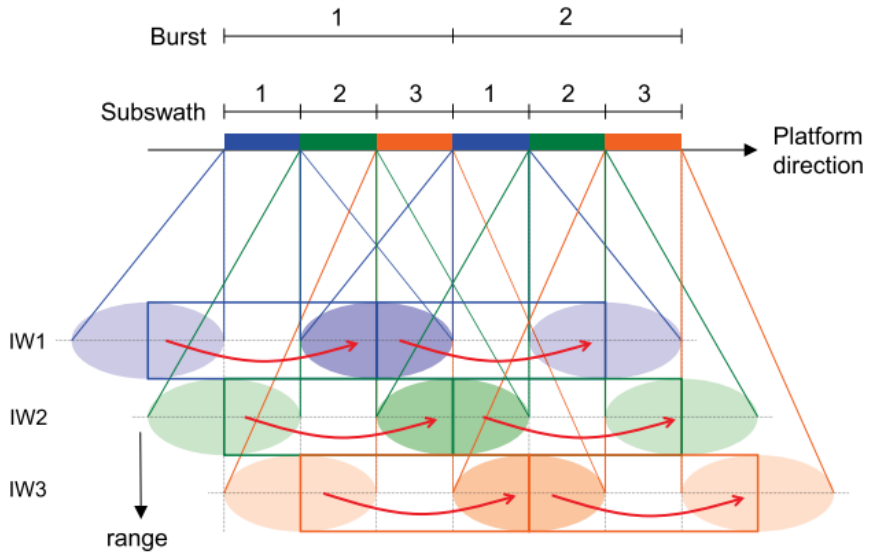


Figure 4.1: TOPS scan pattern for Sentinel-1 (Yagüe- Martınez et al. 2016). The antenna rotation in azimuth is indicated by the red arrows. For burst 1, subswath 1 is acquired first (blue), followed by the second (green) and the third (orange). The antenna steers back to the first subswath and the process repeats for the subsequent bursts.

The AIS_cci GLL dataset contains 11,506 lines delineated on 478 double difference interferograms. Although coherent, not all interferograms were usable either due to missing or corrupted data. For this study, grounding lines from 340 interferograms were used. Table.4.1 shows the distribution of interferograms within the 3 satellite missions.

Table 4.1: Overview of satellite missions. IW - Interferometric Wide Swath, TOPS - Terrain Observations by Progressive Scans, SM - Stripmap. The last column shows the number of double difference interferograms used for this study versus the total processed in the AIS_cci project.

Satellite	**SAR wavelength** [cm]	**Repeat cycle** [days]	**Imaging mode**	**DInSAR interferograms** (used/dataset)
Sentinel 1A/B	C-band (~ 5.5 cm)	6/12	IW, TOPS	199/201
TerraSAR-X	X-band (~ 3.1 cm)	11	SM	13/129
ERS 1/2	C-band (~ 5.6 cm)	1, 3	SM	119/148

Fig. 4.2 shows the simplified processing chain used for the generation of the GLL product (Muir 2021). Following data ordering and download, the images are sorted based on orbit and look direction to determine the repeat pass pairs or triplets that can be formed based on the

required temporal baseline. The interferograms are computed by the established Integrated Wide Area Processor (IWAP) (Gonzalez et al. 2013). The double difference interferograms are generated by the GL preprocessor, followed by the time-consuming manual delineation and creation of ESRI Shapefiles. Metadata such as orbit information, acquisition times, modelled tidal displacement and atmospheric pressure are added as attributes. Detailed list of attributes is available in the product user guide (Groh 2021).

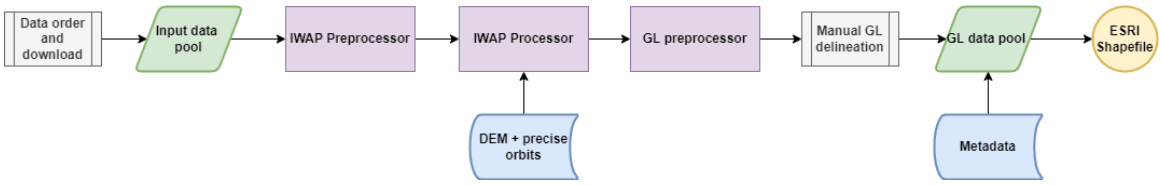


Figure 4.2: Simplified AIS_cci GLL processing chain (Muir 2021). Manual functions shown in grey, intermediate outputs in green, input data/additional information in blue, processors in purple and final product in yellow. Databases not shown.

4.1.2 Training feature stack

In addition to using wrapped phases of the double difference interferograms, in this study six other features are stacked to form the training dataset of the DNN. The grounding line location could potentially also be expressed in these features. They contain information about phase quality, other grounding zone features (break in slope) or physical phenomenon that causes the deformation at the hinge line (tidal amplitudes and atmospheric pressure). Table 4.2 lists the attributes of the selected features.

1. **Wrapped phase and pseudo coherence**

   These are the double difference interferograms on which the grounding lines were manually delineated. The interferograms were resampled to a specific pixel size and reprojected to Polar Stereographic projection (EPSG:3031). In order to preserve cyclic phase variations, the interpolation has to be performed on the real and imaginary components separately. However, due to the absence of corresponding magnitude components,

unit magnitude was added to obtain the complete, complex polar representation of the interferograms. Complex matrices were transformed to their rectangular equivalents on which the resampling operations were performed. The resultant matrices were transformed back to polar form to retrieve the resampled and reprojected phase. This process is illustrated for a sample interferogram in Fig. 4.3. As a consequence of resampling, the resultant magnitude reflects phase stability. As shown in Fig 4.3 g, the magnitude is closer to 1 for slow varying phase and tends to 0 where the fringe frequency is high. A similar quality measure can also be obtained by applying a boxcar filter to the phase or using measured interferometric coherence if available.

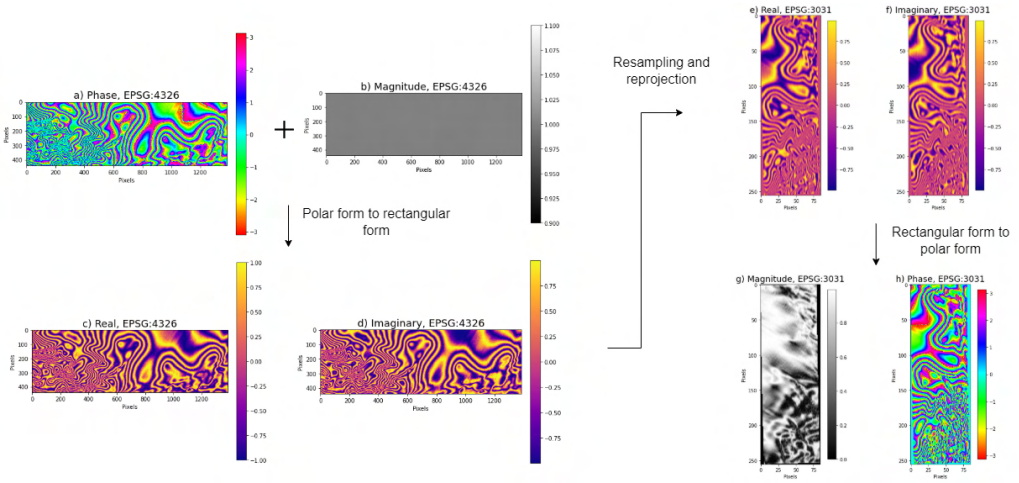


Figure 4.3: Phase preserving reprojection and resampling scheme a) phase with b) added unit magnitude in original projection (EPSG:4326). Complex transformation to c) real and d) imaginary components. This is followed by resampling and reprojection to EPSG:3031 e) real and f) imaginary components. These are transformed back to polar form g) magnitude h) phase. A cyclic colorbar is used for the phase plots to represent -π to π variation. Plot created using Matplotlib version 3.5.0.

2. **DEM**

The 90 m resolution TanDEM-X PolarDEM (Huber 2020) of Antarctica (derived from the global TanDEM-X DEM) was used as a proxy for break in elevation. Acquisitions from April 2013-October 2014 and July 2016-September 2017 were used to compile the DEM.

3. **Ice velocity**

   The ice velocity product from ENVEO IT, generated also within the AIS_cci project, was used to augment the feature stack. Sentinel-1 acquisitions from 2014-2021 were used to create a multi-year averaged ice velocity map in Polar stereographic projection at 200 m resolution. Offset tracking was applied to SAR images to derive velocities along azimuth and LOS. The complete algorithm is detailed in (Nagler et al. 2015).

4. **Tidal displacement**

   Tidal amplitudes at the time of acquisition of the interferogram pairs and triplets were derived from the Circum-Antarctic Tidal Simulation (CATS2008) model (Padman et al. 2008). The model uses assimilated data from several sources, namely tide gauge, GPS, TOPEX/Poseidon radar altimetry measurements, ICESat derived grounding line locations and MOA grounding lines. While the model is gridded at 4 km, pixel-wise interpolated tidal amplitudes were extracted via python APIs provided in the pyTMD module developed by (Alley et al. 2017).

5. **Atmospheric pressure**

   Surface level atmospheric pressure data provided by the National Centers for Environmental Prediction and the National Center for Atmospheric Research, NCEP/NCAR Reanalysis dataset (Kalnay et al. 1996) was used. The dataset was created using a technique called reanalysis which incorporates historical measurements with current observations. The dataset is available as daily values averaged over an year in EPSG:4326 projection with a resolution of 2.5° x 2.5°.

4.2. Methods

Two deep learning models were considered for this study. One is the Holistically-Nested Edge Detection (HED) (Xie and Tu 2015), an architecture that was specifically created to learn boundaries and edges. The model integrates a VGG-Net based fully convolutional neural network (FCN) with deep supervision, wherein intermediate outputs at multiple scales are combined to obtain an edge map. The rationale behind deep supervision is that different con-

Table 4.2: Attributes of input features. CATS2008 assimilates several measurements across various time periods and provides predictions for the required epoch

Feature	**Dataset**	**Projection**	**Resolution**	**Temporal coverage** [years]
Phase	AIS_cci	EPSG:4326	S1A/B: 0.00043° ERS 1/2:0.00055° TSX: 0.00016°	S1 A/B: 2014 - 2021 ERS 1/2: 1992-1999 TSX: 2011-2018
Pseudo coherence/ Noise	Derived from AIS_cci	EPSG:4326	S1A/B: 0.00043° ERS 1/2:0.00055° TSX: 0.00016°	S1 A/B: 2014 - 2021 ERS 1/2: 1992-1999 TSX: 2011-2018
DEM	TanDEM-X PolarDEM	EPSG:3031	90 m	April 2013-Oct 2014 July 2016-Sept 2017
Ice velocity (Vx)	ENVEO IT	EPSG:3031	200 m	2014-2021
Ice velocity (Vy)	ENVEO IT	EPSG:3031	200 m	2014-2021
Tidal amplitudes	CATS2008	EPSG:3031	4000 m	-
Atmospheric pressure	NCEP/NCAR	EPSG:4236	2.5° x 2.5°	1948 -

textual information is captured at different scales, potentially correcting or filling in structures that are missed by successive convolutions.

The second network is UNet (Ronneberger et al. 2015), a popular and 'generic' semantic segmentation approach. It was chosen to compare with the specialised HED network mentioned above, as well as with the existing study of (Mohajerani et al. 2021) that utilised a different semantic segmentation DNN. The network consists of contracting (downsampling) and expansive (upsampling) paths as the left and right arms of the 'U' respectively. Similar to HED, outputs from the convolutional layers of the downsampling path are concatenated to the results of convolutional layers one level below in the upsampling path and fed as inputs to the following layer. The purpose behind this architectural design is that features that were recognised earlier are not lost with progressive convolutions.

4.2.1 Architecture details

The HED network consists of five convolutional blocks with a max pool layer in between. 3x3, padded convolution is applied to all layers. A downsampling by a factor 2 occurs at each maximum pool layer due to the use of a 2 x 2 filter with stride 2. The side outputs are converted to edge maps with a 1x1 convolution and scaled up to the input scale via a deconvolution operation. The side outputs are concatenated and finally fused together by a 1x1 convolution to obtain an array with values ranging from 0 to 1, indicating the probability of each pixel being a grounding line pixel. ReLU activation (Fig. 2.12 b) was applied after each convolution and a sigmoid function (Fig. 2.12 a) was used for the side outputs and final fused output. Fig. 4.4 shows a pictorial representation of the network, with different numbers of convolutional filters for different datasets to overcome memory constraints.

Fig 4.5 illustrates the downsampling/encoder and upsampling/decoder arms of the modified UNet network. Unlike the original network, padded 3 x 3 convolutions are performed to avoid cropping and padding the side outputs. Each encoder block contains two convolutional layers with the same number of filters and a max pooling layer which performs a factor 2 downsampling. The decoder block contains two convolutional layers followed by a deconvolutional layer which performs a factor 2 upsampling. The outputs of convolutional layers of the encoder blocks are appended to the input of the corresponding decoder block. The convolutional layers are ReLU activated, with sigmoid activation for the last 1x1 convolution layer which yields the prediction probabilities.

4.2.2 Loss function

Due to the highly disproportionate distribution of grounding line to non grounding line pixels, the weighted binary cross entropy function [Scott, 2012] was used to compute the loss for both networks. Each class is weighted by the fraction of the other class. With y_{gl} representing

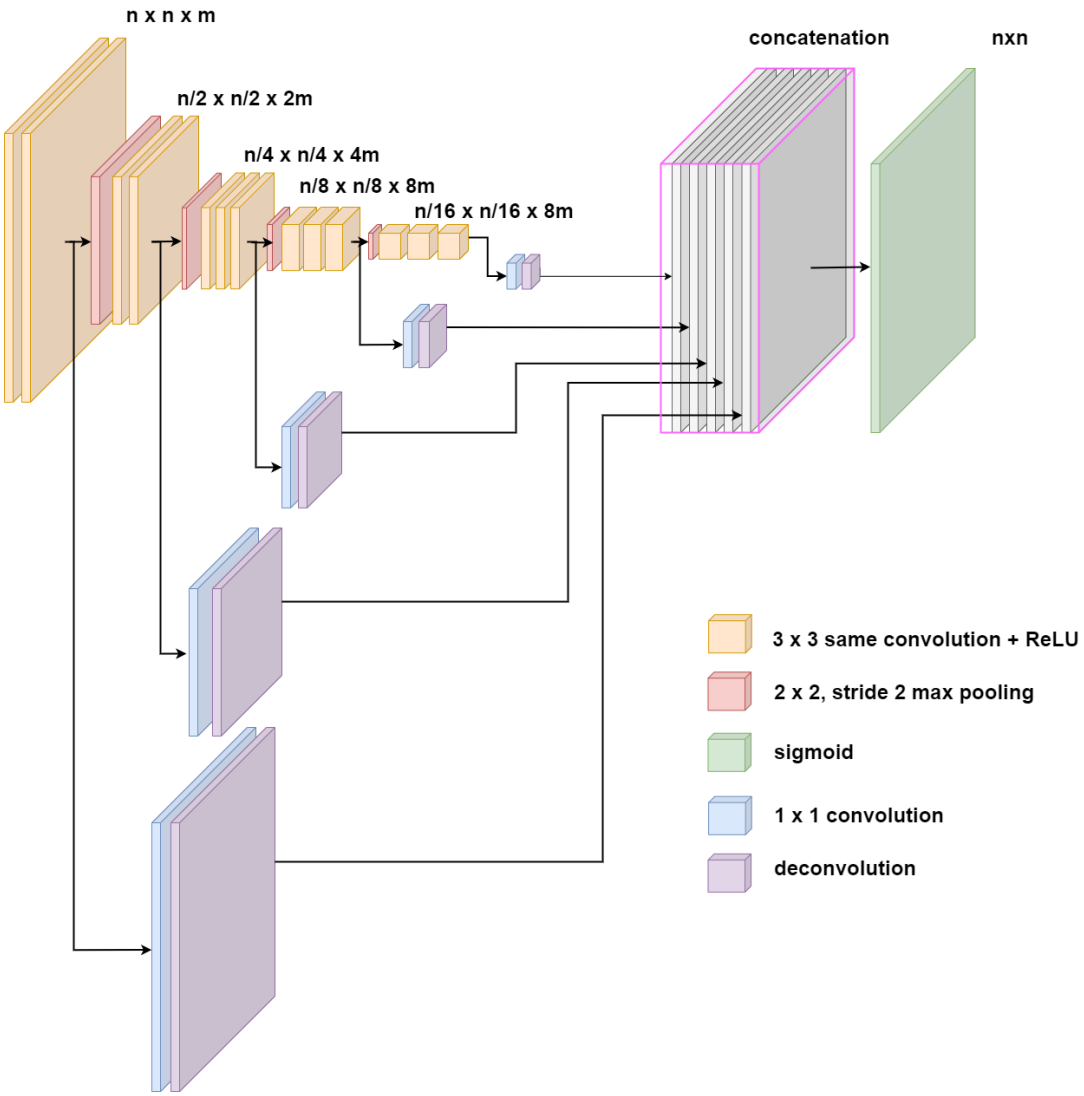


Figure 4.4: Holistically-Nested Edge Detection architecture (Xie and Tu 2015). The superscripts show the dimensions of the resulting tensor after convolution. n is the height/width of the input and m is the number of convolution filters.

grounding line pixels, y_{ngl} non grounding line pixels and operator indicating the number of pixels, the loss for prediction y is:

$$L(\hat{y}) = \frac{-|y_{ngl}| \sum_{j \in y_{gl}} \log \hat{y}_j}{|y|} - \frac{-|y_{gl}| \sum_{j \in y_{ngl}} \log 1 - \hat{y}_j}{|y|} \tag{4.1}$$

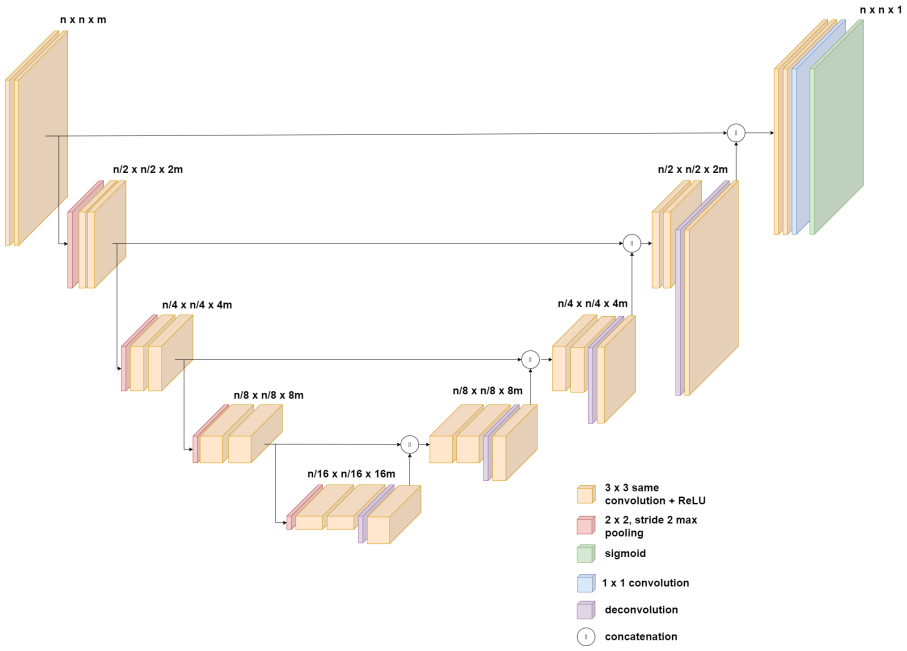


Figure 4.5: UNet architecture (Ronneberger et al. 2015). The superscripts show the dimensions of the resulting tensor after convolution, n being the dimensions of the input and m is the number of convolution filters

5. Experimental design

5.1. Grounding line delineation workflow

Fig. 5.1 shows the processing schematic for automatic grounding line delineation. While the whole chain has the capability of running seamlessly without any intervention, the main components (preprocessing, training and post-processing) were carried out separately due to constraints on memory and computing resources.

1. **Preprocessing**

   Double difference interferograms are resampled and reprojected to EPSG:3031. As mentioned in Section 4.1.1, this process generates the proxy coherence or noise feature. Along with the rasterized grounding lines (ground truth labels), they are divided into overlapping square tiles. A 20% overlap in four directions was used for all datasets in this study. This augments the dataset as well as facilitates merging of predictions later in the processing chain.

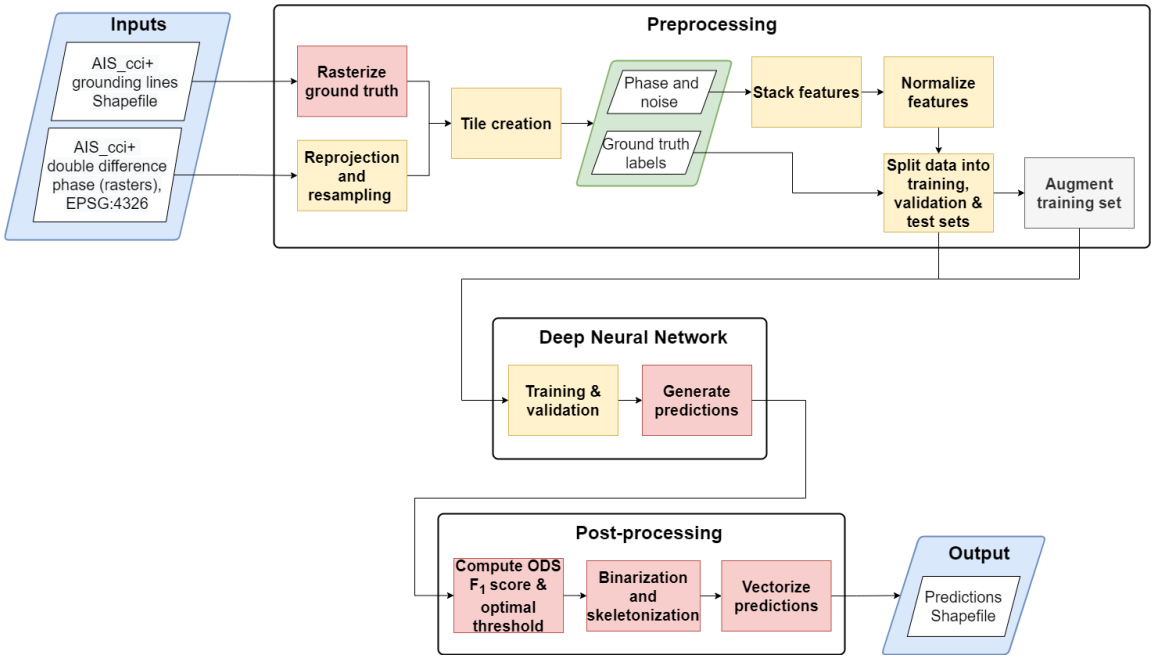


Figure 5.1: Automatic grounding line delineation flowchart. Data blocks are shown in blue, intermediate outputs in green, optional functions in grey, fixed operations in pink and flexible/configurable operations in yellow.

The rest of the features are stacked behind phase and pseudo coherence features and stored as tensors. Interpolation of missing values and removal of invalid tiles are handled simultaneously. This is followed by feature normalisation and division into training, validation and test sets (Section 5.2). An optional augmentation of the training set (random flipping along horizontal and vertical axes) is also possible.

2. **Training and validation**

   Model training and validation are handled by the neural network module. The component is configurable and provides generic functionalities applicable to both networks. The input tiles are run through the trained model to generate arrays of the same dimensions (height and width), each pixel value representing the probability of the pixel belonging to the grounding line class.

3. **Post processing**

   The probability maps are filtered to remove uncertain predictions. The threshold value is computed as the probability at which the maximum Optimal Dataset Scale (ODS) F1 score is obtained. The formulation of ODS F1 score is discussed in Section 5.4. The

predictions are binarized by converting all pixels with values lesser than the threshold to 0 and those equal to or greater than the threshold to 1. Spurious branches are removed by applying a median filter.

Thick and blurry lines are filtered out using a skeletonization algorithm (Zhang and Suen 1984), wherein pixels are iteratively removed until only connected pixels that represent the skeleton of the original pattern remain. Using the affine transform matrix of the ground truth rasters as reference, the skeletonized prediction rasters are converted to vectors and subsequently saved in the ESRI Shapefile format. Post-processing stages for a sample prediction are shown in Fig. 5.2.

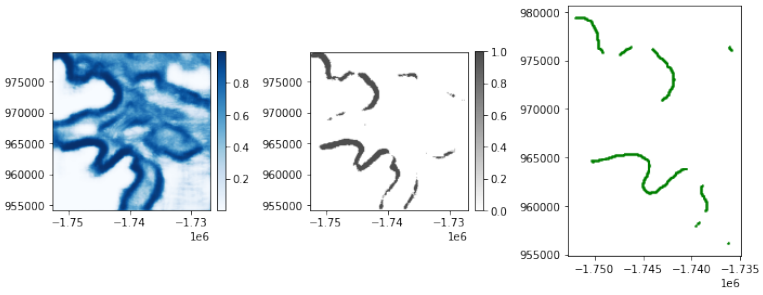


Figure 5.2: Post-processing steps a) probability map b) after thresholding + median filtering c) skeletonized + vectorized prediction. Plotted using Matplotlib version 3.5.0

5.2. Dataset variants

Several versions of the stacked features were created to explore the performance of the deep learning networks. Due to the large variation in the numerical ranges of the features, normalisation is necessary to prevent invalid or Not-a-Number (NaN) predictions. All features except phase (to preserve $-\pi$ to π variation), pseudo coherence (already in range 0 - 1), real and imaginary components of phase were normalised. The rest of the features were scaled to 0 - 1 range across the dataset. The goal of further experiments was to examine three different aspects:

1. **Feature importance**

   As explained in Section 1.2, one of the objectives of this study is to investigate the contributions of additional features to grounding line delineation. Two feature stacks with tiles dimension 256 x 256 pixels and pixel size of 100 m were used for training the HED network. The first dataset variant contained all the features described in Section 4.1.1. In the second variant, the phase and pseudo coherence were replaced by real and imaginary components of the phase obtained during the resampling process (Fig. 4.3). Importance of individual features was gauged by training several models for which one feature was excluded from the feature stack (Table 6.1). The contribution of each feature was determined indirectly by comparing metrics and visual inspections of predictions.

2. **Effect of tile size and pixel size**

   Due to the localised operating extent of convolutional layers, not all the pixels of the input have the same influence on each output pixel. The number of input pixels that contribute to one output pixel is called the receptive field (RF) of the CNN (Araujo et al. 2019), (Luo et al. 2016). The RF is a critical parameter for semantic segmentation, even more so in the case of extreme class imbalance because there is a limited number of informative pixels. Fig. 5.3 shows the receptive field for a four layer CNN given a one dimensional input.

   Pixels near the centre of the input tensor influence nearly all output pixels. This means that the DNN does not learn much from the pixels near the edges. An example of such a case is illustrated in Fig. 5.4 (a) which shows the RF (white rectangle) for HED. A 256 x 256 pixels tile of random numbers was used as the input. The gradients of the central pixel of the prediction (red cross) with respect to all input pixels was calculated. Pixels with non zero gradients were counted in the RF. The advantage of a large tile is apparent when considering the RF for a corner pixel of the 256 x 256 pixels tile (Fig. 5.4 (b)). Such a pixel would only affect a few of the output pixels. Therefore it was of interest to see if the performance of the networks improve when they are trained on large tiles. In this view, the networks were trained on two dataset variants that contain tiles with

extents 256 x 256 pixels and 1024 x 1024 pixels. The pixel spacing for both variants is 100 m.

A variant with tiles of 1024 x 1024 pixels and 50 m pixel size was created to check if the predictions improve when the networks are trained on higher resolution data. Additionally, HED and UNet were trained on a feature stack that contained only two features, the phase and the pseudo coherence.

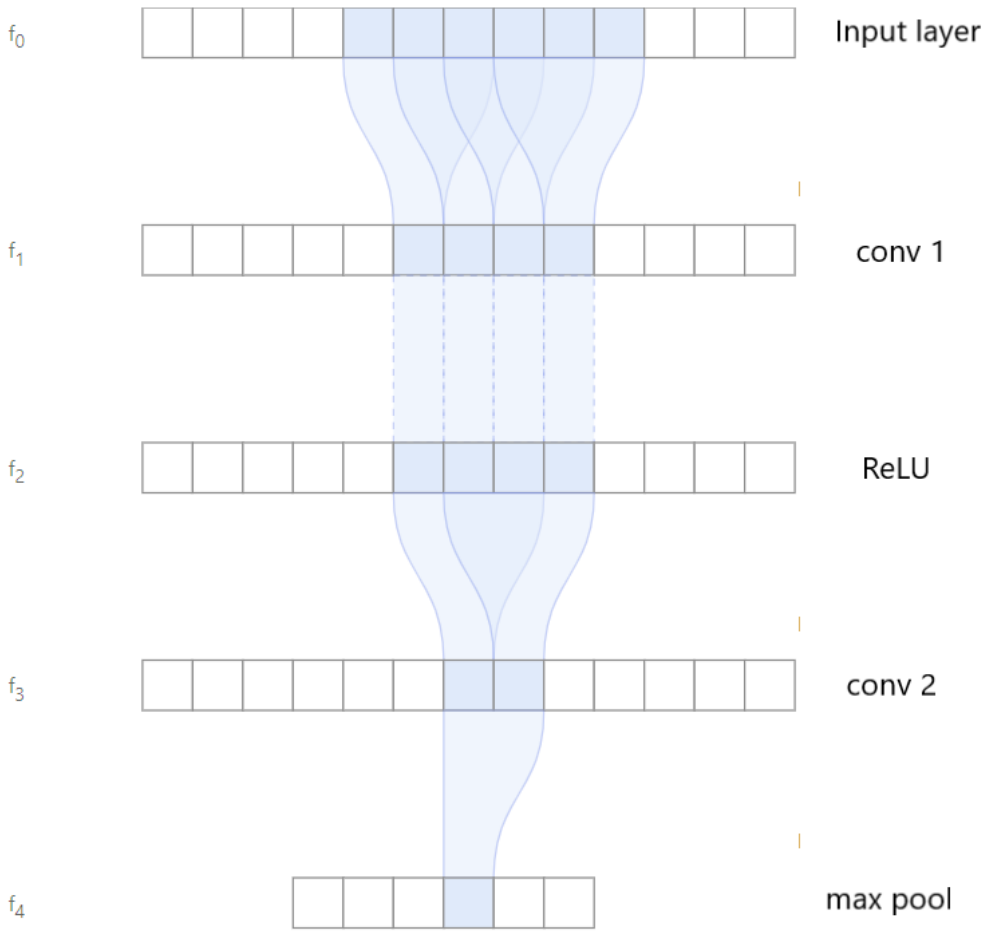


Figure 5.3: Receptive field (highlighted in blue) for a four layer CNN. f_0 is the input layer followed by a convolutional layer (f_1), ReLU activation function (f_2), a second convolutional layer (f_3) and a max pooling layer (f_4). The convolutional layers have a kernel size = 3, padding = 1, stride = 1 and the max pooling layer has a kernel size = 2, stride = 2. Figure adapted from (Araujo et al. 2019).

3. **Augmentation schemes**

The larger the number of training samples, the greater is the generalising capability

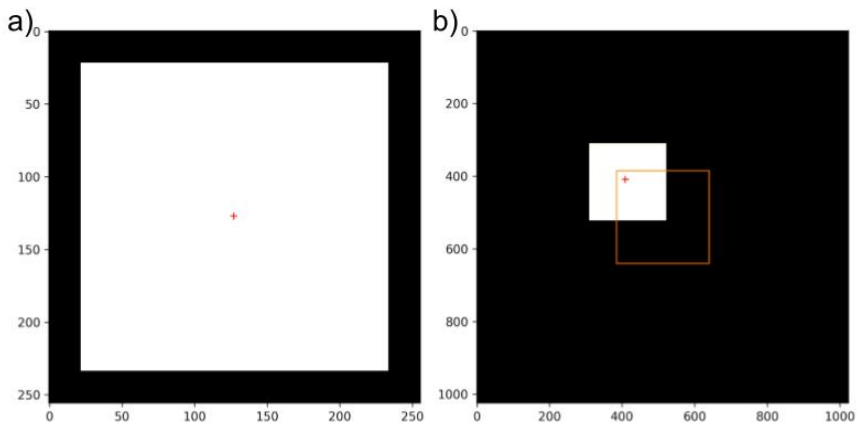


Figure 5.4: Receptive field (white rectangle, 212 x 212 pixels) for HED a) computed for the centre pixel (red cross) for an input tile of size 256 x 256 pixels b) computed for a corner pixel of the 256 x 256 tile (represented as an orange outline). The same pixel in a 1024 x 1024 tile is much closer to the centre of the tile and therefore exerts more influence on the output pixel.

of a DNN, with model performance increasing logarithmically with the volume of training data (Sun et al. 2017). Data augmentation refers to introducing additional training samples by applying label preserving transformations. Techniques range from applying simple geometric or affine transformations (random rotations, flips, cropping and scaling), oversampling minority class labels (Synthetic Minority Oversampling Technique (SMOTE)), to using neural networks that can generate 'believable' samples by learning the distribution of the original dataset (Generative Adversarial Networks (GANs))(Shorten and Khoshgoftaar 2019).

Two types of augmentation were considered: random flipping of 'valid' tiles and inclusion of 'non-valid' tiles, with tiles being referred to as non-valid if they do not contain any grounding line pixels. Both augmentation schemes were applied to UNet and HED networks trained on the full feature stack with tiles of 256 x 256 pixels, 100 m pixel size. The augmentation doubled the number of training samples for each variant.

5.3. Training details

All dataset versions were split into training and test sets in the ratio 90:10 (Table 5.1). All models were trained on a GeForce RTX 3090 GPU with 24 GB VRAM for a maximum of 30 epochs. Early stopping was implemented to prevent overfitting (Chollet 2021 5.4.3).

Adam optimizer was used for both models, with learning rates mentioned in Table 5.1 and default values for the parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^-8$ as specified in (Kingma and Ba 2014). Convolution kernel weights were initialised with the uniform Xavier initializer (Glorot and Bengio 2010). Fig. 5.5 shows the spatial distribution of training (lines from 296 double difference interferograms) and test set (lines from 35 double difference interferograms) used for all dataset variants. Due to unavailability of a few interferograms some GLLs (black) were not used.

Table 5.1: Training details. The split shows numbers before augmentation

Dataset (n x n), pixel size	**# Train/test tiles**	Neural network hyperparameters					
		HED			**UNet**		
		# Filters	Batch	Learning	# Filters	Batch	Learning
256 x 256, 100 m	4142/558	64	128	10^{-03}	8	16	10^{-05}
1024 x 1024, 100 m	788/110	32	16	10^{-03}	8	4	10^{-05}
1024 x 1024, 50 m	1762/246	32	16	10^{-03}	8	4	10^{-05}

5.4. Metrics

The predictions were quantified using two metrics. F_1 score at the optimal dataset scale (ODS) was used to measure the pixel wise accuracy of the models. The F_1 score is defined as the harmonic mean of precision and recall:

$$F_1 = 2\frac{precision * recall}{precision + recall} \tag{5.1}$$

with $precision = \frac{true\ positives}{true\ positives + false\ positives}$ and $recall = \frac{true\ positives}{true\ positives + false\ negatives}$

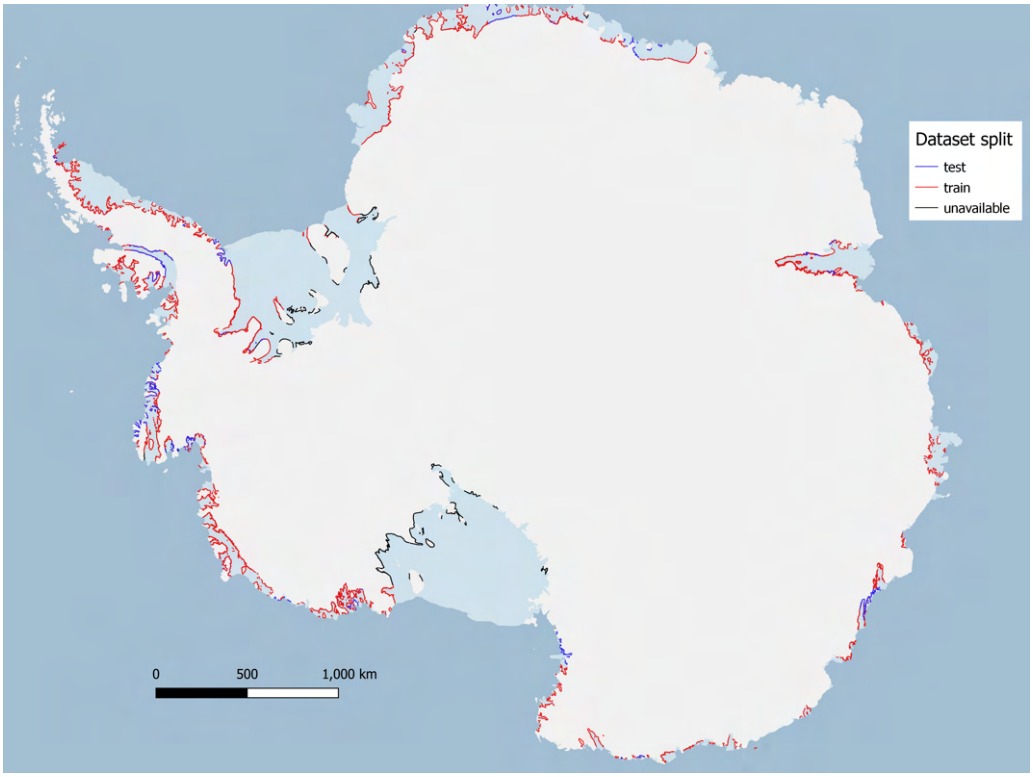


Figure 5.5: Spatial distributions of train (red) and test (blue) split for all dataset variants. Black lines are present in the AIS_cci dataset but are not used in the experiments. Plot created using Quantarctica 3 (Matsuoka et al. 2021).

Precision quantifies the proportion of pixels identified as grounding line to actually be grounding line pixels. Recall shows the proportion of correctly identified grounding line pixels. The ODS F_1 score was calculated for a range of thresholds to find the optimum threshold (at maximum ODS F_1 score) for binarization of the probability maps.

Additionally, the deviation between predictions and ground truth was computed using the metric for polygons and line segments (PoLiS) (Avbelj et al. 2014). For two line segments A and B with point sets $a_j \in A, j = 1, 2, \ldots j$ and $b_k \in B, k = 1, 2 \ldots . r$ respectively (Fig. 5.6), the PoLiS distance is calculated as:

$$p(A, B) = \frac{\vec{p}(A, B)}{2q} + \frac{\vec{p}(B, A)}{2r} \tag{5.2}$$

with $\vec{p}(A, B) = \sum_{a_j \in A}^{q} \min_{b_k \in \partial B} ||a - b||$ being the PoLiS distance from line A to B. ∂B are points on B closest to a_j.

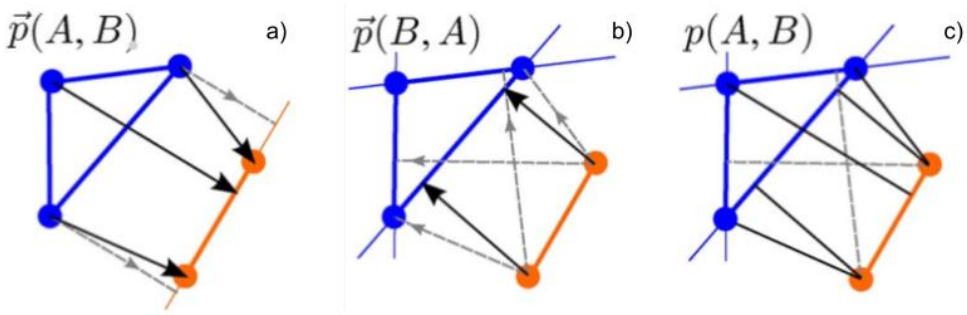
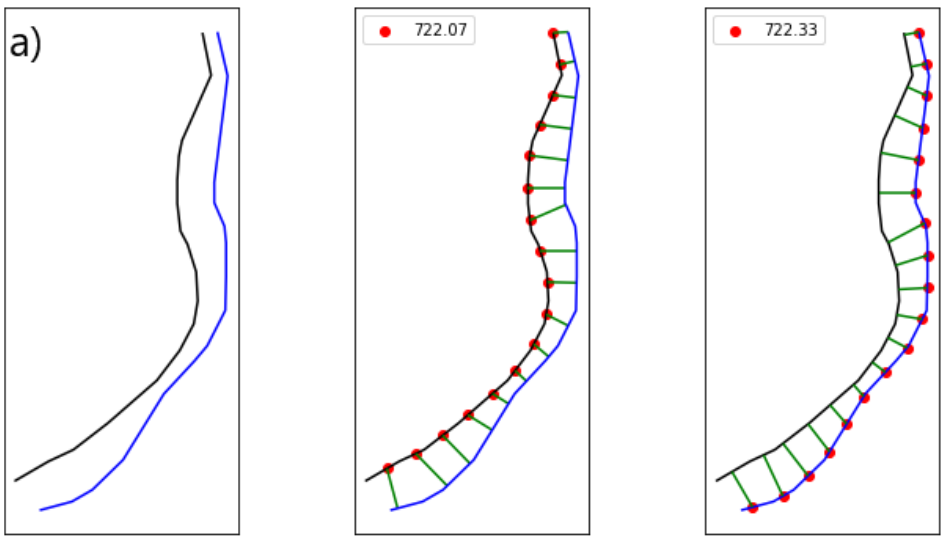


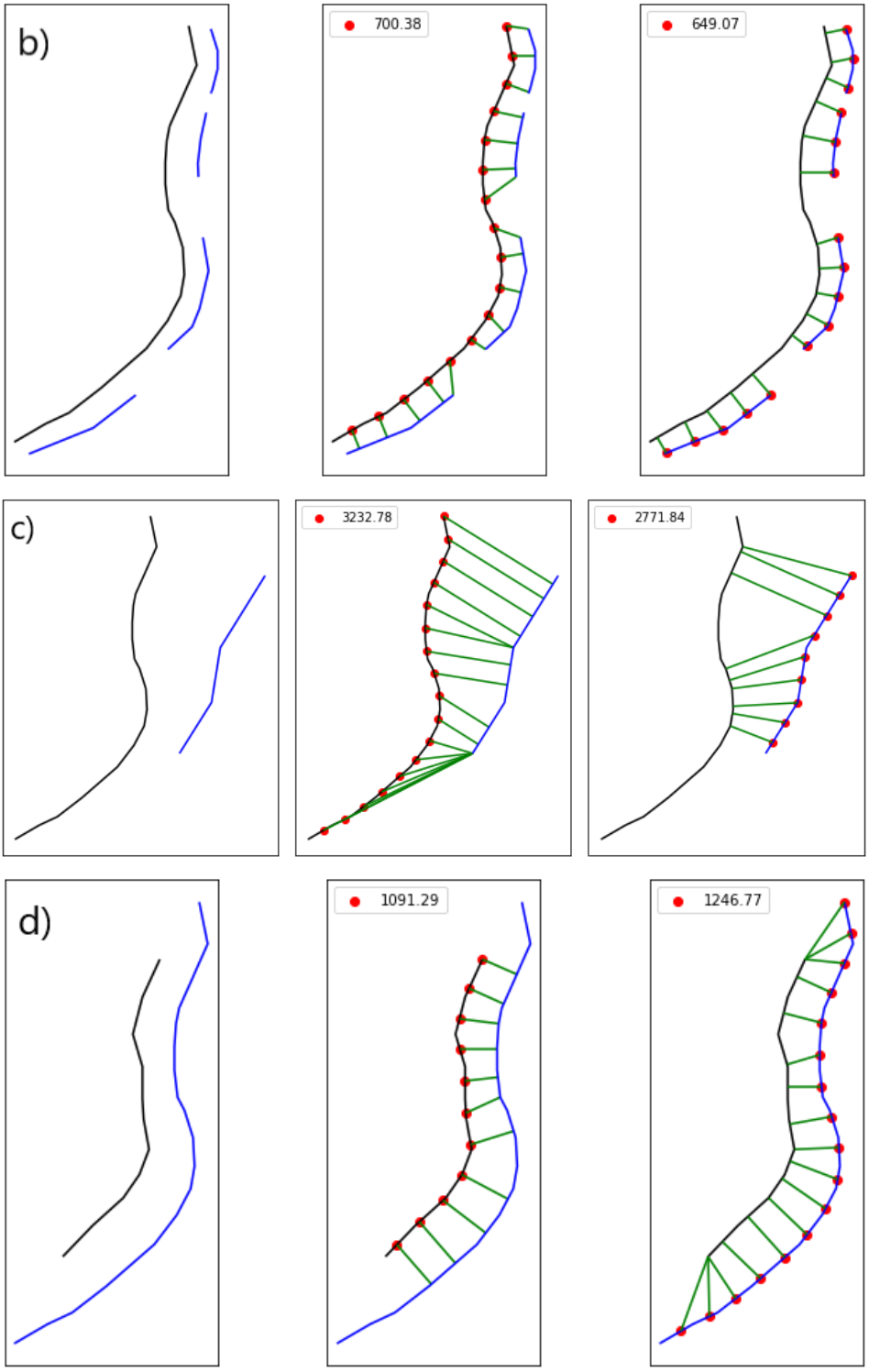
Figure 5.6: PoLiS metric calculation (adapted from (Avbelj et al. 2014)) a) from polygon A (blue) to line segment B (orange) b) from line segment B to polygon A c) symmetric distance. Black arrows show direction in which the distance is calculated, black lines show point to line measurements.

6. Results and discussion

The outcomes of experiments described in Section 5.2 are enumerated in this chapter in terms of numerical results (with the metrics described in Section 5.4) and visualisations of the predicted GLLs. The median values of prediction to ground truth, ground truth to prediction and overall PoLiS distances are reported. While the overall deviation provides an overview of the prediction quality, the one way distances provide an insight into the geometry of the predictions. Fig 6.2 shows examples of ground truth and predicted GLL position configurations. When the prediction follows the GLL in length and shape, the one-way distances are similar (Fig. 6.2 a). Prediction to ground truth deviation is smaller when the predicted line is either fragmented (Fig. 6.2 b) or shorter (Fig. 6.2 c) than the ground truth. Ground truth to prediction distance is smaller when the predicted line is longer (Fig. 6.2 d) or contains branches that are close or cross the ground truth (Fig. 6.2 e).

Median Absolute Deviation (MAD) is used as a measure of dispersion due to the right skew in the distributions of the deviations (A.1). The ODS F1 scores (A.3) for all experiments are fairly low due to the large class imbalance and therefore are not used to evaluate the quality of the predictions.




b)
700.38
649.07
c)
3232.78
2771.84
d)
1091.29
1246.77

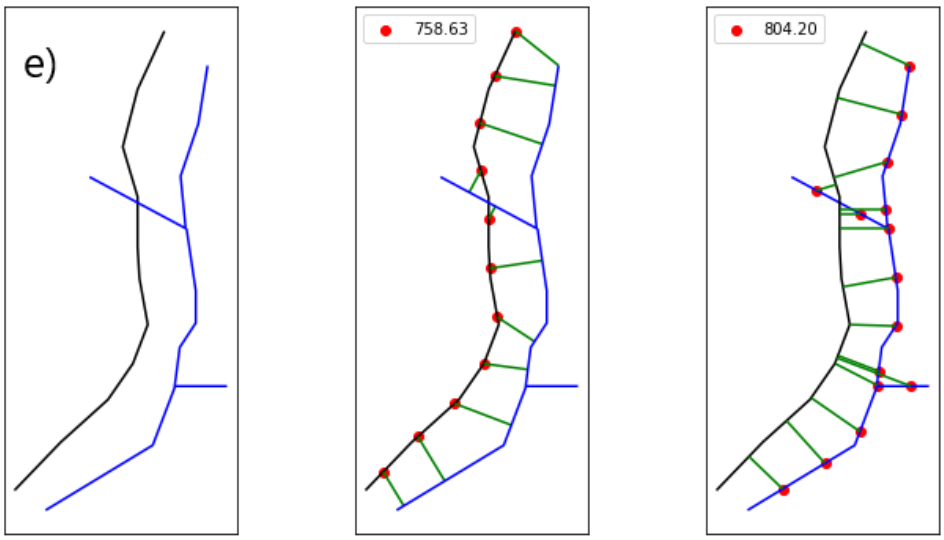


Figure 6.2: Computation of PoLiS for several ground truth (black) and prediction (blue) configurations. Column 2 visualises the ground truth to prediction distance and column 3 shows the prediction to ground truth distance. The distances are measured from red points on one line to its closest point on the other line (Section 5.4). Respective distances (metres) are shown in the legend. Configuration a) ideal case where prediction follows the ground truth closely b) prediction is fragmented but still within the length of ground truth c) prediction is far away and shorter than the ground truth, d) prediction length exceeds ground truth distance and e) spurious branches on the prediction. Plotted using Matplotlib version 3.5.0.

6.1. Feature exploration

This section presents the results from training several HED networks on the 256 x 256 pixels, 100 m pixel size dataset variant, following the experimental design described in Section 5.2. The dataset was augmented by training samples that were randomly flipped along either their horizontal or vertical axis.

6.1.1 Deviation from AIS_cci

Table 6.1 summarises the metrics computed on the test set (Fig. 5.5) (Table 5.1, row 1) for the feature exploration experiments. The experiments revealed that the model trained on the feature stack with real and imaginary components (Section 5.2) results in the least overall deviation. HED has the worst performance when either DEM features, phase or pseudo coherence are absent.

Table 6.1: Metrics for the test set (Fig. 5.5) of feature exclusion experiments. Model IDs are mentioned in brackets in the first column.

Experiment HED (256 x 256), 100 m	**Prediction to ground truth** [m]	**Ground truth to prediction** [m]	**Overall deviation** [m]	**False negatives** (lines/1232)
All features **(hed_256_100_flipped)**	202.83 MAD = 147.19	541.77 MAD = 477.72	328.06 MAD = 270.63	0
Phase and pseudo coherence only **(hed_256_100_phase_noise)**	219.84 MAD = 165.8	554.28 MAD = 483.77	337.92 MAD = 280.44	0
Pseudo coherence absent **(hed_256_100_no_noise)**	201.16 MAD = 143.96	735.98 MAD = 666.42	437.04 MAD = 376.58	0
Phase absent **(hed_256_100_no_phase)**	248.68 MAD = 183.35	908.49 MAD = 810.22	539.39 MAD = 462.45	0
DEM absent **(hed_256_100_no_dem)**	205.77 MAD = 150.89	862.65 MAD = 779.82	519.33 MAD = 452.79	0
Vx absent **(hed_256_100_no_vx)**	198.7 MAD = 143.99	464.89 MAD = 404.95	302.99 MAD = 247.45	0
Vy absent **(hed_256_100_no_vy)**	201.10 MAD = 145.95	422.69 MAD = 360.82	276.59 MAD = 221.12	0
Tidal displacement absent **(hed_256_100_no_td)**	202.66 MAD = 145.50	413.89 MAD = 347.5	273.5 MAD = 214.69	1
Air pressure absent **(hed_256_100_no_ap)**	190.47 MAD = 135.03	560.91 MAD = 498.83	356.23 MAD = 299.67	1
Real and imaginary components **(hed_256_100_real_imaginary_all)**	194.09 MAD = 136.49	323.03 MAD = 258.01	**209.23** MAD = 152.91	0
Real and imaginary components only **(hed_256_100_ real_imaginary)**	276.69 MAD = 223.52	371.59 MAD = 310.35	241.50 MAD = 186.14	0

Generally, the GLL appears as an edge in the DEM and pseudo coherence features. An example of this can be seen in the feature stack for a test sample from the Abbot Ice Shelf (Fig. 6.3). Consequently, these two features might help in fine tuning the position of overall predicted GLL. The exclusion of DEM and pseudo coherence from the experiments results in deviations >1 km (Fig. 6.4). Even the prediction from the best model (hed_256_100_-real_imaginary_all) has a deviation of 1.8 km due to a labelling error for the part of the GLL enclosed in the red rectangle. Here the AIS_cci GLL was incorrectly mapped on the seaward

extent of the grounding zone fringe belt. This example shows the robustness of HED against labelling errors.

In contrast, the GLL is almost never visible in the ice velocity, tidal displacement and air pressure features. Although it is unclear the degree to which they influence the predictions, the lack of strong gradients in these features (Fig. 6.3) and the deviations from hed_256_-100_no_vx, hed_256_100_vy, hed_256_100_no_ap and hed_256_no_td models for the test sample from Abbot Ice Shelf (Fig. 6.5) suggest that these features seem to have an overall negative impact on the predictions.

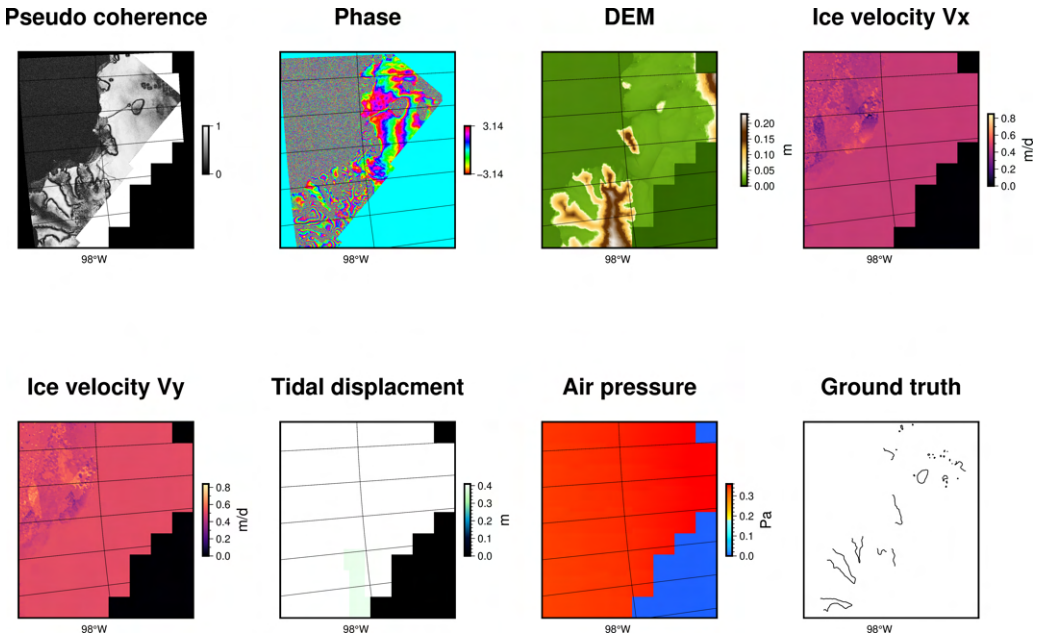


Figure 6.3: Feature stack of the test sample shown in Fig. 6.5. The GLL is not expressed as an edge in the ice velocity, tidal displacement and air pressure features.

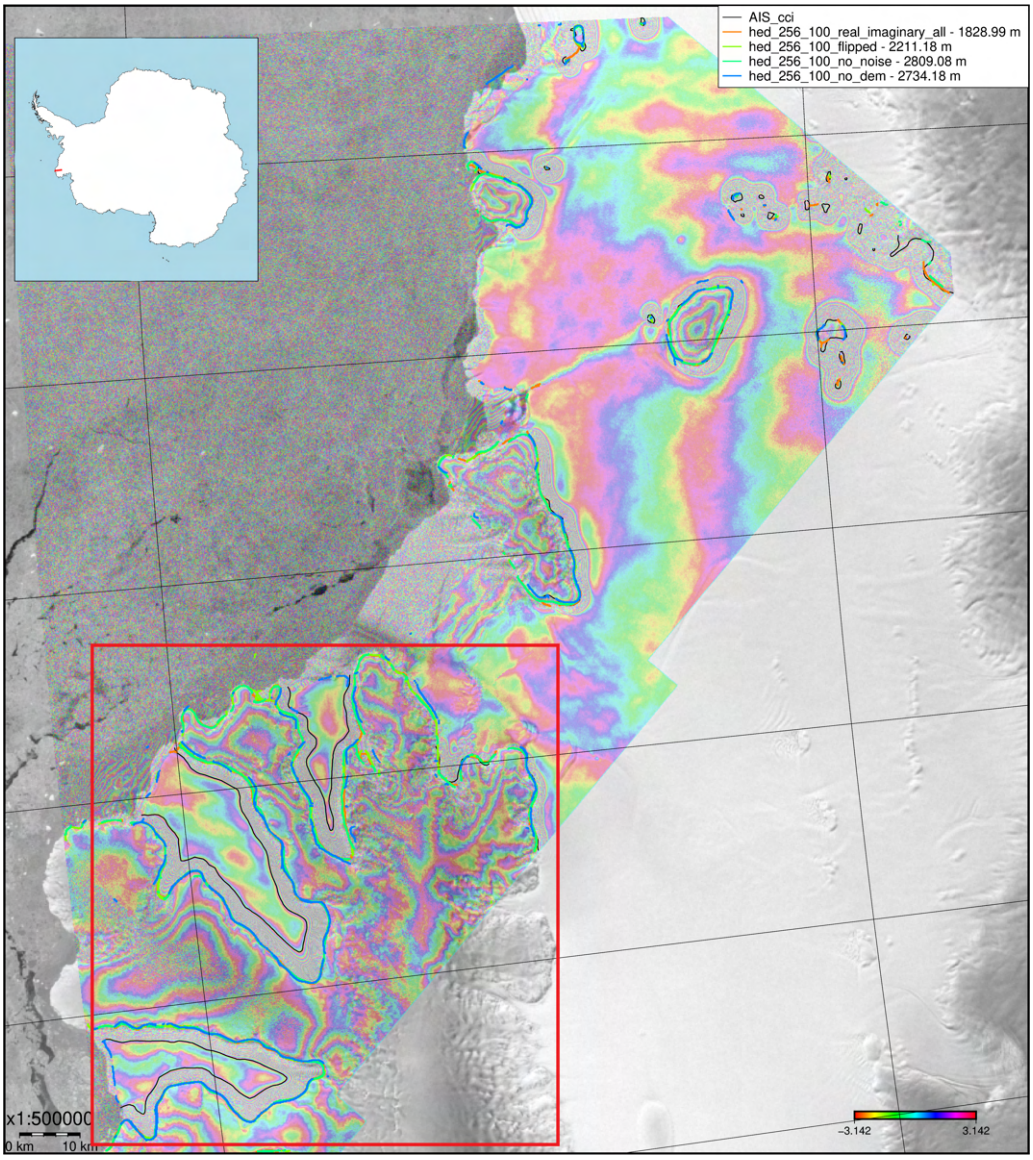


Figure 6.4: Predictions for test sample taken from Abbot Ice Shelf. The red rectangle shows an example of incorrect manual delineation. The predictions are plotted on the RAMP backscatter scatter mosaic (in greyscale) (Jezek 2002) and the corresponding DInSAR interferogram (formed from Sentinel-1 scenes). Numbers in the legend show the average overall deviations. The plot was created using A Python interface for the Generic Mapping Tools (PyGMT) (Uieda et al. 2021) version 0.5.0.

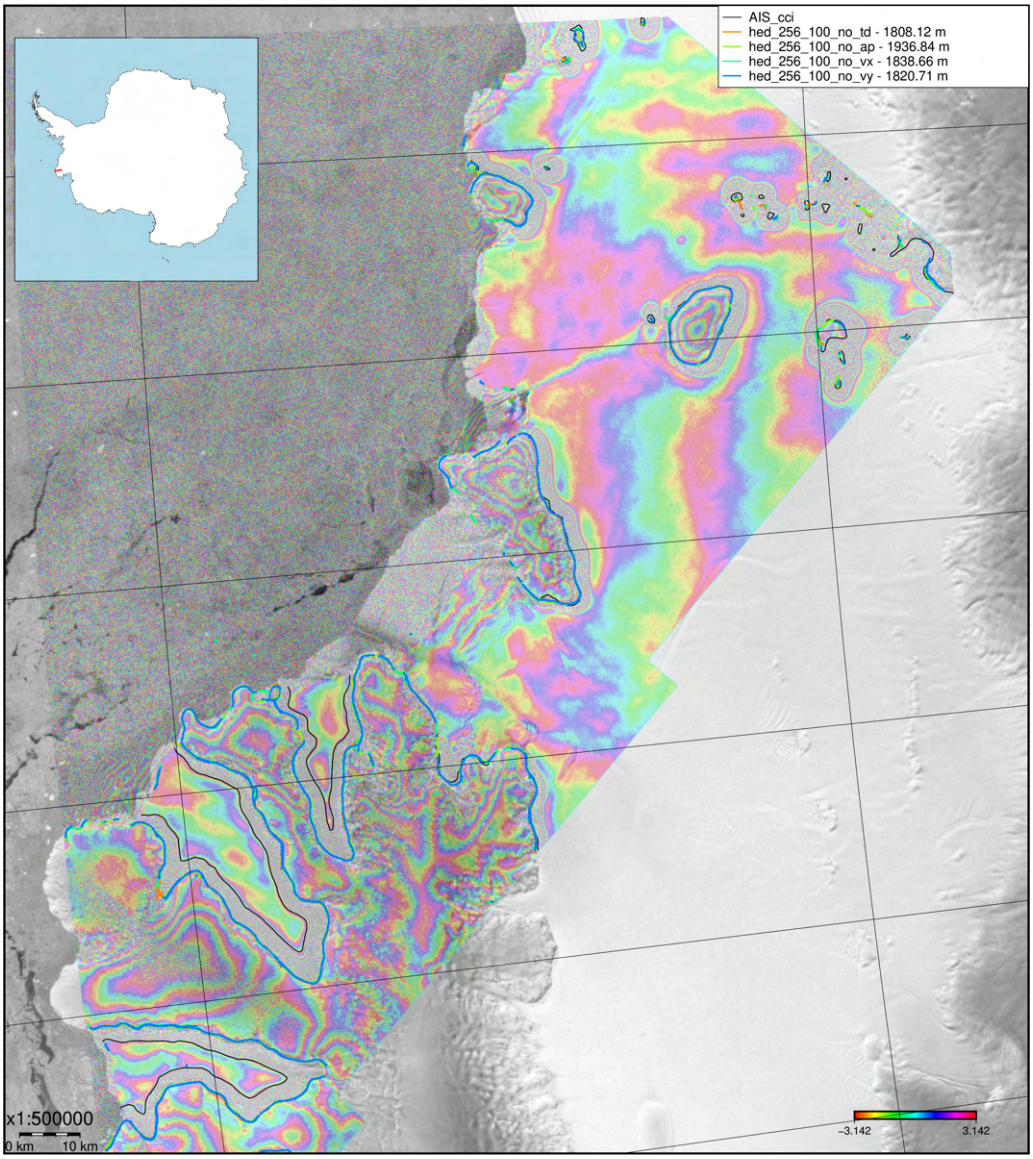


Figure 6.5: Predictions for test sample taken from Abbot Ice Shelf, from experiments in which velocity components, tidal displacement and air pressure were withheld from the training stack, implying negligible contribution from the above mentioned features.

Deviations computed for predictions from hed_256_100_flipped do not differ greatly from hed_256_100_phase_noise, suggesting that the contributions from the rest of the features cancel each other out. A similar behaviour is also exhibited by hed_256_100_real_imaginary_-all and hed_256_100_real_imaginary experiments. Visual inspection of predicted GLLs for test samples from Getz Ice Shelf (Fig. 6.6), Totten Ice Shelf (Fig. 6.7), echo the metrics specified in Table 6.1. The above mentioned models seem to fail in areas where the GLL curves around loose fringes (enclosed by the red rectangles in Fig. 6.6 and Fig. 6.7). Additionally, false positives and spurious branches (enclosed by black rectangles in Fig. 6.6, and Fig. 6.7) bias the overall deviation by hundreds of metres. hed_256_100_real_imaginary and hed_-256_100_real_imaginary_all do not seem to generate spurious branches. The models seem to perform well on Getz Ice Shelf, which could be attributed to the relatively high coherence of the training samples in this region.

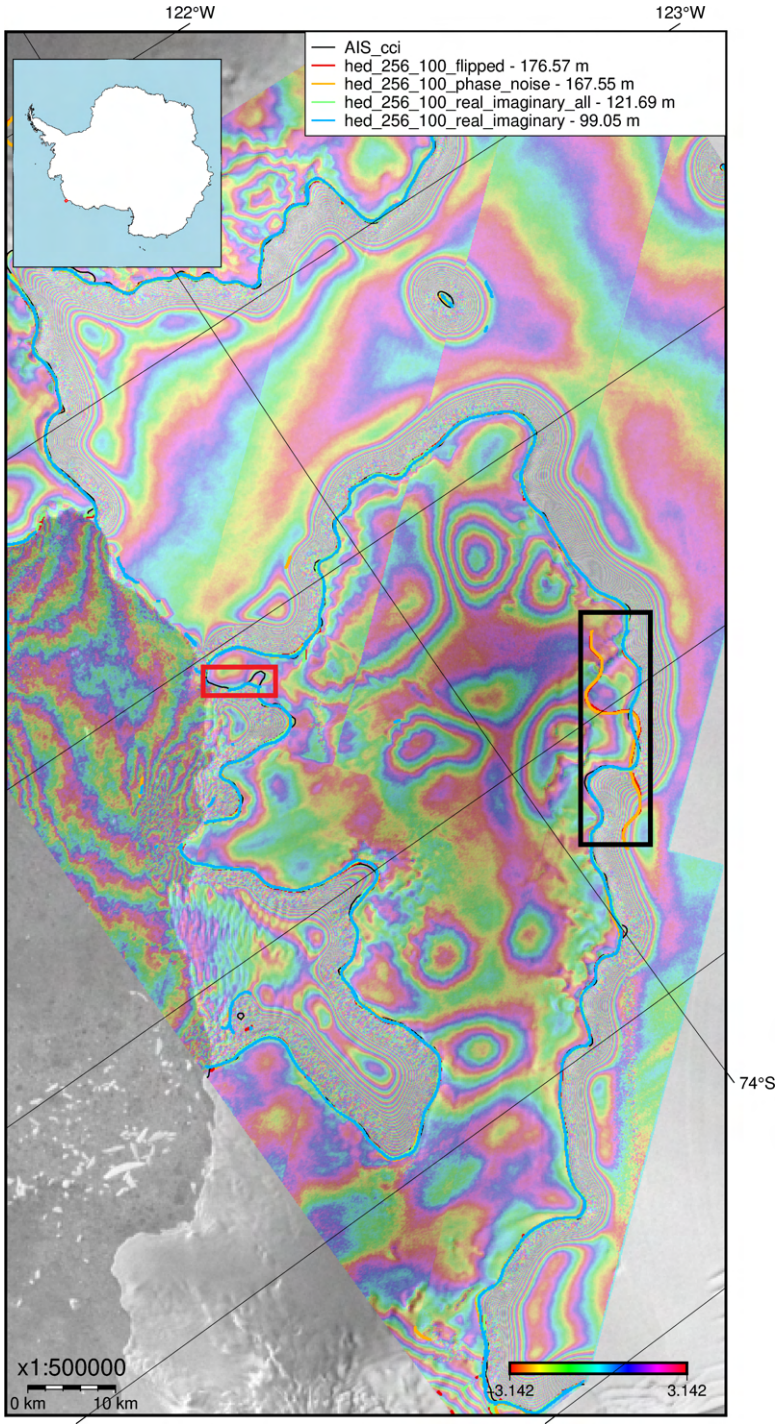


Figure 6.6: Predictions for a test sample from Getz Ice Shelf.Area enclosed in the black rectangle shows spurious predictions, area enclosed by the red rectangle show parts of GLL where the models fail to make a prediction.

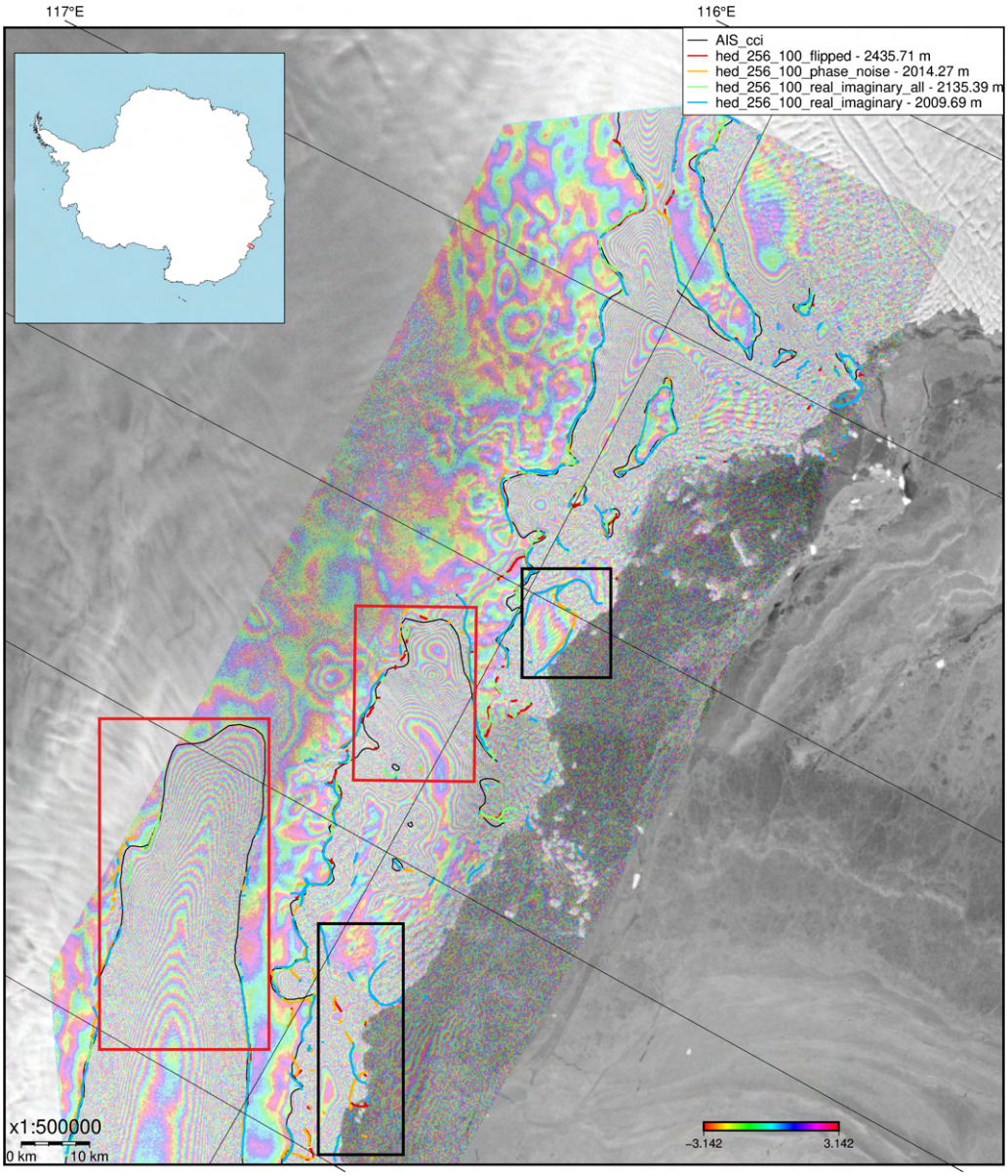


Figure 6.7: Predictions for a test sample from Totten Ice Shelf. Large portions of the GLL were missed by the models (red rectangles). There are also more false positives (black rectangles) for this sample, compared to Fig 6.6

Nearly all GLLs were predicted except for one that was missed by the models that excluded tidal displacement and air pressure. A buffer was created around all the predictions to calculate the percentage of ground truth area covered by them on the whole. About 75% of the GLLs are covered by most models within 2 km (Fig. 6.8), 80% is covered by predictions from hed_256_100_real_imaginary_all. The predictions from hed_256_100_no_dem only cover 65% of the ground truth pixels, implying that most of the predictions are fragmented. This is also reflected by the large difference in the one way distances (Table 6.1) for this model.

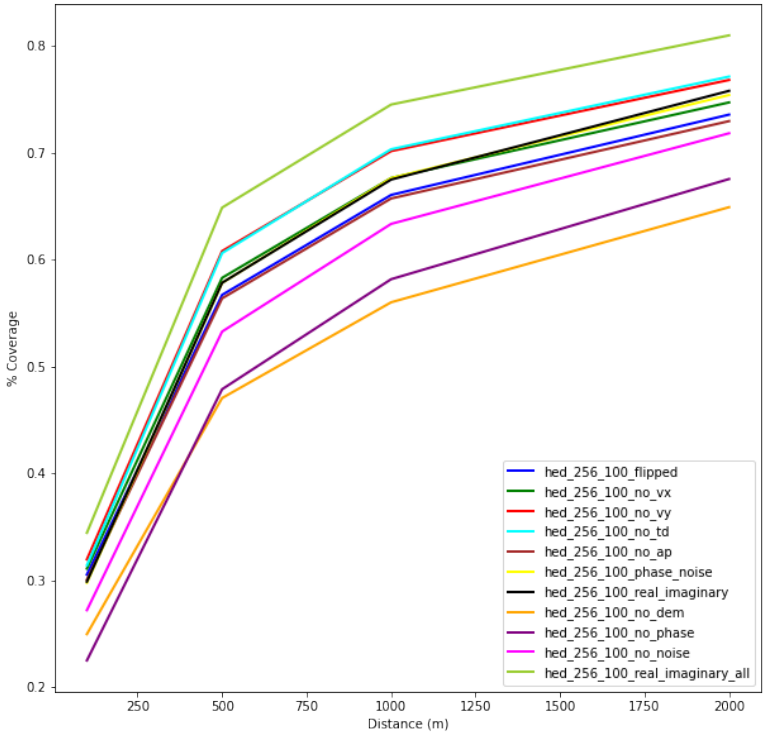


Figure 6.8: Cumulative ratio showing the ground truth coverage by model predictions

6.1.2 Time series analysis

The hed_256_100_real_imaginary model was used to generate predictions for a set of unlabelled DInSAR interferograms over the Totten Ice Shelf. Sentinel-1 triplets with a 6 day temporal baseline were used to generate 57 DInSAR interferograms spanning from January 2019 to December 2019. The processing chain specified in Section 4.1.1 was used to generate the double differences.

Real and imaginary components from the DInSAR phases were generated according to the procedure shown in Fig. 4.3. The interferograms were tiled and resampled into 8200 overlapping tiles following the preprocessing procedure detailed in Section 5.1.

Although there is a significant number of false detections landwards of the GLL (Fig. 6.9), the DNN is able to identify large portions of the grounding line for the parts of the ice shelf without or slow moving tributaries (enclosed by the blue rectangles in Fig. 6.10). However, the model still fails to predict the GLL in some regions where dense fringes due to tidal displacement are clearly visible (marked by red rectangles in Fig. 6.10). Fig 6.11 shows an example of a strongly decorrelated interferogram for which the model generated a large number of false positives. A similar pattern of false and fragmented predictions resulted for most of the highly decorrelated interferograms acquired between January and mid October (Fig. 6.9).

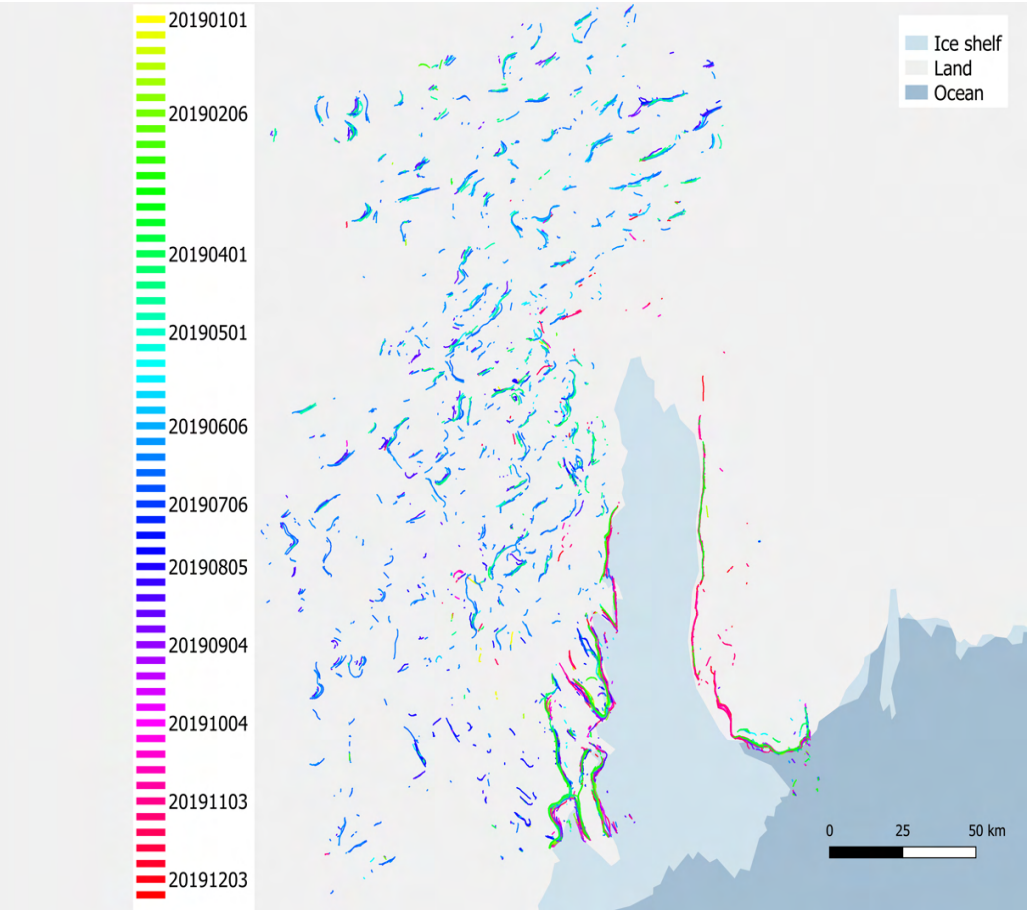


Figure 6.9: Predictions generated from the hed_256_100_real_imaginary model spanning January 2019 - December 2019

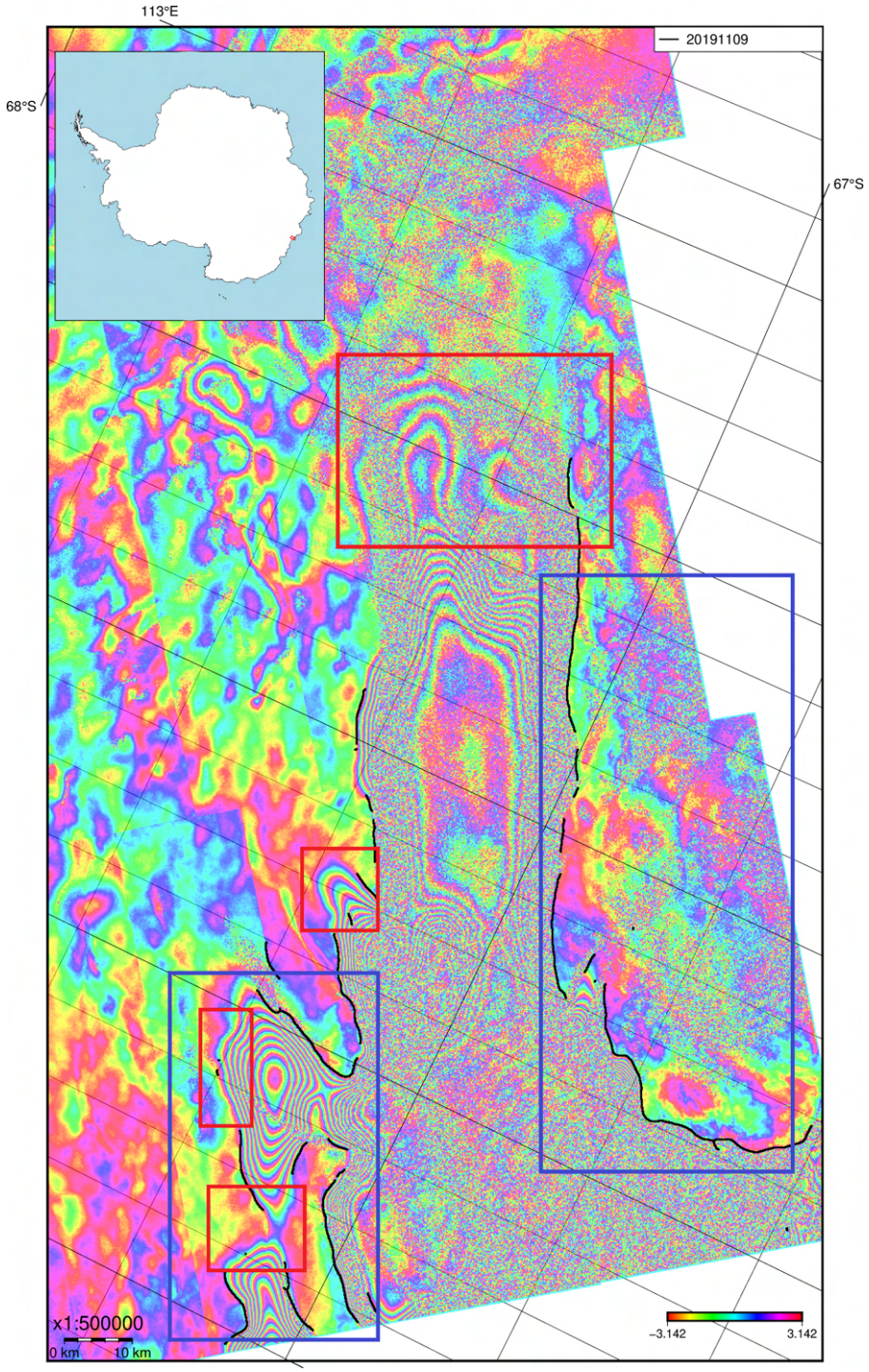


Figure 6.10: Example of a good prediction (primary acquisition at the beginning of November 2019) from the time series generated for the Totten Ice Shelf. Blue rectangles show regions where the GLL is predicted consistently for all DInSAR interferograms. Red rectangles show areas where the model fails, despite the presence of coherent fringes.

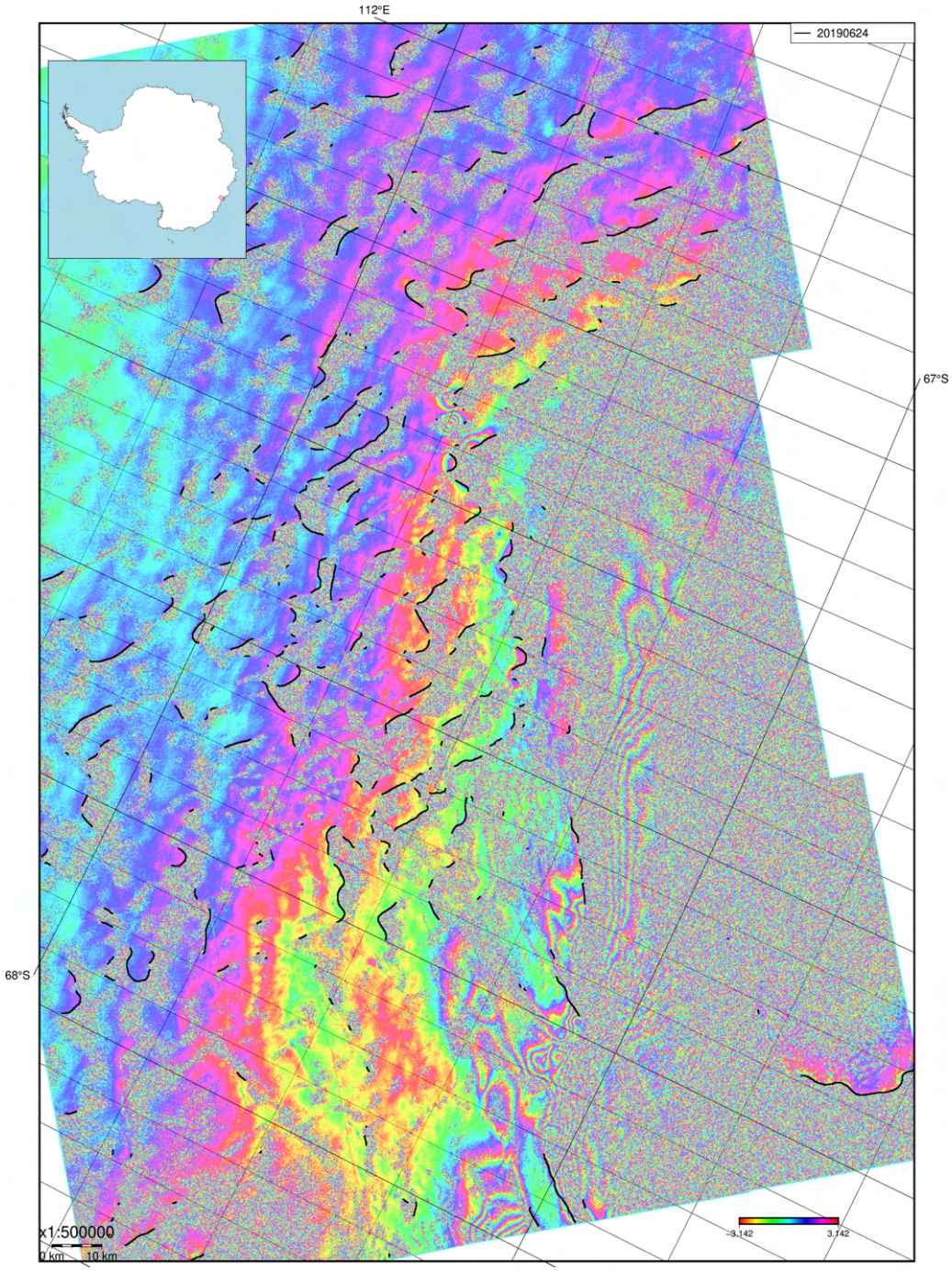


Figure 6.11: Example of a poor prediction (primary acquisition at the end June 2019) from the time series generated for the Totten Ice Shelf. The interferogram is highly decorrelated, even in the stable regions indicated in Fig. 6.10

6.2. Comparing networks

This section presents the results from the experiments which explored different augmentation techniques (Section 5.2, Augmentation Schemes) and the impact of tile and pixel size (Section 5.2, Effect of tile size and pixel size). The test sites for the metrics were computed are visualised in Fig. 5.5, with the dataset split shown in Table 5.4.

6.2.1 Effect of augmentation

Augmentation of the training set with flipped training samples results in a greater overall deviation for HED, while adding tiles that do not contain grounding line pixels yields better results for both networks (Table 6.2). For HED, augmentation with empty tiles yields long spurious branches that cross the ground truth (Fig. 6.12 a), which could bias the deviation as demonstrated in Fig 6.2 e. For UNet however, the effect of augmentation is not clear due to the highly fragmented predictions (Fig. 6.12 b) generated by all the UNet models.

Table 6.2: Metrics (computed for test samples) for experiments that used different augmentation schemes.

Experiment	**HED**			**UNet**		
(256 x 256), 100m,	**Prediction to ground truth deviation** [m]	**Ground truth to prediction deviation** [m]	**Overall** [m]	**Prediction to ground truth truth** [m]	**Ground truth to prediction deviation** [m]	**Overall** [m]
All features **(256_100)**	213.32 MAD = 152.53	455.04 MAD = 382.12	288.99 MAD = 225.56	1955.91 MAD = 1474.94	2186.67 MAD = 1609.00	1400 MAD = 1040.14
All features, random flips **(256_100_flipped)**	202.83 MAD = 147.19	541.77 MAD = 477.72	328.06 MAD = 270.63	1814.27 MAD = 1539.49	1930.77 MAD = 1585	1217.25 MAD = 1005.39
All features, empty tiles **(256_100_empties)**	327.24 MAD = 265.51	336.91 MAD = 260.08	225.91 MAD = 161.22	2098.96 MAD = 1447.27	1358.65 MAD = 744.65	973.69 MAD = 563.66

The predictions shown in Fig 6.12 demonstrate a potential weakness of the PoLiS metric. Although the deviations for both the networks are considerably large (5 km-6 km), the quality

of the predicted GLLs in terms of completion and continuity of the lines is not reflected in the numbers. Overall, the predictions from the HED models are visibly closer to the ground truth (AIS_cci GLL) and less fragmented when compared to those from the UNet models. Spurious branches are not completely avoidable for HED either.

6.2.2 Effect of tile size and pixel size

Training on larger tiles produces results in smaller overall deviations (Table 6.3) for both networks. The large difference between the one direction distances, however, suggests that the corresponding predictions would be fragmented (refer Fig. 6.2 b). An example of this can be seen in the predictions for a test sample from George IV Ice Shelf (Fig 6.13 and Fig. 6.14). Predictions from the models trained on the larger tiles tend to follow the seaward extent of the fringe belt (black rectangles in Fig. 6.13 and Fig 6.14) for both networks.

Table 6.3: Metrics (computed for the test set) for the experiments exploring the effects of tile size and pixel size.

Dataset		**HED**			**UNet**	
Pseudo, Coherence and Phase	**Prediction to ground truth deviation** [m]	**Ground truth to prediction deviation** [m]	**Overall deviation** [m]	**Prediction to ground truth truth** [m]	**Ground truth to prediction deviation** [m]	**Overall deviation** [m]
256 x 256, 100 m **(256_100_phase_noise)**	202.83 MAD = 147.19	541.77 MAD = 477.72	328.06 MAD = 270.63	1845.99 MAD = 1567.41	2100.53 MAD = 1764.21	1386.21 MAD = 1186.87
1024 x 1024, 100 m **(1024_100_phase_noise)**	655.59 MAD = 597.54	267.33 MAD = 161.02	193.65 MAD = 110.08	8289.4 MAD = 6200.91	1529.07 MAD = 749.28	1342.95 MAD = 753.52
1024 x 1024, 50 m **(1024_50_phase_noise)**	538.63 MAD = 500.25	291.82 MAD = 203.38	190.52 MAD = 130.60	3359.68 MAD = 2786.83	1094.76 MAD = 744.95	718.95 MAD = 484.15

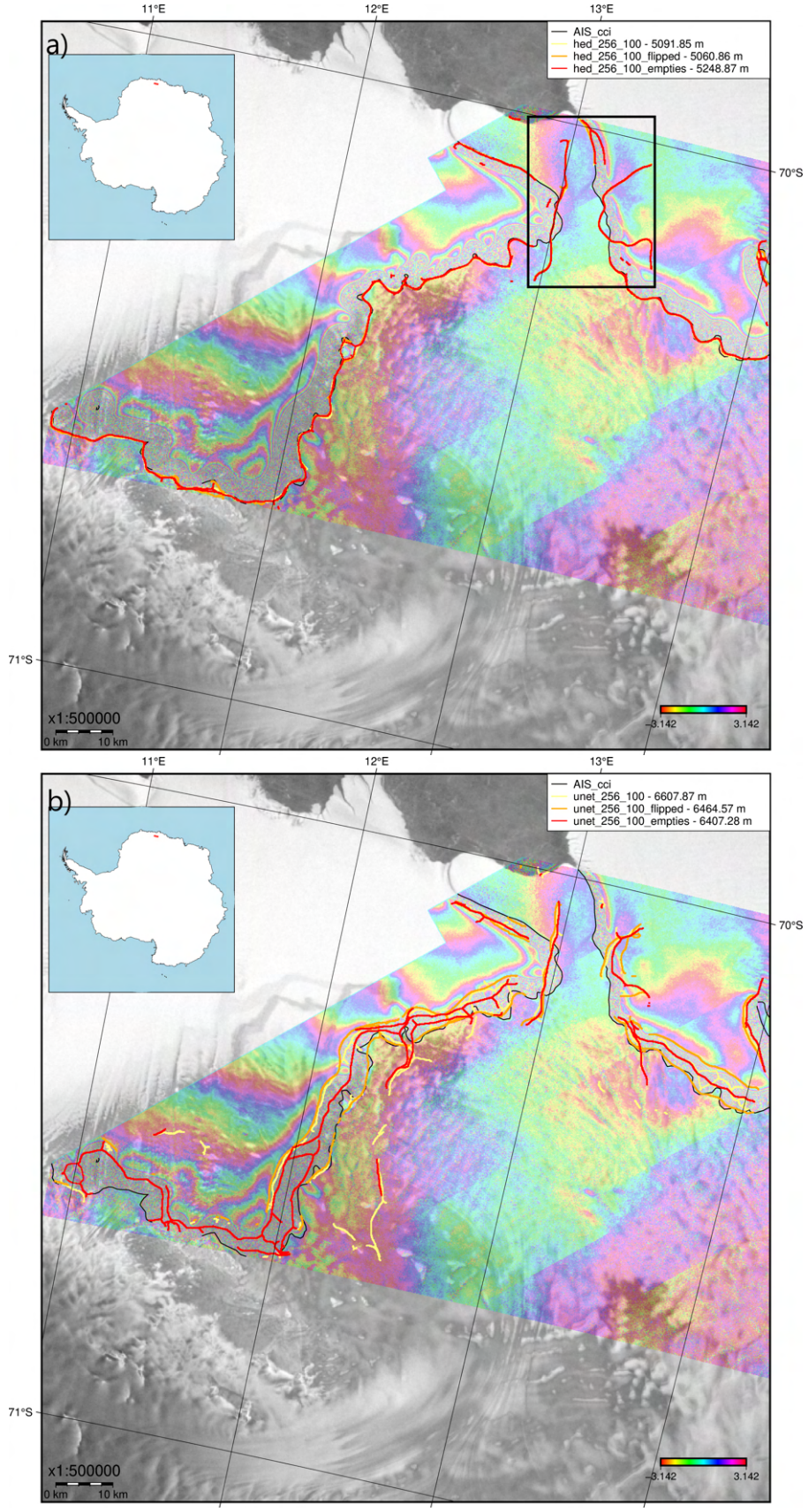


Figure 6.12: GLL predictions for a sample near Dronning Maud Land from models trained on datasets without augmented samples (yellow), augmented with random flips of training samples (orange) and augmented with empty tiles (red). a) shows predictions from HED, the black rectangle highlighting spurious branches b) predictions from UNet, nearly all of which are fragmented.

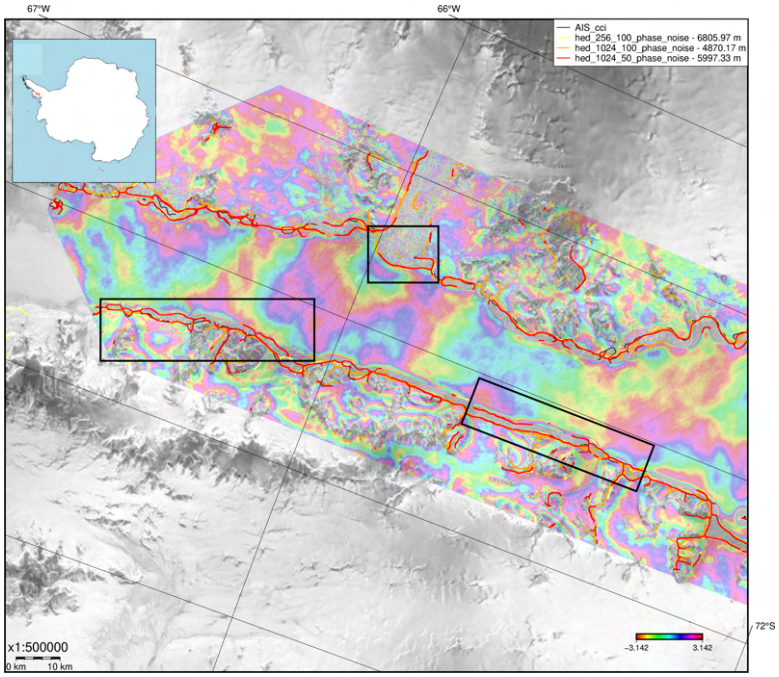


Figure 6.13: Predictions from HED tile extent experiments for a sample from George IV Ice Shelf. The black rectangles show regions where the seaward extent of the grounding zone was predicted.

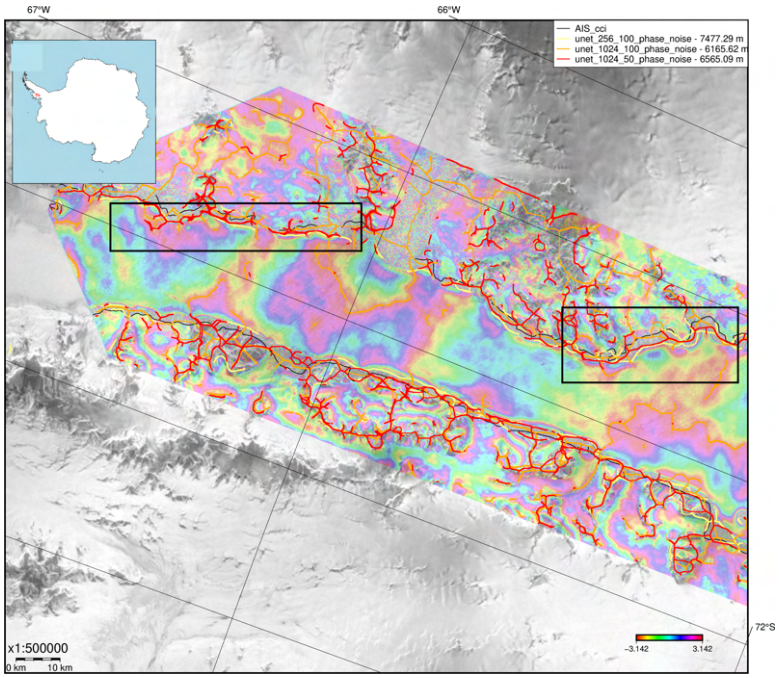


Figure 6.14: Predictions from UNet tile extent experiments for a sample from George IV Ice Shelf. The black rectangles show regions where the seaward extent of the grounding zone was predicted.

Fragmented and false positive detections from hed_1024_100_phase_noise, hed_1024_50_-phase_noise, unet_1024_100_phase_noise and unet_1024_50_phase_noise models seem to contradict the initial hypothesis stated in Section 5.2 on the effect of tile size and pixel size. In order to gain a deeper understanding of the behaviour of the networks, predictions for sample tiles containing a part of the same GLL in all three datasets were generated. This chosen part (dark patches in row 1 of Fig 6.15 and Fig. 6.16) of the grounding line is situated in the corner for the sample with an extent 256 x 256 pixels, 100 m pixel size whereas it is closer to the centre for the larger tile samples. The last rows of Fig. 6.15 and Fig. 6.16 show calculated gradients of the patch with respect to the input sample tiles during backpropagation. These gradients are used to adjust the weights for the same small patch of the grounding line. The light patches in row two of Fig. 6.15 and Fig. 6.16 show the receptive fields (Section 5.2) of the networks.

Pixels in the RF with large gradient magnitudes push the prediction probabilities towards either 0 or 1 and are considered to be more important relative to pixels for which the gradients are close to 0.

Although the RF far exceeds the extent of the considered output patch, only pixels within the patch seem to have a strong influence on the predictions, for both networks. For the 256 x 256, pixel, 100 m pixel size sample, there is a clear separation between the blue and red pixels, indicating a 'line-like' prediction. Interestingly, this behaviour is also seen for the larger tiles for UNet. However the corresponding predictions still contain branches. For HED, most of the pixels are red for the 1024 x 1024 pixels, 100 m sample, which indicates a large number of false positives. This would result in a thicker blob-like prediction, which leads to a branch-like prediction after post processing (Section 5.1). It is not clear why the higher resolution of the 1024 x 1024 pixels, 50 m pixel size sample does not improve the overall prediction.

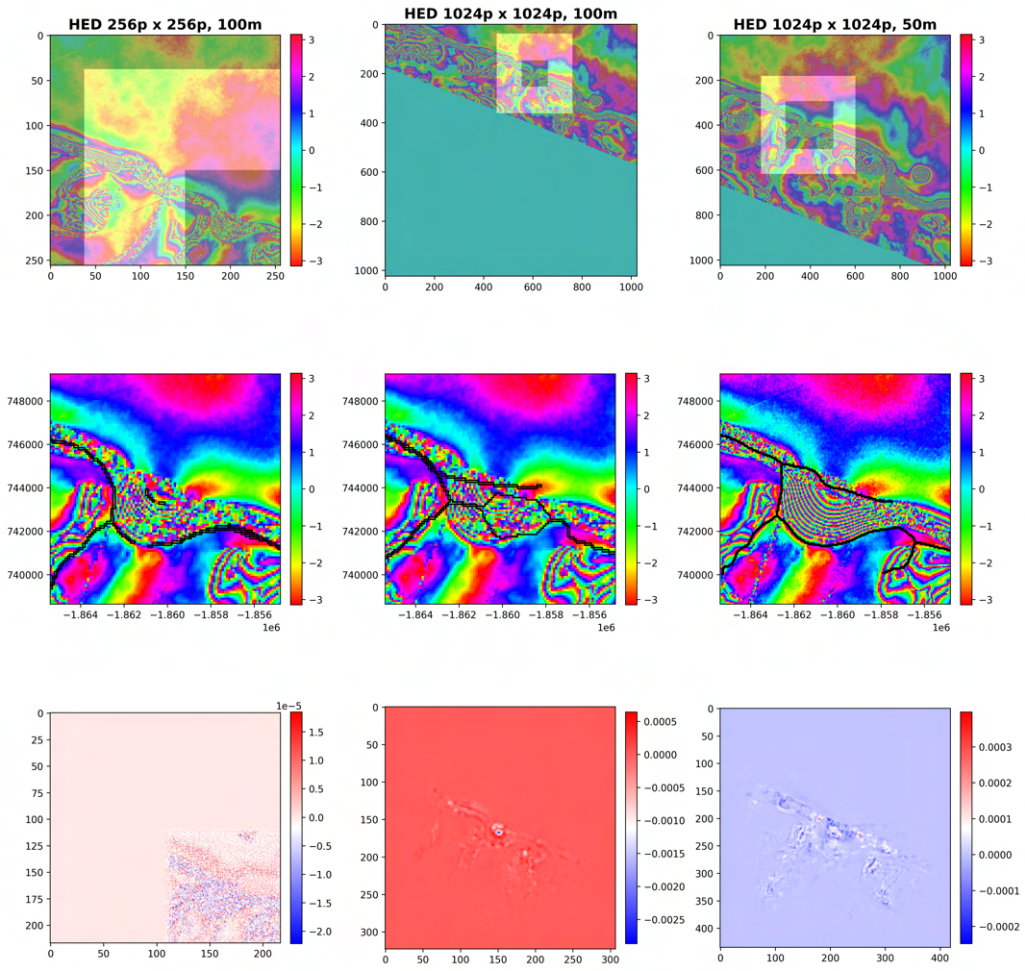


Figure 6.15: Impact of the RF of HED on predictions for different sample tile and pixel sizes. Row 1 shows the sample phases which contain the same part of a GLL indicated by the dark rectangle and the corresponding RF (light yellow rectangle) that influence this patch. Row 2 shows the corresponding predictions plotted on the phase. Row 3 shows the pixel wise gradients computed for the pixels in the highlighted rectangle in row 1.

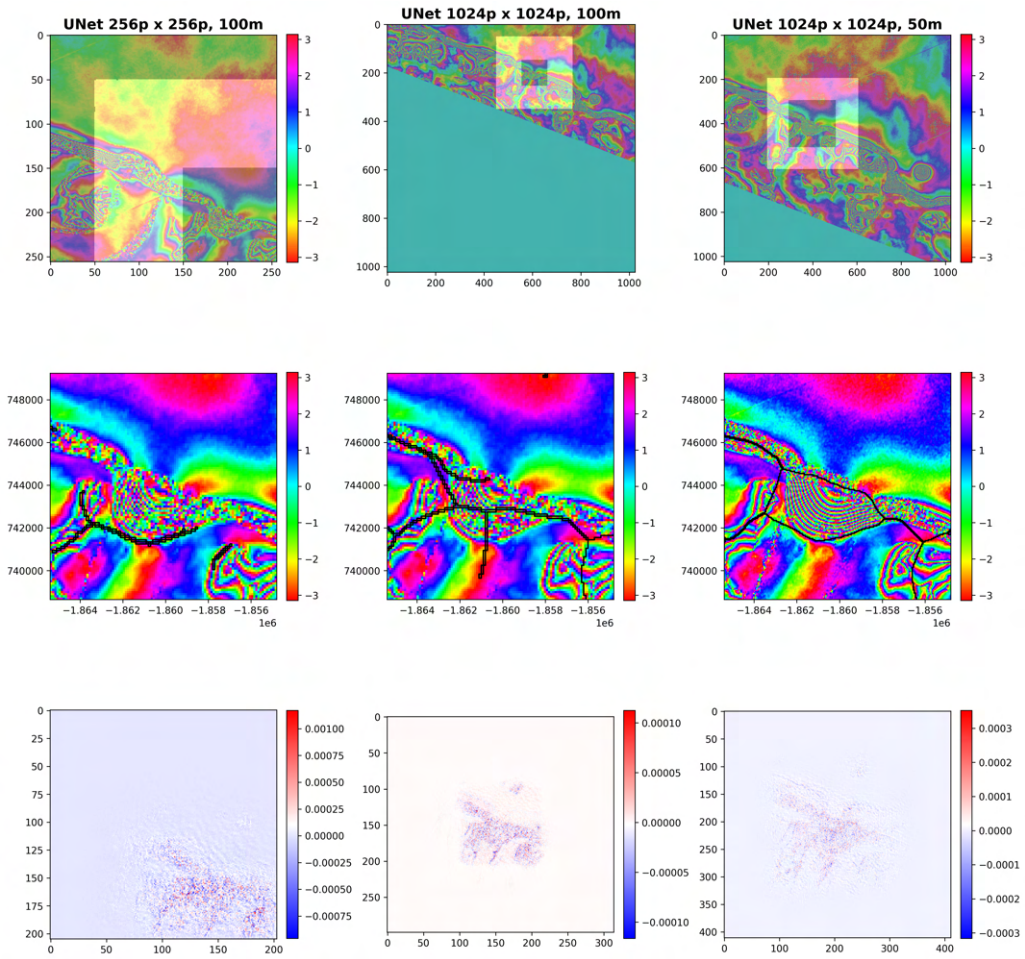


Figure 6.16: Impact of the RF of UNet on predictions for different sample tile and pixel sizes. Row 1 shows the sample phases which contain the same part of a GLL indicated by the dark rectangle and the corresponding RF (light yellow rectangle). Row 2 shows the corresponding predictions plotted on the phase. Row 3 shows the pixel wise gradients computed for the pixels in the highlighted rectangle in row 1.

7. Conclusion

7.1. Summary and conclusion

The grounding line location (GLL) is an important parameter for ice sheet mass balance and indicator of ice thickness change. Satellite based remote sensing methods provide spatially extended and temporally dense data over the margins of the Antarctic Ice Sheet. With the exception of the study from (Mohajerani et al. 2021), none of the other current grounding line detection and delineation techniques (Section 3.1) are able to delineate spatially detailed, continent-wide GLLs in a reasonable time frame. Hence there is a need to develop new techniques for efficient processing of these large datasets.

The aim of this study was to explore the potential of machine learning algorithms in achieving this task and simultaneously investigating the importance of several variables that directly or indirectly influence the GLL.

GLLs from the AIS_cci project (Section 4.1.1) with corresponding DInSAR interferograms from ERS-1/2, Sentinel-1 and TerraSAR-X missions formed the basis of the dataset on which the DNNs were trained on. An end-to-end pipeline was developed during the course of the thesis, which is capable of modifying DInSAR interferograms to be suitable for training a DNN and appending additional features to create a feature stack. Training the DNN and processing the resulting predictions are also handled by the pipeline. Two neural networks (Section 4.2) were explored for automatic delineation. The performance of the networks were quantified by means of several metrics (Section 5.4) in addition to visual inspection of predicted GLLs.

Considering the research objectives (Section 1.2) and results obtained from the conducted experiments (Chapter 6), the following conclusions can be drawn:

- In terms of feasibility, HED is able to predict complex GLL geometries reasonably well, with the best performing model achieving a median deviation of 2.1 pixels (on

interferograms of 100 m pixel size) from the ground truth. This is comparable to the results obtained in the study of (Mohajerani et al. 2021). Notwithstanding the time taken to train the DNN, generation of predictions with the developed pipeline takes a fraction of the time required for manual delineation. Even if the effort to clean up false predictions is considered, the speed up gained from automation is beneficial.

- A key finding in this work is that the DEM, ice velocity, tidal displacement and air pressure features do not significantly impact the prediction quality. In fact, the results summarised in Section 6.1.1 suggest that including some of these features might even confound the network, resulting in poor predictions. This is significant in terms of saving time and effort spent on creating the feature stack for future DInSAR interferograms. I suggest to avoid using time-invariant features and improve the perfomance of HED with a focus on phase and coherence which are provided directly by the InSAR processing.
- Computed metrics, visual inspection of test samples (Section 6.2) and the trends of training set and validation set losses (A.2) clearly show that HED outperforms UNet, despite the reported robustness and accurate classification capability of the latter. One explanation for this could be the large class imbalance (grounding line vs. non grounding line pixels) inherent in the GLL detection problem. This supports the cause for choosing DNNs that are specifically designed to deal with such issues.
- Predictions generated for a set of unlabelled DInSAR interferograms (Section 6.1.2) legitimises the feasibility of DNNs for GLL delineation. These results reinforce the results obtained in Section 6.1.1, proving that HED did not overfit to the selected test set (Fig. 5.5) and can provide useful delineations for large stacks of double difference interferograms.

7.2. Outlook

Although the experiments conducted during the course of this work have shown promising results, they provide only a glimpse into the numerous possibilities for automatic GLL

delineation on DInSAR interferograms. Enumerated below are some interesting research directions:

1. **Dataset variants**

   An advantage of the Fully Convolutional Neural Networks (FCNs) is their invariance to input tile dimensions. The training samples used in this study were resampled and tiled in order to augment the dataset. Using the training samples at their original resolution was not possible in this study due to memory constraints. In the absence of such restrictions, it would be interesting to see if the DNNs capture the intricate GLL geometries when trained on high resolution DInSAR interferograms. Additionally, this would save time spent on resampling the interferograms.

   A three class labelling system consisiting of grounded ice, grounding line and ocean in favour of the current grounding line and non grounding line classes could help reduce the number of false detections offsides (inland and seawards) of the GLL. Although it requires a one-time manual effort, the clear boundaries between land and sea that result from this kind of labelling could also yield sharper predictions, maybe even remove the need for post processing. This might be essential to improving the performance of UNet, given its prior success with similar three class problems (Section 3.2.1).

2. **Model optimization**

   Default configurations of both networks were used in this work. Rigorous hyperparameter tuning could yield better results (Chollet 2021, 5.2.1). Another aspect of optimising model performance is the choice of loss function. Experiments from (Zhang et al. 2020) and (Abraham and Khan 2019) have shown that the performance of DNNs on datasets with large class imbalance improved when using loss functions that were designed to appropriately weight labels of the minority class. These loss functions could drastically improve the performance of HED and UNet.

3. **Transfer learning**

   Given that the DNNs used in this study were trained on a dataset that spans most of

Antarctica, there are only few other regions where model transferability could be tested. Although Greenland contains only a very limited number of sites with identifiable GLLs (Hill et al. 2018, Fig.1), it would nevertheless be interesting to see if the models developed in this study are capable of predicting GLLs on these glaciers.

4. **Model explainability**

   Although the results of the feature exploration experiments revealed that the networks seem to extract the most information from pseudo coherence and wrapped phases, it was not possible to numerically quantify the contributions of individual features. Model interpretation techniques discussed in (Lundberg and Lee 2017) aim to find the influence exerted by each feature towards each prediction. These tools could provide insight into the function that is approximated by DNNs.

5. **Predicting the evolution of the grounding line position**

   Several studies have attempted to use DNNs in modelling various aspects of the cryosphere (Jouvet et al. 2021), (Hu et al. 2021), (Bolibar et al. 2020), (Leong and Horgan 2020). In a similar vein, training models to identify trends in the GLL and model ice sheet dynamics could help in understanding short term and potentially predicting long term GLL migration.

Bibliography

Abraham, Nabila and Naimul Mefraz Khan (2019). „A novel focal tversky loss function with improved attention u-net for lesion segmentation“. In: *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, pp. 683–687.

Alley, K, K Brunt, S Howard, L Padman, M Siegfried, and T Sutterly (2017). *pyTMD: Python based tidal prediction software*. URL: `https://github.com/tsutterley/pyTMD`.

Araujo, André, Wade Norris, and Jack Sim (2019). „Computing receptive fields of convolutional neural networks“. In: *Distill* 4.11, e21.

Avbelj, Janja, Rupert Müller, and Richard Bamler (2014). „A metric for polygon comparison and building extraction evaluation“. In: *IEEE Geoscience and Remote Sensing Letters* 12.1, pp. 170–174.

Bamber, Jonathan L, Richard M Westaway, Ben Marzeion, and Bert Wouters (2018). „The land ice contribution to sea level during the satellite era“. In: *Environmental Research Letters* 13.6, p. 063008.

Bamler, Richard (2000). „Principles of synthetic aperture radar“. In: *Surveys in Geophysics* 21.2, pp. 147–157.

Bamler, Richard and Philipp Hartl (1998). „Synthetic aperture radar interferometry“. In: *Inverse problems* 14.4, R1.

Baumhoer, Celia A, Andreas J Dietz, Stefan Dech, and Claudia Kuenzer (2018). „Remote sensing of antarctic glacier and ice-shelf front dynamics—A review“. In: *Remote Sensing* 10.9, p. 1445.

Baumhoer, Celia A, Andreas J Dietz, Christof Kneisel, and Claudia Kuenzer (2019). „Automated extraction of antarctic glacier and ice shelf fronts from sentinel-1 imagery using deep learning“. In: *Remote Sensing* 11.21, p. 2529.

Bindschadler, RA and PL Vornberger (1994). „Detailed elevation map of Ice Stream C, Antarctica, using satellite imagery and airborne radar“. In: *Annals of glaciology* 20, pp. 327–335.

Bindschadler, R, Hyeungu Choi, ASAID Collaborators, et al. (2011). „High-resolution image-derived grounding and hydrostatic lines for the Antarctic Ice Sheet“. In: *Digital media, National Snow and Ice Data Center, Boulder, Colorado, USA* 4932.7, pp. 4913–4936.

Bolibar, Jordi, Antoine Rabatel, Isabelle Gouttevin, Clovis Galiez, Thomas Condom, and Eric Sauquet (2020). „Deep learning applied to glacier evolution modelling“. In: *The Cryosphere* 14.2, pp. 565–584.

Brunt, KM, HA Fricker, and L Padman (2011). „Analysis of ice plains of the Filchner–Ronne Ice Shelf, Antarctica, using ICESat laser altimetry“. In: *Journal of Glaciology* 57.205, pp. 965–975.

Brunt, KM, HA Fricker, L Padman, TA Scambos, and S O'Neel (2010). „Mapping the grounding zone of the Ross Ice Shelf, Antarctica, using ICESat laser altimetry“. In: *Annals of Glaciology* 51.55, pp. 71–79.

Cheng, Daniel, Wayne Hayes, Eric Larour, Yara Mohajerani, Michael Wood, Isabella Velicogna, and Eric Rignot (2021). „Calving Front Machine (CALFIN): glacial termini dataset and automated deep learning extraction method for Greenland, 1972–2019“. In: *The Cryosphere* 15.3, pp. 1663–1675.

Choi, Dami, Christopher J Shallue, Zachary Nado, Jaehoon Lee, Chris J Maddison, and George E Dahl (2019). „On empirical comparisons of optimizers for deep learning“. In: *arXiv preprint arXiv:1910.05446*.

Chollet, Francois (2021). *Deep learning with Python*. Simon and Schuster.

Corr, Hugh FJ, CSM Doake, A Jenkins, and David G Vaughan (2001). „Investigations of an "ice plain" in the mouth of Pine Island Glacier, Antarctica“. In: *Journal of Glaciology* 47.156, pp. 51–57.

Dawson, GJ and JL Bamber (2017). „Antarctic Grounding Line Mapping From CryoSat-2 Radar Altimetry“. In: *Geophysical Research Letters* 44.23, pp. 11–886.

Deng, Zhipeng, Hao Sun, Shilin Zhou, Juanping Zhao, Lin Lei, and Huanxin Zou (2018). „Multi-scale object detection in remote sensing imagery with convolutional neural networks“. In: *ISPRS journal of photogrammetry and remote sensing* 145, pp. 3–22.

Dong, Hao, Hao Dong, Zihan Ding, Shanghang Zhang, and Chang (2020). *Deep Reinforcement Learning*. Springer.

Dong, Yuting, Ji Zhao, Dana Floricioiu, and Lukas Krieger (2022). „Automatic calving front extraction from digital elevation model-derived data“. In: *Remote Sensing of Environment* 270, p. 112854.

Fox, Adrian J, A Paul, and R Cooper (1994). „Measured properties of the Antarctic ice sheet derived from the SCAR Antarctic digital database“. In: *Polar Record* 30.174, pp. 201–206.

Fox-Kemper, B, HT Hewitt, C Xiao, G Aðalgeirsdóttir, SS Drijfhout, TL Edwards, NR Golledge, M Hemer, RE Kopp, G Krinner, A Mix, D Notz, SS Nowicki, IS Nurhati, L Ruiz, JB Sallée, ABA Slangen, and Y Yu (2021). „Ocean, Cryosphere and Sea Level Change (Ch. 9 of Climate Change 2021: The Physical Science Basis)“. In: *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Fricker, HA, H Amanda, and L Padman (2006). „Ice shelf grounding zone structure from ICESat laser altimetry“. In: *Geophysical Research Letters* 33.15.

Fricker, HA, R Coleman, L Padman, TA Scambos, J Bohlander, and KM Brunt (2009). „Mapping the grounding zone of the Amery Ice Shelf, East Antarctica using InSAR, MODIS and ICESat“. In: *Antarctic Science* 21.5, pp. 515–532.

Fricker, Helen Amanda, Ian Allison, Mike Craven, Glenn Hyland, Andrew Ruddell, Neal Young, Richard Coleman, Matt King, Kim Krebs, and Sergey Popov (2002). „Redefinition of the Amery ice shelf, East Antarctica, grounding zone“. In: *Journal of Geophysical Research: Solid Earth* 107.B5, ECV–1.

Friedl, Peter, Frank Weiser, Anke Fluhrer, and Matthias H Braun (2020). „Remote sensing of glacier and ice sheet grounding lines: A review“. In: *Earth-Science Reviews* 201, p. 102948.

Gardner, Alex S, Geir Moholdt, Ted Scambos, Mark Fahnstock, Stefan Ligtenberg, Michiel Van Den Broeke, and Johan Nilsson (2018). „Increased West Antarctic and unchanged East Antarctic ice discharge over the last 7 years“. In: *The Cryosphere* 12.2, pp. 521–547.

Glorot, Xavier and Yoshua Bengio (2010). „Understanding the difficulty of training deep feedforward neural networks“. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, pp. 249–256.

Goldstein, Richard M, Hermann Engelhardt, Barclay Kamb, and Richard M Frolich (1993). „Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream“. In: *Science* 262.5139, pp. 1525–1530.

Gonzalez, Fernando Rodriguez, Nico Adam, Alessandro Parizzi, and Ramon Brcic (2013). „The integrated wide area processor (IWAP): A processor for wide area persistent scatterer interferometry“. In: *ESA Living Planet Symposium*. Vol. 722, p. 353.

Groh, A (2021). *Product User Guide (PUG) for the Antarctic Ice Sheet cci project of ESA's Climate Cha*
URL: https://climate.esa.int/media/documents/ST-UL-ESA-AISCCI-PUG-0001.pdf.

Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. (2018). „Recent advances in convolutional neural networks“. In: *Pattern Recognition* 77, pp. 354–377.

Haseloff, Marianne and Olga V Sergienko (2018). „The effect of buttressing on grounding line dynamics“. In: *Journal of Glaciology* 64.245, pp. 417–431.

He, Nanjun, Leyuan Fang, Shutao Li, Javier Plaza, and Antonio Plaza (2019). „Skip-connected covariance network for remote sensing scene classification“. In: *IEEE transactions on neural networks and learning systems* 31.5, pp. 1461–1474.

Heidler, Konrad, Lichao Mou, Celia Baumhoer, Andreas Dietz, and Xiao Xiang Zhu (2021). „HED-UNet: Combined segmentation and edge detection for monitoring the Antarctic coastline“. In: *IEEE Transactions on Geoscience and Remote Sensing* 60, pp. 1–14.

Hill, Emily A, J Rachel Carr, Chris R Stokes, and G Hilmar Gudmundsson (2018). „Dynamic changes in outlet glaciers in northern Greenland from 1948 to 2015“. In: *The Cryosphere* 12.10, pp. 3243–3263.

Holdsworth, Gerald (1969). „Flexure of a floating ice tongue“. In: *Journal of Glaciology* 8.54, pp. 385–397.

Horwath, M and A Groh (2018). *Product Validation and Intercomparison Report (PVIR) for the Antarctic_Ice_Sheet_cci project of ESA's Climate Change Initiative, version* 1.3, 04. URL: http://www.esa-icesheets-antarctica-cci.org/.

Hu, Zhongyang, Peter Kuipers Munneke, Stef Lhermitte, Maaike Izeboud, and Michiel Van Den Broeke (2021). „Improving surface melt estimation over the Antarctic Ice Sheet using deep learning: a proof of concept over the Larsen Ice Shelf“. In: *The Cryosphere* 15.12, pp. 5639–5658.

Huber, M (2020). *TanDEM-X PolarDEM Product Description, prepared by German remote sensing data center (DFD) and Earth Observation Center*. URL: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-11882/20871_read-66374.

Jezek, Kenneth C (2002). „RADARSAT-1 Antarctic Mapping Project: change-detection and surface velocity campaign“. In: *Annals of Glaciology* 34, pp. 263–268.

Joughin, Ian (2002). „Ice-sheet velocity mapping: a combined interferometric and speckle-tracking approach“. In: *Annals of Glaciology* 34, pp. 195–201.

Joughin, Ian, David E Shean, Ben E Smith, and P Dutrieux (2016). „Grounding line variability and subglacial lake drainage on Pine Island Glacier, Antarctica“. In: *Geophysical Research Letters* 43.17, pp. 9093–9102.

Jouvet, Guillaume, Guillaume Cordonnier, Byungsoo Kim, Martin Lüthi, Andreas Vieli, and Andy Aschwanden (2021). „Deep learning speeds up ice flow modelling by several orders of magnitude“. In: *Journal of Glaciology*, pp. 1–14.

Kalnay, Eugenia, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. (1996). „The NCEP/NCAR 40-year reanalysis project“. In: *Bulletin of the American meteorological Society* 77.3, pp. 437–472.

Kingma, Diederik P and Jimmy Ba (2014). „Adam: A method for stochastic optimization“. In: *arXiv preprint arXiv:1412.6980*.

Klinger, Tobias, Marcel Ziems, Christian Heipke, Hans Werner Schenke, and Norbert Ott (2011). „Antarctic coastline detection using snakes“. In: *Photogrammetrie-Fernerkundung-Geoinformation*, pp. 421–434.

Krieger, Lukas and Dana Floricioiu (2017). „Automatic calving front delienation on TerraSAR-X and Sentinel-1 SAR imagery“. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). „Deep learning“. In: *nature* 521.7553, pp. 436–444.

LeCun, Yann, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard, and Lawrence Jackel (1989). „Handwritten digit recognition with a back-propagation network“. In: *Advances in neural information processing systems* 2.

Lederer, Johannes (2021). „Activation functions in artificial neural networks: A systematic overview“. In: *arXiv preprint arXiv:2101.09957*.

Leong, Wei Ji and Huw Joseph Horgan (2020). „DeepBedMap: a deep neural network for resolving the bed topography of Antarctica“. In: *The Cryosphere* 14.11, pp. 3687–3705.

Li, T, GJ Dawson, SJ Chuter, and JL Bamber (2021). *ICESat-2-derived grounding zone product for Antarctica, University of Bristol [data set]*.

Li, Tian, Geoffrey J Dawson, Stephen J Chuter, and Jonathan L Bamber (2022). „A high-resolution Antarctic grounding zone product from ICESat-2 laser altimetry“. In: *Earth System Science Data* 14.2, pp. 535–557.

Liu, H and KC Jezek (2004). „Automated extraction of coastline from satellite imagery by integrating Canny edge detection and locally adaptive thresholding methods“. In: *International journal of remote sensing* 25.5, pp. 937–958.

Lundberg, Scott M and Su-In Lee (2017). „A unified approach to interpreting model predictions“. In: *Advances in neural information processing systems* 30.

Luo, Wenjie, Yujia Li, Raquel Urtasun, and Richard Zemel (2016). „Understanding the effective receptive field in deep convolutional neural networks“. In: *Advances in neural information processing systems* 29.

Marsh, OJ, Wolfgang Rack, D Floricioiu, Nicholas R Golledge, and W Lawson (2013). „Tidally induced velocity variations of the Beardmore Glacier, Antarctica, and their representation in satellite measurements of ice velocity“. In: *The Cryosphere* 7.5, pp. 1375–1384.

Matsuoka, Kenichi, Richard CA Hindmarsh, Geir Moholdt, Michael J Bentley, Hamish D Pritchard, Joel Brown, Howard Conway, Reinhard Drews, Gaël Durand, Daniel Goldberg, et al. (2015). „Antarctic ice rises and rumples: Their properties and significance for ice-sheet dynamics and evolution“. In: *Earth-science reviews* 150, pp. 724–745.

Matsuoka, Kenichi, Anders Skoglund, George Roth, Jean de Pomereu, Huw Griffiths, Robert Headland, Brad Herried, Katsuro Katsumata, Anne Le Brocq, Kathy Licht, et al. (2021). „Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands“. In: *Environmental Modelling & Software* 140, p. 105015.

Mätzler, Christian (1987). „Applications of the interaction of microwaves with the natural snow cover“. In: *Remote sensing reviews* 2.2, pp. 259–387.

McCulloch, Warren S and Walter Pitts (1943). „A logical calculus of the ideas immanent in nervous activity“. In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133.

Meredith, M, M Sommerkorn, S Cassotta, C Derksen, A Ekaykin, A Hollowed, G Kofinas, A Mackintosh, J Melbourne-Thomas, MMC Muelbert, et al. (2019). „Polar Regions. Chapter 3, IPCC Special Report on the Ocean and Cryosphere in a Changing Climate“. In:

Milillo, P, E Rignot, P Rizzoli, B Scheuchl, J Mouginot, JL Bueso-Bello, P Prats-Iraola, and L Dini (2022). „Rapid glacier retreat rates observed in West Antarctica“. In: *Nature Geoscience*, pp. 1–6.

Minetto, Rodrigo, Maurıcio Pamplona Segundo, and Sudeep Sarkar (2019). „Hydra: An ensemble of convolutional neural networks for geospatial land classification“. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.9, pp. 6530–6541.

Mohajerani, Yara, Seongsu Jeong, Bernd Scheuchl, Isabella Velicogna, Eric Rignot, and Pietro Milillo (2021). „Automatic delineation of glacier grounding lines in differential interferometric synthetic-aperture radar data using deep learning“. In: *Scientific reports* 11.1, pp. 1–10.

Mohajerani, Yara, Michael Wood, Isabella Velicogna, and Eric Rignot (2019). „Detection of glacier calving margins with convolutional neural networks: A case study“. In: *Remote Sensing* 11.1, p. 74.

Moreira, Alberto, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou (2013). „A tutorial on synthetic aperture radar“. In: *IEEE Geoscience and remote sensing magazine* 1.1, pp. 6–43.

Muir, A (2021). *System Specification Document (SSD) for the Antarctic Ice Sheet Climate project of ESA's Cl* URL: https://climate.esa.int/media/documents/ST-UL-ESA-AISCCI-SSD-001-v1.1.pdf.

Murphy, Kevin P (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nagler, Thomas, Helmut Rott, Markus Hetzenecker, Jan Wuite, and Pierre Potin (2015). „The Sentinel-1 mission: New opportunities for ice sheet observations“. In: *Remote Sensing* 7.7, pp. 9371–9389.

Oppenheimer, M, B.C Glavovic, J Hinkel, R van de Wal, AK Magnan, A Abd-Elgawad, R Cai, M Cifuentes-Jara, RM DeConto, T Ghosh, J Hay, F Isla, B Marzeion, B Meyssignac, and Z Sebesvari (2019). „2019: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities“. In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*.

Padman, Laurie, Svetlana Y Erofeeva, and Helen Amanda Fricker (2008). „Improving Antarctic tide models by assimilation of ICESat laser altimetry over ice shelves“. In: *Geophysical Research Letters* 35.22.

Pang, Long-Gang, Kai Zhou, Nan Su, Hannah Petersen, Horst Stöcker, and Xin-Nian Wang (2018). „An equation-of-state-meter of quantum chromodynamics transition from deep learning“. In: *Nature communications* 9.1, pp. 1–6.

Paterson, WSB (2016). *The Physics of Glaciers 3rd edition*. Vol. 41. 2. [Oxford, Oxfordshire]: Published for the International Association of . . ., pp. 272–272.

Pattyn, Frank (2018). „The paradigm shift in Antarctic ice sheet modelling“. In: *Nature communications* 9.1, pp. 1–3.

Pegler, Samuel S (2018). „Marine ice sheet dynamics: the impacts of ice-shelf buttressing“. In: *Journal of Fluid Mechanics* 857, pp. 605–647.

Riedel, Björn, Uwe Nixdorf, Michael Heinert, Alfons Eckstaller, and Christoph Mayer (1999). „The response of the Ekströmisen (Antarctica) grounding zone to tidal forcing“. In: *Annals of Glaciology* 29, pp. 239–242.

Rignot, E (1996). „Tidal motion, ice velocity and melt rate of Petermann Gletscher, Greenland, measured from radar interferometry“. In: *Journal of Glaciology* 42.142, pp. 476–485.

Rignot, E, J Mouginot, and B Scheuchl (2011). „Antarctic grounding line mapping from differential satellite radar interferometry“. In: *Geophysical Research Letters* 38.10.

– (2016). „MEaSUREs Antarctic Grounding Line from Differential Satellite Radar Interferometry, Version 2“. In: *NASA*. URL: https://doi.org/10.5067/IKBWW4RYHF1Q.

Rignot, EJ (1998). „Fast recession of a West Antarctic glacier“. In: *Science* 281.5376, pp. 549–551.

Rignot, Eric, J Mouginot, Bernd Scheuchl, Michiel Van Den Broeke, Melchior J Van Wessem, and Mathieu Morlighem (2019). „Four decades of Antarctic Ice Sheet mass balance from 1979–2017“. In: *Proceedings of the National Academy of Sciences* 116.4, pp. 1095–1103.

Rignot, Eric, Jeremie Mouginot, Mathieu Morlighem, Helene Seroussi, and Bernd Scheuchl (2014). „Widespread, rapid grounding line retreat of Pine Island, Thwaites, Smith, and Kohler glaciers, West Antarctica, from 1992 to 2011“. In: *Geophysical Research Letters* 41.10, pp. 3502–3509.

Rignot, Eric, Laurence Padman, Douglas R MacAyeal, and M Schmeltz (2000). „Observation of ocean tides below the Filchner and Ronne Ice Shelves, Antarctica, using synthetic aperture radar interferometry: Comparison with tide model predictions“. In: *Journal of Geophysical Research: Oceans* 105.C8, pp. 19615–19630.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). „U-net: Convolutional networks for biomedical image segmentation“. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Rosen, Paul A, Scott Hensley, Ian R Joughin, Fuk K Li, Soren N Madsen, Ernesto Rodriguez, and Richard M Goldstein (2000). „Synthetic aperture radar interferometry“. In: *Proceedings of the IEEE* 88.3, pp. 333–382.

Rosenau, R, E Schwalbe, H-G Maas, Michael Baessler, and R Dietrich (2013). „Grounding line migration and high-resolution calving dynamics of Jakobshavn Isbræ, West Greenland“. In: *Journal of Geophysical Research: Earth Surface* 118.2, pp. 382–395.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986). „Learning representations by back-propagating errors“. In: *nature* 323.6088, pp. 533–536.

Scambos, Ted A, Terry M Haran, MA Fahnestock, TH Painter, and Jennifer Bohlander (2007). „MODIS-based Mosaic of Antarctica (MOA) data sets: Continent-wide surface morphology and snow grain size“. In: *Remote Sensing of Environment* 111.2-3, pp. 242–257.

Schoof, Christian (2007). „Ice sheet grounding line dynamics: Steady states, stability, and hysteresis“. In: *Journal of Geophysical Research: Earth Surface* 112.F3.

Seale, Anthony, Poul Christoffersen, Ruth I Mugford, and Martin O'Leary (2011). „Ocean forcing of the Greenland Ice Sheet: Calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers“. In: *Journal of Geophysical Research: Earth Surface* 116.F3.

Shepherd, Andrew, Erik Ivins, Eric Rignot, Ben Smith, Michiel Van Den Broeke, Isabella Velicogna, Pippa Whitehouse, Kate Briggs, Ian Joughin, Gerhard Krinner, et al. (2018). „Mass balance of the Antarctic Ice Sheet from 1992 to 2017“. In: *Nature* 558, pp. 219–222.

Shimada, Masanobu (2010). „Ortho-rectification and slope correction of SAR data using DEM and its accuracy evaluation“. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3.4, pp. 657–671.

Shorten, Connor and Taghi M Khoshgoftaar (2019). „A survey on image data augmentation for deep learning“. In: *Journal of big data* 6.1, pp. 1–48.

Smith, AM (1991). „The use of tiltmeters to study the dynamics of Antarctic ice-shelf grounding lines“. In: *Journal of Glaciology* 37.125, pp. 51–58.

Smith, Ben, Helen A Fricker, Alex S Gardner, Brooke Medley, Johan Nilsson, Fernando S Paolo, Nicholas Holschuh, Susheel Adusumilli, Kelly Brunt, Bea Csatho, et al. (2020). „Pervasive ice sheet mass loss reflects competing ocean and atmosphere processes“. In: *Science* 368.6496, pp. 1239–1242.

Sohn, Hong-Gyoo and KC Jezek (1999). „Mapping ice sheet margins from ERS-1 SAR and SPOT imagery“. In: *International Journal of Remote Sensing* 20.15-16, pp. 3201–3216.

Sun, Chen, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta (2017). „Revisiting unreasonable effectiveness of data in deep learning era“. In: *Proceedings of the IEEE international conference on computer vision*, pp. 843–852.

Tasar, Onur, Yuliya Tarabalka, and Pierre Alliez (2019). „Incremental learning for semantic segmentation of large-scale remote sensing data“. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.9, pp. 3524–3537.

Uieda, L, D Tian, WJ Leong, L Toney, W Schlitzer, M Grund, D Newton, M Ziebarth, M Jones, and P Wessel (2021). „PyGMT: A Python interface for the generic mapping tools“. In:

Vaughan, DG, JC Comiso, I Allison, J Carrasco, G Kaser, R Kwok, P Mote, T Murray, F Paul, J Ren, et al. (2013). „Climate change 2013: the physical science basis“. In: *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Observations: cryosphere*, pp. 317–382.

Wild, Christian T, Oliver J Marsh, and Wolfgang Rack (2019). „Differential interferometric synthetic aperture radar for tide modelling in Antarctic ice-shelf grounding zones“. In: *The Cryosphere* 13.12, pp. 3171–3191.

Wouters, Bert, Alba Martin-Español, Velt Helm, Thomas Flament, Jan M van Wessem, Stefan RM Ligtenberg, Michiel Roland Van den Broeke, and Jonathan L Bamber (2015). „Dynamic thinning of glaciers on the Southern Antarctic Peninsula“. In: *Science* 348.6237, pp. 899–903.

Wuite, Jan, Thomas Nagler, Noel Gourmelen, Maria Jose Escorihuela, Anna E Hogg, and Mark R Drinkwater (2019). „Sub-Annual calving front migration, area change and calving

rates from Swath Mode CryoSat-2 altimetry, on Filchner-Ronne Ice Shelf, Antarctica“. In: *Remote Sensing* 11.23, p. 2761.

Xie, Saining and Zhuowen Tu (2015). „Holistically-nested edge detection“. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1395–1403.

Xie, Surui, Timothy H Dixon, Denis Voytenko, Fanghui Deng, and David M Holland (2018). „Grounding line migration through the calving season at Jakobshavn Isbræ, Greenland, observed with terrestrial radar interferometry“. In: *The Cryosphere* 12.4, pp. 1387–1400.

Yagüe- Martınez, Néstor, Pau Prats- Iraola, Fernando Rodriguez Gonzalez, Ramon Brcic, Robert Shau, Dirk Geudtner, Michael Eineder, and Richard Bamler (2016). „Interferometric processing of Sentinel-1 TOPS data“. In: *IEEE Transactions on Geoscience and Remote Sensing* 54.4, pp. 2220–2234.

Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi (2018). „Convolutional neural networks: an overview and application in radiology“. In: *Insights into imaging* 9.4, pp. 611–629.

Zebker, Howard (2021). „Accuracy of a model-free algorithm for temporal InSAR tropospheric correction“. In: *Remote Sensing* 13.3, p. 409.

Zhang, Linbin, Caiguang Zhang, Sinong Quan, Huaxin Xiao, Gangyao Kuang, and Li Liu (2020). „A class imbalance loss for imbalanced object recognition“. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, pp. 2778–2792.

Zhang, Tongjie Y and Ching Y. Suen (1984). „A fast parallel algorithm for thinning digital patterns“. In: *Communications of the ACM* 27.3, pp. 236–239.

Zhu, Xiao Xiang, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer (2017). „Deep learning in remote sensing: A comprehensive review and list of resources“. In: *IEEE Geoscience and Remote Sensing Magazine* 5.4, pp. 8–36.

A. Appendix

A.1. Distribution of PoLiS distances

The figures below show the distribution of prediction to ground truth, ground truth to prediction and overall deviations. In general, all the distributions are right skewed.

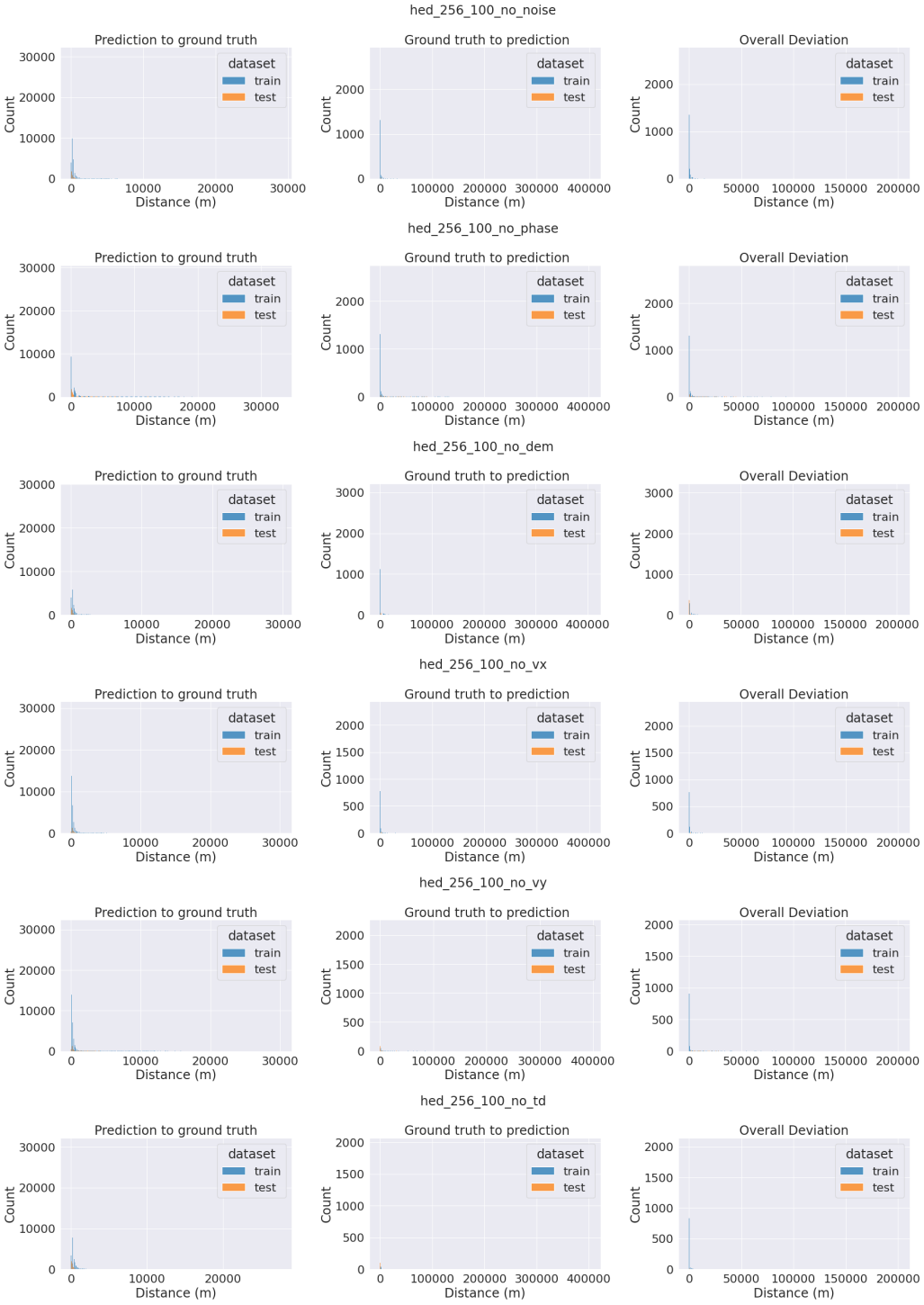


Figure A.1: Distribution of PoLiS deviations for feature exploration experiments.

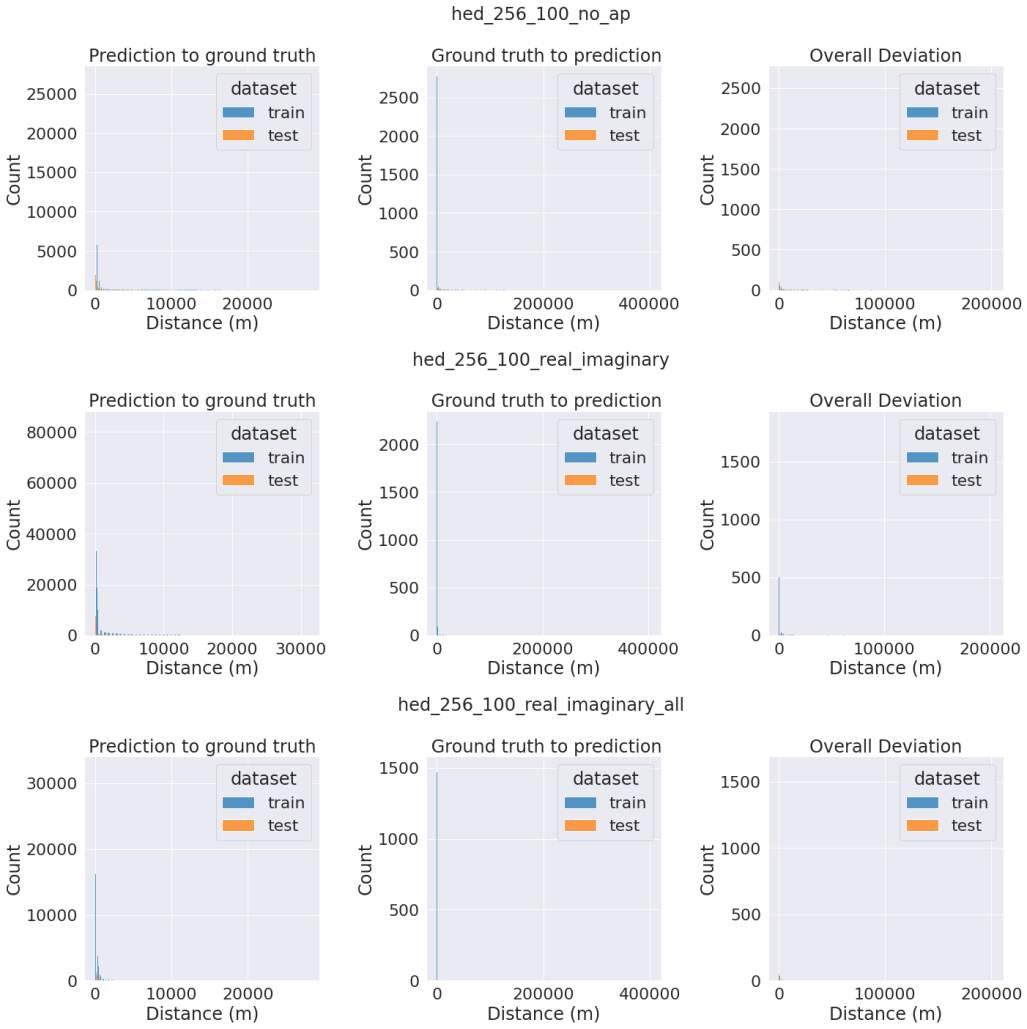


Figure A.2: Distribution of PoLiS deviations for feature exploration experiments.

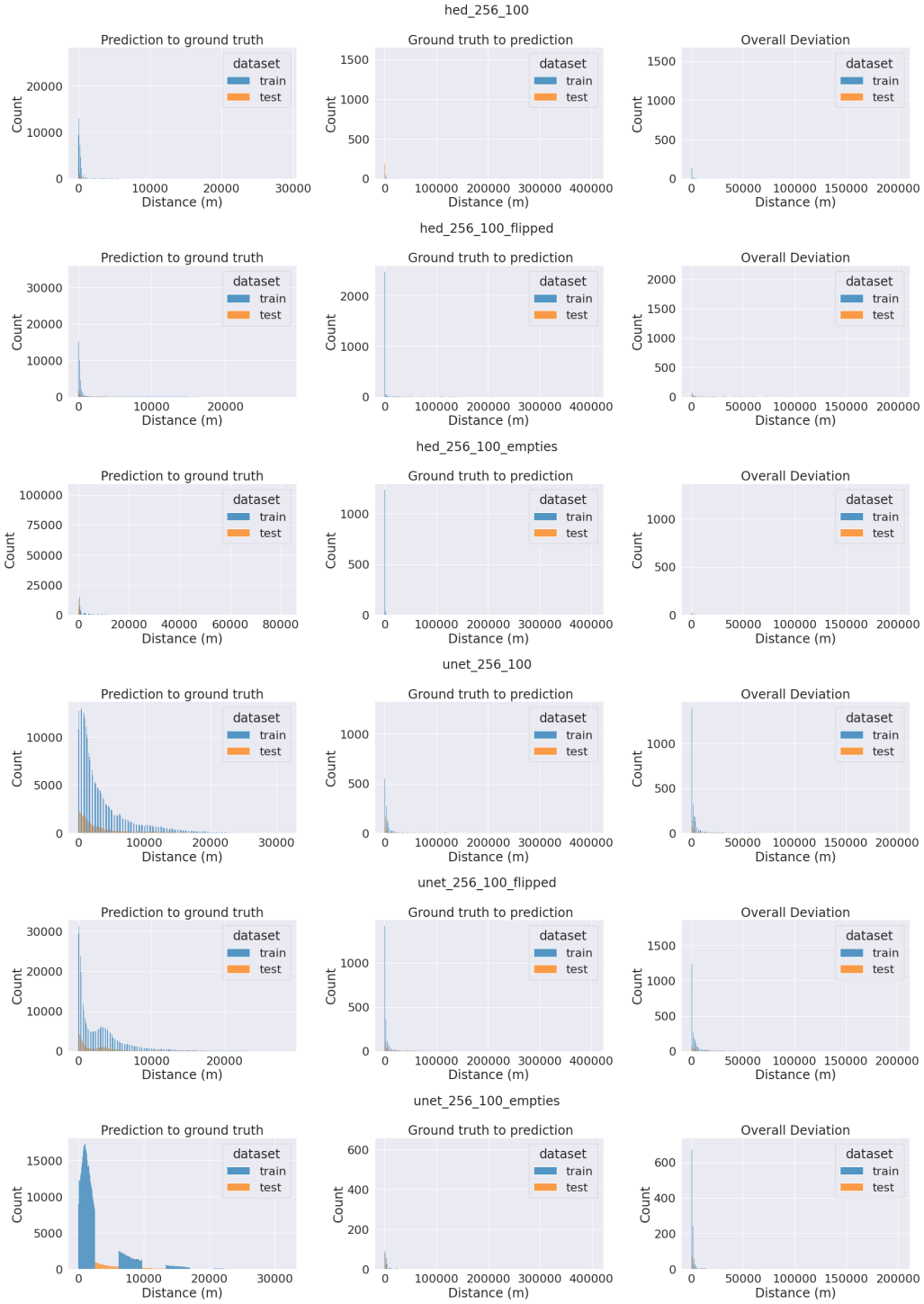


Figure A.3: Distribution of PoLiS deviations for augmentation experiments.

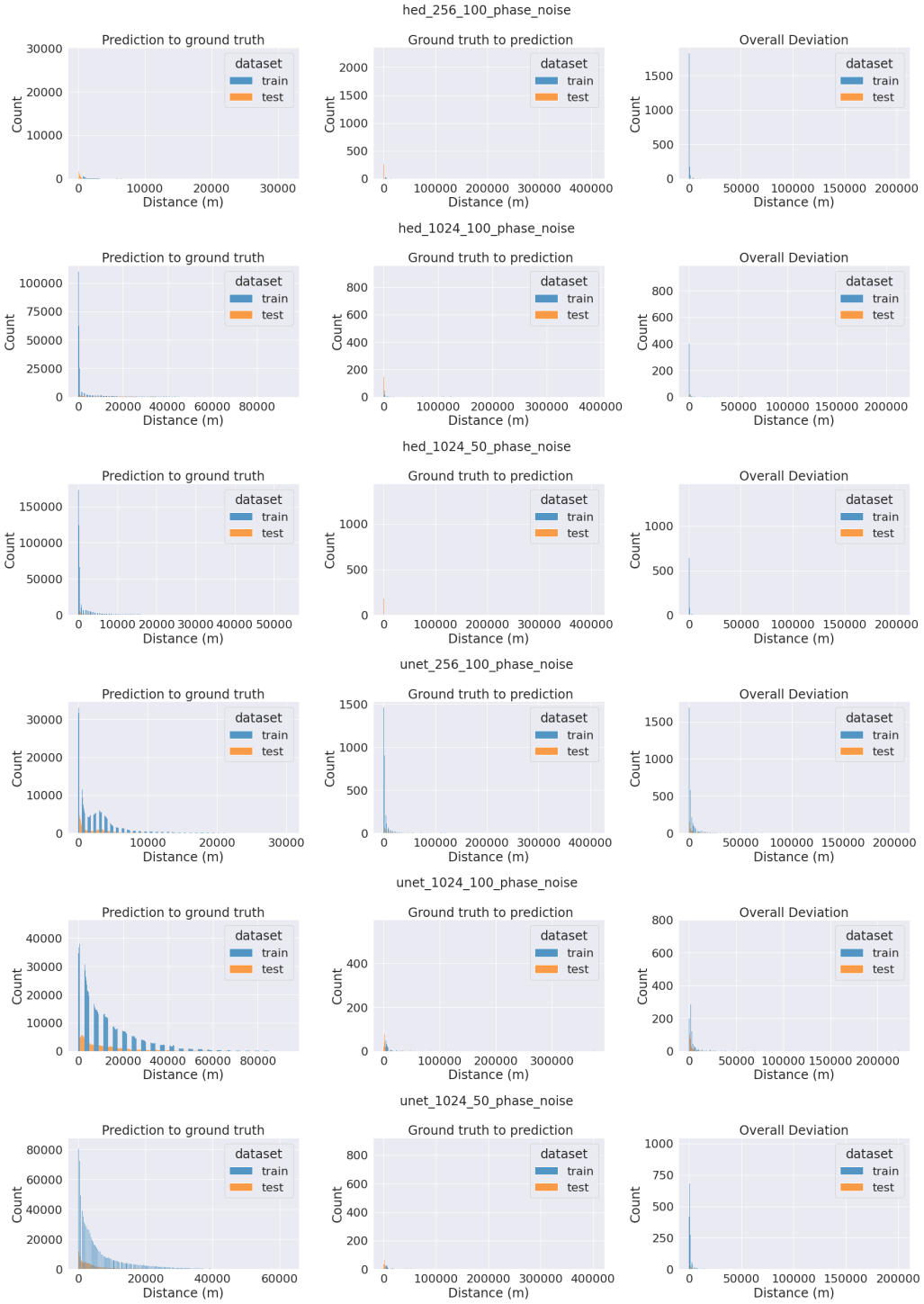


Figure A.4: Distribution of PoLiS deviations for tile extent experiments.

A.2. Training and validation loss curves

The figures below show the trend of model loss during training and validation phases, The differences in the between the training and validation loss indicates that the model does not generalise to the test set samples.

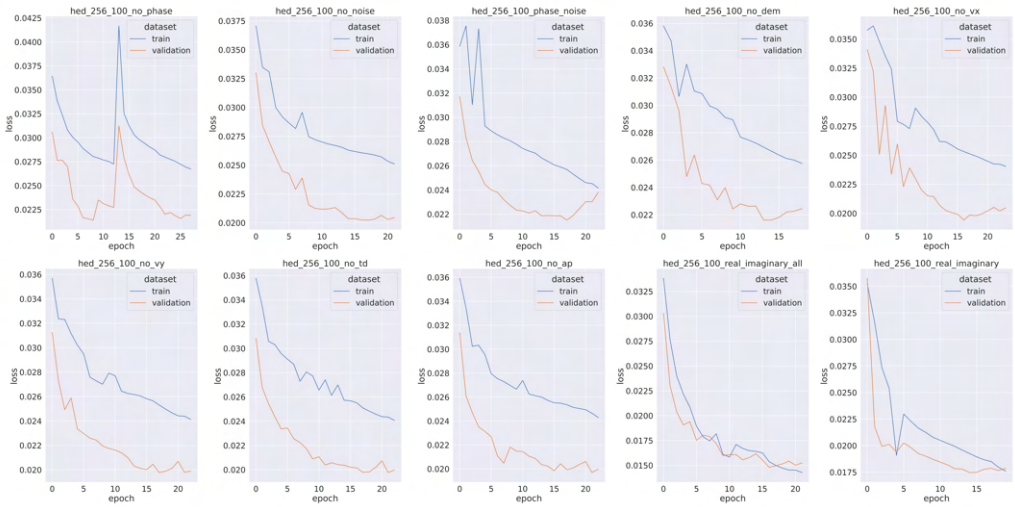


Figure A.5: Training and validation loss curves for feature exploration experiments

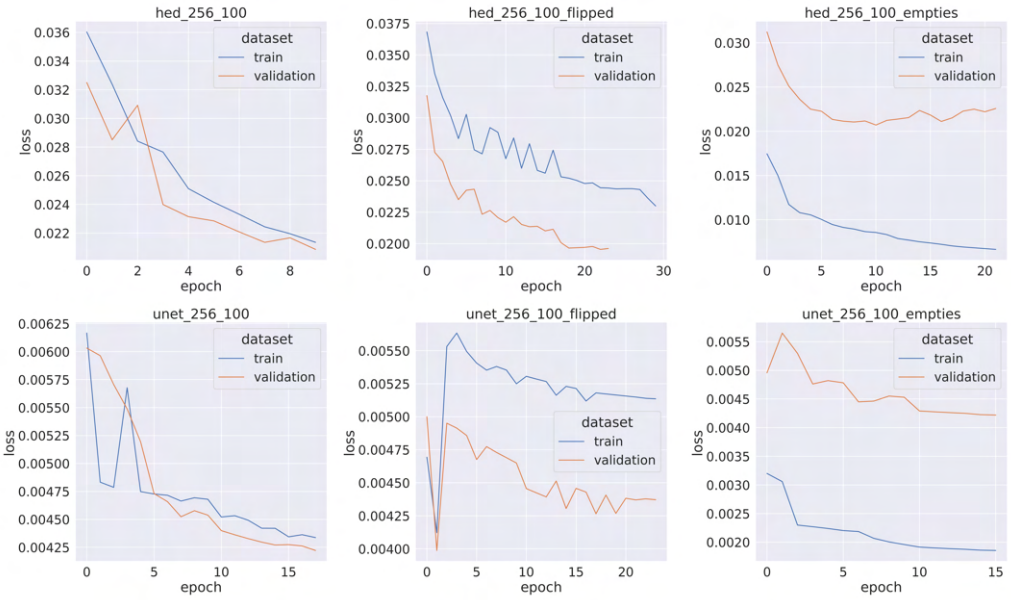


Figure A.6: Training and validation loss curves for augmentation experiments

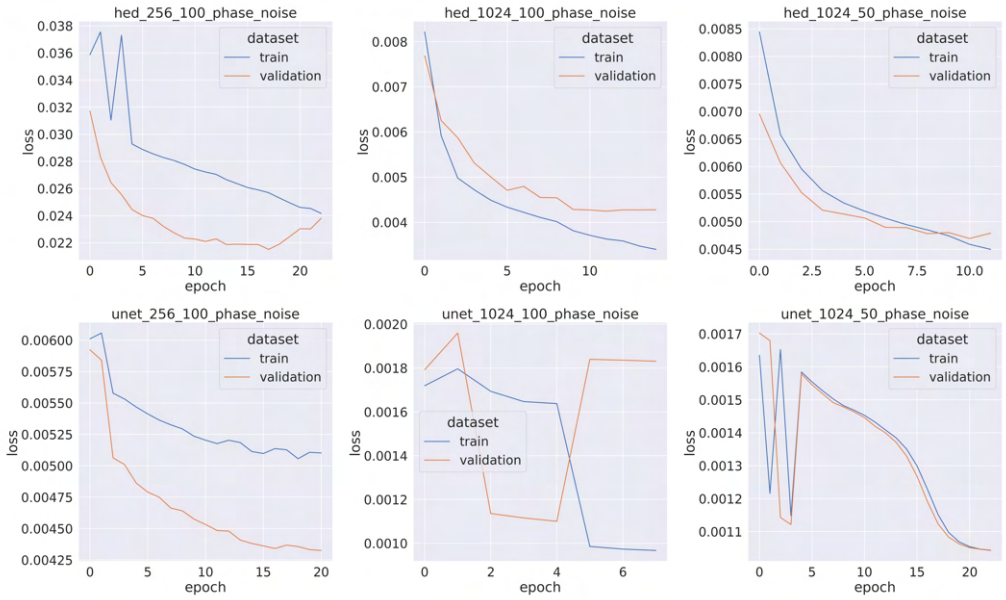


Figure A.7: Training and validation loss curves for tile extent experiments

A.3. ODS F_1 scores

The below tables list the ODS F_1 scores and false negatives for all experiments (section 5.2)

Table A.1: ODS F_1 scores for feature exploration experiments

Experiment ID	ODS F_1 score	False negatives (lines/1232)
hed_256_100_flipped	0.19	0
hed_256_100_phase_noise	0.19	0
hed_256_100_no_noise	0.2	0
hed_256_100_no_phase	0.17	0
hed_256_100_no_dem	0.19	0
hed_256_100_no_vx	0.19	0
hed_256_100_no_vy	0.19	0
hed_256_100_no_td	0.19	1
hed_256_100_no_ap	0.2	1
hed_256_100_real_imaginary	0.15	0
hed_256_100_real_imaginary_all	0.2	0

Table A.2: ODS F_1 scores for augmentation experiments

Experiment ID	ODS F_1 score	False negatives (lines/1232)
hed_256_100	0.19	0
hed_256_100_flipped	0.19	0
hed_256_100_empties	0.13	0
unet_256_100	0.0359	1
unet_256_100_flipped	0.054	1
unet_256_100_empties	0.032	0

Table A.3: ODS F_1 scores for tile extent experiments

Experiment ID	ODS F_1 score	False negatives (lines/1232)
hed_256_100_phase_noise	0.19	0
hed_1024_100_phase_noise	0.10	0
hed_1024_50_phase_noise	0.069	1
unet_256_100_phase_noise	0.065	0
unet_1024_100_phase_noise	0.011	0
unet_1024_50_phase_noise	0.016	1

A.4. Implementation

The implementation is available within the DLR Gitlab network. Access can be provided on request.