Advancing Satellite Radar Altimetry for Ice Sheet Monitoring using High Resolution Digital Elevation Models and Deep Learning

Joe Phillips, BSc, MSc

Lancaster Environment Centre

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

April, 2026

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university.

This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is: 60603

Joe Phillips

Acknowledgements

First and foremost, I want to thank my supervisor, Mal McMillan, whose support and guidance throughout this journey has been invaluable. You have been caring, capable, and always there when I needed it most. Thank you for your patience with everything I have thrown your way, and I look forward to continuing to work with you.

I am also immensely grateful to Jenny Maddalena for the support you have provided throughout this PhD, and especially for the countless hours you have given me over these final months. Your advice and encouragement kept me going through to the end.

Thank you to all the Lancaster Glaciology Group, who are the warmest and most caring people I have ever worked with. You have all contributed significantly to my decision to stay on, and I am grateful to be part of such a supportive community.

To everyone in office A524 throughout the years, thank you for making work genuinely enjoyable, whether through rusty scooter circuits or lengthy games of chess. You made the difficult days easier.

To Jack and Alice, thank you for welcoming me to Manchester, and especially for the support and friendship you have shown me over the years.

To the Lancaster Bombers American Football Team, thank you for getting me out of the house and providing a much-needed escape from the PhD. Though I could have done without the concussions.

To my friends Joe, Will, and Nathan, thank you for keeping me company most nights, despite the ever-growing distances between us in person. Your presence, even from afar, has been a constant source of support.

To Monty, thank you for keeping me company, especially as hand-in grew nearer.

To my Granny and Grandad, thank you for everything. And to my family - Mum, Dad, Hannah, Benji, Sammy, and Esther - thank you for always being there for me and for contributing non-negligibly to the word count. Your love and support have meant everything.

Advancing Satellite Radar Altimetry for Ice Sheet Monitoring using High Resolution Digital Elevation Models and Deep Learning

Joe Phillips, BSc, MSc.

Lancaster Environment Centre, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy*. April, 2026.

Abstract

Satellite radar altimeters have provided a near-continuous record of polar ice sheets since 1991, delivering new understanding of ice sheet contributions to sea level rise. The complex topography of the ice margins, however, presents several deep-seated challenges; namely, difficulty in tracking the ice surface, and fundamental ambiguity in determining the origin of surface reflections. This thesis develops methods to better understand and overcome these deficiencies, through the use of high-resolution Digital Elevation Models and novel deep learning methodologies. In doing so, it aims to characterise key limitations in conventional processing methods, and demonstrate how new data-driven approaches can better exploit the information encoded within radar waveforms.

Three interconnected objectives were pursued: assessment of current operational systems, development of novel methods for topographic characterisation, and creation of probabilistic frameworks that embrace waveform ambiguity as an information source rather than a processing obstacle.

First, comprehensive evaluation of Sentinel-3 SAR altimetry over Antarctica using high-resolution topographic datasets revealed systematic breakdown of core assumptions in complex terrain. Novel Singular Value Decomposition methods created continent-wide slope and roughness datasets, establishing quantifiable relationships between topography and instrument performance.

Second, a probabilistic deep learning framework was developed for CryoSat-2, using ensemble models trained with quantile regression to predict full elevation

distributions across the altimeter swath from power waveforms alone. The framework demonstrates robust performance whilst explicitly quantifying both fundamental physical ambiguity and model confidence.

Third, the practical application of this new probabilistic approach was investigated, by using the framework to generate swath predictions of ice sheet elevation change. The framework successfully reproduced known elevation change trends, and was benchmarked against a suite of other datasets.

These advances demonstrate that new machine learning methods and high-resolution datasets can address longstanding limitations in SAR altimetry, impacting past, current, and future missions and, ultimately, our understanding of ice sheet change.

Abbreviations and Acronyms

AIS Antarctic Ice Sheet.

AMPLI Altimeter Data Modelling and Processing for Land Ice.

ArcticDEM Arctic Digital Elevation Model.

ATLAS Advanced Topographic Laser Altimeter System.

BC Baseline Collection.

CMIP Coupled Model Intercomparison Project.

CNN Convolutional Neural Network.

COG Centre of Gravity.

CRISTAL Copernicus Polar Ice and Snow Topography Altimeter.

CS2 CryoSat-2.

DEM Digital Elevation Model.

EAIS East Antarctic Ice Sheet.

ENVISAT Environmental Monitoring Satellite.

EOLIS Elevation Over Land Ice from Swath.

ERS European Remote-Sensing Satellite.

ESA European Space Agency.

EU European Union.

GDAL Geospatial Data Abstraction Library.

GIA Glacial Isostatic Adjustment.

GIS Geographic Information System.

GRACE Gravity Recovery and Climate Experiment.

GRASS Geographic Resources Analysis Support System.

GrIS Greenland Ice Sheet.

ICESat-2 Ice, Cloud, and land Elevation Satellite-2.

IMBIE Ice sheet Mass Balance Inter-comparison Exercise.

IPCC Intergovernmental Panel on Climate Change.

iSTAR Investigating the Stability of the West Antarctic Ice Sheet.

LEO Low Earth Orbit.

LEPTA Leading Edge Point Based.

LPC L2-derived POCA Class.

LRM Low Resolution Mode.

LS Least Squares.

MAD Median Absolute Deviation.

mCDW Modified Circumpolar Deep Water.

MEaSUREs Making Earth System Data Records for Use in Research Environments.

MODIS Moderate Resolution Imaging Spectroradiometer.

MPI MultiPeak Ice.

MWR Microwave Radiometer.

NASA National Aeronautics and Space Administration.

NIG Normal-Inverse-Gamma.

NIW Normal-Inverse-Wishart.

OCOG Offset Centre of Gravity.

OLCI Ocean and Land Colour Imager.

OLTC Open-Loop Tracking Command.

OOD Out-of-Distribution.

PICP Prediction Interval Coverage Probability.

PIG Pine Island Glacier.

POCA Point of Closest Approach.

PRF Pulse Repetition Frequency.

QGIS Quantum GIS.

RCP Representative Concentration Pathway.

ReLU Rectified Linear Unit.

REMA Reference Elevation Model of Antarctica.

ResNet Residual Network.

ResNet-D Residual Network - Variant D.

ResNet-RS Residual Network – Rescaled.

RGB Red Green Blue.

RMSprop Root Mean Square Propagation.

RPC REMA-derived POCA Class.

S3 Sentinel-3.

S3NG Sentinel-3 Next Generation.

SAMOSA SAR Altimetry MOde Studies and Applications.

SAR Synthetic Aperture Radar.

SARIn SAR Interferometric.

SE Squeeze-and-Excitation.

SIRAL Synthetic Aperture Radar/Interferometric Radar Altimeter.

SLSTR Sea and Land Surface Temperature Radiometer.

SMB Surface Mass Balance.

SRAL Synthetic Aperture Radar Altimeter.

SVD Singular Value Decomposition.

TFMRA Threshold First Maximum Re-tracker Algorithm.

TPI Topographic Position Index.

TRI Terrain Ruggedness Index.

UF-SAR Unfocused Synthetic Aperture Radar.

USO Ultra-Stable Oscillator.

VGG Visual Geometry Group.

VRM Vector Ruggedness Measure.

WAIS West Antarctic Ice Sheet.

Contents

List of Figures

List of Tables

Chapter 1

Polar Ice Sheets

This chapter provides the context and motivation for this thesis by examining the current state of Earth's ice sheets and the observational techniques used to monitor them. It begins by describing the Antarctic and Greenland Ice Sheets, their contrasting characteristics, and their recent changes in response to climate forcing. The chapter then reviews the primary methods used to measure ice sheet mass balance, discussing the strengths and limitations of each approach and establishing why satellite-based observations have become the preferred framework for contemporary ice sheet monitoring. This foundation establishes the scientific context necessary to understand the technical developments and analyses presented in subsequent chapters.

1.1 Overview

At present, the Earth hosts two continental-scale ice sheets: the Antarctic Ice Sheet (AIS) and the Greenland Ice Sheet (GrIS) (Fig. 1.1). Together, they contain over 68% of the Earth's freshwater reserves, stored as glacial ice, and represent some of the most critical components of the global climate system (Stephens et al., 2020). The AIS extends over approximately 13.66 million km^2 and contains 26.5 million km^3 of ice, equivalent to 58.3 m of potential sea level rise (Fretwell et al., 2013;

IMBIE Team, 2018). The GrIS, though smaller, covers 1.7 million km^2 with 2.9 million km^3 of ice, representing 7.4 m of sea level equivalent (Morlighem et al., 2017; IMBIE Team, 2020).

These ice sheets exhibit fundamentally different structural and environmental settings that shape their distinct responses to external forcing. Antarctica is a continent surrounded by the Southern Ocean, with most of its ice grounded below sea level and extensive fringing ice shelves that buttress ice flow. Its vulnerability is primarily tied to processes at the ice-ocean interface beneath these shelves (Pritchard et al., 2012; Dinniman et al., 2016). Greenland, in contrast, is an island at lower latitudes where summer temperatures regularly exceed the melting point (Hanna et al., 2008; Hanna et al., 2016; Hanna et al., 2021). Surface melt is therefore prominent, with mass loss enhanced by marine-terminating glaciers discharging into narrow fjords (Broeke et al., 2009; Mouginot et al., 2017; Otosaka et al., 2023). This makes Greenland uniquely sensitive to atmospheric warming and associated hydrological processes (Hanna et al., 2008; Slater et al., 2021).

In recent decades, both ice sheets have exhibited increasing signs of instability and accelerated change. Satellite observations show they have been losing mass at accelerating rates since the early 1990s, now contributing roughly one-third of observed global sea level rise (Shepherd et al., 2012; IMBIE Team, 2018; IMBIE Team, 2020; Otosaka et al., 2023). Beyond direct sea level contributions, ice sheets also influence regional and global climate through multiple additional pathways. Their high albedo reflects solar radiation, freshwater discharge perturbs ocean circulation, and their presence modifies atmospheric patterns. Under high-emission scenarios, the combined influence of the Antarctic and Greenland Ice Sheets is expected to dominate sea level rise by the end of the twenty-first century (Slater et al., 2020), with consequences extending well beyond coastal inundation to threaten freshwater supplies, agriculture, and infrastructure in low-lying regions (Fox-Kemper et al., 2021).

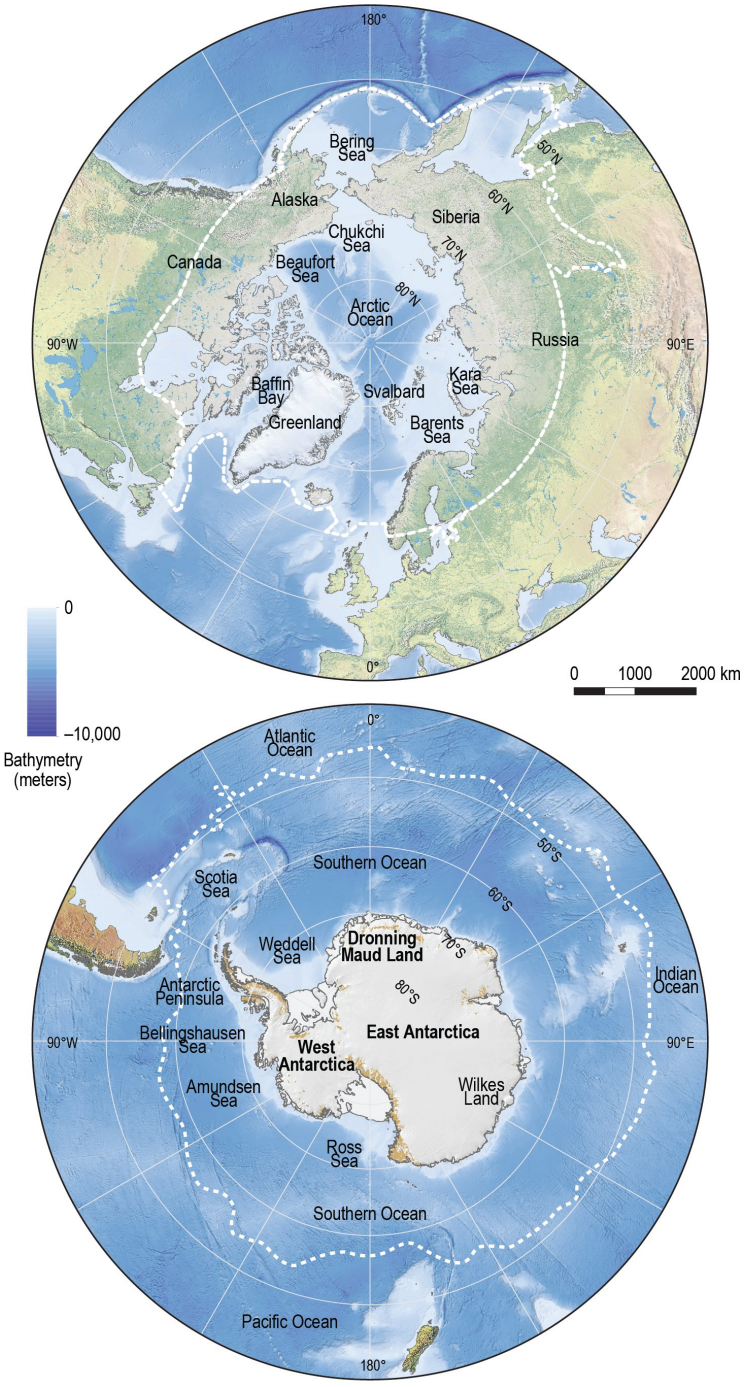

Figure 1.1: The Antarctic Ice Sheet (bottom) and Arctic region including Greenland (top). The Antarctic region encompasses the Antarctic continent and ice sheet, bounded approximately by the Antarctic Circumpolar Current and extending to include subantarctic islands. The Arctic region includes Greenland and its ice sheet, along with surrounding areas characterised by permafrost, glacial ice, and persistent seasonal snow cover. Significant place names are marked, with dashed lines indicating the approximate boundaries of the Arctic and Antarctic polar regions. From Core Writing Team et al. (2014).

1.2 Antarctic Ice Sheet

The Antarctic Ice Sheet can be divided into three distinct sectors with fundamentally different characteristics. East Antarctica (EAIS) contains approximately four-fifths of the continent's ice, grounded largely above sea level on stable continental bedrock. It has remained relatively stable with mass changes of $+3 \pm 15$ Gt yr^{-1} on average between 1992-2020, essentially zero within uncertainties (Otosaka et al., 2023). However, localised coastal regions show vulnerability, particularly the Totten Glacier catchment, which has exhibited sustained mass loss since the 2000s ($\sim 7 \pm 2$ Gt yr^{-1} over 1989-2015) due to warm ocean water intrusion and its basin being grounded largely below sea level, raising concern for marine ice sheet instability (Li et al., 2016).

West Antarctica (WAIS) represents the most vulnerable sector, with ice grounded hundreds of meters below sea level over bedrock that slopes downward inland. This marine-based configuration creates inherently unstable conditions favoring rapid grounding line retreat once initiated (Fretwell et al., 2013). WAIS dominates Antarctic mass balance signals, with mass loss rising from 37 ± 19 Gt yr^{-1} between 1992 to 1996 to a maximum of 131 ± 21 Gt yr^{-1} between 2012 and 2016, before reducing modestly to 94 ± 25 Gt yr^{-1} between 2017 and 2020 (Otosaka et al., 2023). The mass imbalance is driven by acceleration, retreat, and the thinning of principal outlet glaciers (Otosaka et al., 2023). In the Amundsen Sea Embayment, basal melting of the Pine Island and Thwaites ice shelves is particularly strong due to the intrusion of warm modified Circumpolar Deep Water (mCDW) through deep submarine troughs (Nakayama et al., 2019). Variability in the thermocline depth, on-shelf circulation, and topographically steered pathways modulate how much warm water reaches the ice-ocean interface, driving enhanced melt under the ice shelves (Park et al., 2024; Kimura et al., 2017; Boehme et al., 2021). The Antarctic Peninsula, extending northward to about 63°S, has undergone some of the most dramatic recent cryospheric changes, largely through the disintegration of its fringing ice shelves. The abrupt collapses of Larsen A in 1995 and Larsen B

in 2002, followed by the large-scale calving event of Larsen C in 2017, exemplify irreversible ice loss events that immediately accelerated tributary glaciers. These remain ∼26% faster than pre-collapse rates despite subsequent stabilisation (Seehaus et al., 2018). These ice shelf losses removed key buttressing, amplifying glacier discharge, and drove a marked increase in Peninsula ice mass loss after the early 2000s (Otosaka et al., 2023), which attributed to subsequent ice shelf collapse (Rignot et al., 2004; Cook et al., 2010; Adusumilli et al., 2018). However, this longer-term acceleration was briefly interrupted between 2012 and 2016, when average mass loss slowed from $\sim$ 15 Gt yr^{-1} to 6 ± 13 Gt yr^{-1}, largely owing to an extreme snowfall event in 2016 (Wang et al., 2021; Chuter et al., 2022), before rebounding to 21 ± 12 Gt yr^{-1} in 2017-2020 (Otosaka et al., 2023). Between 1992-2020, the AIS lost $2{,}671 \pm 530$ Gt, contributing 7.4 ± 1.5 mm to global sea level rise, with loss rates doubling in the mid-2000s from 62 ± 41 to 130 ± 45 Gt yr^{-1} (Otosaka et al., 2023). The primary driver is an acceleration in ice discharge, particularly in the Amunsen Sea sector (Mouginot et al., 2014; Otosaka et al., 2023). This is driven by ice shelf thinning through basal melting, which reduces buttressing and allows outlet glacier acceleration. Antarctic ice shelf basal meltwater flux varies between $\sim$ $1{,}090 \pm 150$ Gt yr^{-1} to $1{,}570 \pm 140$ Gt yr^{-1} with strong spatial and temporal variability related to surrounding water temperatures (Adusumilli et al., 2020).

Future projections indicate continued Antarctic mass loss under all emission scenarios. Although snowfall (surface mass gain) is expected to increase, dynamic losses from ice discharge, grounding-line retreat, ice-shelf collapse, and ocean-driven basal melting are projected to outpace the gains from increased accumulation, especially under high-warming pathways (Seroussi et al., 2024; Fox-Kemper et al., 2021). The Ice Sheet Model Intercomparison Project for CMIP6 and the Linear Antarctic Response Model Intercomparison Project, project contributions of 0.11-0.12 m by 2100 across emission scenarios, though upper bounds remain poorly constrained due to uncertainties in basal melting sensitivity and marine ice cliff instability mechanisms (Levermann et al., 2020; Edwards et al., 2021). Critical

thresholds emerge on longer timescales: WAIS instability may occur near 1.5°C-2°C warming (Fox-Kemper et al., 2021), with complete WAIS loss projected under 2°C-3°C warming over a millennia (Fox-Kemper et al., 2021), and EAIS vulnerability beyond 3°C potentially contributing 6-12 m total sea level equivalent (Fox-Kemper et al., 2021).

1.3 Greenland Ice Sheet

The Greenland Ice Sheet covers approximately 80% of Greenland's landmass, representing the largest ice mass in the Northern Hemisphere. Its position between Arctic and Atlantic Oceans, and southward extension to ∼60°N, makes it highly sensitive to large-scale atmospheric circulation patterns. In particular, the North Atlantic Oscillation and Greenland Blocking Index, which strongly modulate regional summer temperature, precipitation, and melt conditions over the ice sheet (Hanna et al., 2016; Hahn et al., 2018).

The ice sheet comprises two glacier types which respond through different mechanisms: land-terminating glaciers and marine-terminating outlet glaciers. Land-terminating glaciers lose mass primarily through surface mass balance (SMB) changes, which is the difference between accumulation and ablation. These glaciers respond directly to atmospheric warming, where rising air temperatures drive substantially increased surface melting and runoff (Fettweis et al., 2020; Slater et al., 2021; Tedesco et al., 2020). At the same time, the firn layer's capacity to retain and refreeze meltwater weakens under warmer scenarios (Noël et al., 2022). The combination of these effects increases the area, number, and inland extent of supraglacial hydrology features (Leeson et al., 2014; Williamson et al., 2018; Glen et al., 2025). Furthermore, the system exhibits self-reinforcing feedbacks: melt-albedo feedback where darker bare ice absorbs more radiation than snow (Sellers, 1969; Oerlemans et al., 1989), and melt-elevation feedback where surface lowering exposes ice to warmer temperatures (Weertman, 1961; Oerlemans, 1982; Levermann

et al., 2016). Marine-terminating glaciers experience the same surface processes as land-terminating glaciers but they are additionally subject to dynamic mass loss through ice-ocean interactions (Rignot et al., 2011). Submarine melting at glacier fronts is a key driver of retreat and thinning, particularly where warm Atlantic Water reach fjord termini (Straneo et al., 2010; Motyka et al., 2011; Straneo et al., 2013). These melt rates are further enhanced when subglacial discharge plumes from surface meltwater rise along the calving front, mixing fjord waters and warm Atlantic layers (Jackson et al., 2014; Chauché et al., 2014; Everett et al., 2021). Major Greenland outlets such as Jakobshavn Isbræ have undergone rapid changes since the 1990s, including the loss of their floating ice tongues and dramatic acceleration (Joughin et al., 2004; Motyka et al., 2011), while others like Helheim and Kangerdlugssuaq have shown similar dynamic responses tied to ocean forcing (Straneo et al., 2010; Joughin et al., 2010). However, the magnitude and timing of retreat and thinning vary considerably between drainage basins, reflecting local controls such as fjord geometry, bed topography, and the extent to which warm Atlantic Water penetrates to the ice front (Carr et al., 2013; Straneo et al., 2013; Enderlin et al., 2014).

The Greenland Ice Sheet has experienced a cumulative loss of 4,892 ± 457 Gt of ice between 1992 and 2020, contributing approximately 13.5 mm to global sea level rise (Otosaka et al., 2023). The average annual mass loss during this period was 169 ± 16 Gt yr^{-1} (Otosaka et al., 2023). However, the rate of ice loss has shown significant inter-annual variability, ranging from 86 ± 75 Gt yr^{-1} in 2017 to a peak of 444 ± 93 Gt yr^{-1} in 2019, driven by exceptional surface melting (Tedesco et al., 2020). Mass loss initially occurred through equal SMB and dynamic contributions (1992-2018), but recent observations show SMB losses now dominate total loss (Broeke et al., 2016; Slater et al., 2021; Otosaka et al., 2023). Surface warming has been pronounced, with coastal regions experiencing +1.7°C summer and +4.4°C winter warming between 1991-2019 (Hanna et al., 2020), driving increased meltwater runoff that averaged 357 ± 58 Gt/yr in 2011-2020, representing a 21% increase over the preceding three decades (Slater et al., 2021).

Under all emission scenarios, the Greenland Ice Sheet is projected to continue losing mass due to both surface and ocean-driven processes. By 2100, projected contributions to global sea level rise from mass loss are estimated at 30 (1-70) mm under a low-emission pathway (RCP2.6, radiative forcing stabilising at 2.6 W/m^2) and 70 (30-160) mm under a high-emission pathway (RCP8.5, radiative forcing reaching 8.5 W/m^2). Under continued high emissions (RCP8.5), cumulative contributions could increase to 310-1740 mm by 2200 (Fox-Kemper et al., 2021). Surface mass loss will continue to accelerate as warming expands meltwater features inland and degrades the firn layer, which currently buffers approximately 50% of surface meltwater through storage and refreezing (Thompson-Munson et al., 2024). Warming ocean waters will drive enhanced submarine melting at marine-terminating glacier fronts, promoting dynamic ice loss through frontal retreat and calving (Nias et al., 2023).

1.4 Measuring Mass Loss

The vulnerabilities of Antarctic and Greenlandic ice sheets - oceanic and atmospheric sensitivity respectively - reflect their heterogeneous responses to climate change, with regional variability in processes and rates driven by the interplay of atmospheric forcing, ocean circulation, and local boundary conditions. Ice sheet mass balance monitoring is vital for quantifying sea level contributions and understanding process drivers, as freshwater input from ice loss alters ocean salinity and circulation systems like the Atlantic Meridional Overturning Circulation, potentially affecting global weather patterns, monsoons, and storm tracks (Wouters et al., 2022; Frajka-Williams et al., 2023; Andernach et al., 2025). This spatial and temporal complexity reinforces the need for sustained monitoring capable of resolving basin-by-basin changes in both surface and dynamic processes.

Given the central role of ice sheets in the climate system, accurately measuring their mass balance is a scientific priority. To address this need, several approaches

are used to monitor ice sheets, ranging from in situ measurements to satellite-based observations. In situ methods involve direct field measurements of accumulation, melt, and ice thickness at specific locations using techniques such as stake networks, ice cores, and ground-penetrating radar. While these provide high-precision local observations and detailed process understanding, they are limited by logistical challenges, high costs, and sparse spatial coverage across vast ice sheet regions (Lenaerts et al., 2019). Airborne surveys offer improved spatial coverage through techniques like laser altimetry and ice-penetrating radar, enabling detailed measurements along flight lines and filling gaps between ground-based observations (Zemp et al., 2013). However, airborne campaigns are expensive, weather-dependent, and typically provide only episodic rather than continuous monitoring.

Satellite-based approaches have become the primary method for ice sheet monitoring due to their ability to provide comprehensive, spatially continuous coverage of entire ice sheets at regular intervals. Unlike point-based in situ measurements or linear airborne surveys, satellites can systematically observe the complete spatial extent of ice sheets, capturing both large-scale trends and regional variability that would be impossible to resolve through ground-based networks alone. This comprehensive coverage is essential for understanding ice sheet behavior, as mass changes are highly heterogeneous and processes can have significant implications for overall mass balance (Otosaka et al., 2023).

Contemporary ice sheet mass balance assessment leverages these satellite capabilities through three main approaches, each with distinct strengths and limitations. Satellite gravimetry provides a direct measure of total mass change by recording variations in Earth's gravity field. The Gravity Recovery and Climate Experiment (GRACE) and its successor GRACE Follow-On have provided monthly resolved records of ice sheet mass loss since 2002, with measurement uncertainties typically less than 10 Gt/month for ice sheet-wide estimates (Chen et al., 2006; Luthcke et al., 2006; Velicogna et al., 2014; Velicogna et al., 2020). However, the

∼300 km spatial resolution of these measurements limits their utility for studying regional processes or attributing causes to specific mechanisms (Loomis et al., 2021; Schlegel et al., 2016). Additionally, gravimetric observations record the combined effect of recent ice sheet mass changes and glacial isostatic adjustment (GIA) - the ongoing solid Earth response to ice mass changes since the last glaciation (Velicogna et al., 2013). Since GIA signals are of the same order of magnitude as contemporary ice sheet mass balance signals, accurate GIA models are essential for isolating the ice sheet mass change component.

Satellite altimetry provides an alternative approach by measuring changes in surface elevation over time. Converting elevation changes to mass changes requires information about ice and firn density to account for the volume-to-mass relationship, along with corrections for firn compaction processes and, to a lesser extent, GIA-related vertical land motion (McMillan et al., 2016; IMBIE Team, 2018). Since the launch of ERS-1 in 1991, radar and laser altimetry have offered multi-decadal records of surface change, forming a cornerstone of contemporary ice sheet assessments (Zwally, 1989; Zwally et al., 2005; Zwally et al., 2011; Hurkmans et al., 2014; McMillan et al., 2016; IMBIE Team, 2018; Sandberg Sørensen et al., 2018; Smith et al., 2023). Altimetry is particularly valuable because it resolves the spatial heterogeneity of change across basins and sectors, enabling direct links between observations and physical processes such as ice shelf thinning (Gudmundsson et al., 2019) and grounding line retreat (Li et al., 2023). However, measurements are sensitive to surface topography effects (Bamber, 1994; Simonsen et al., 2017), radar signal penetration depth (Nilsson et al., 2015; Nilsson et al., 2016; Slater et al., 2019), and uncertainties in density assumptions for volume-to-mass conversion (McMillan et al., 2016; IMBIE Team, 2018).

The input-output method, or mass budget method, represents a third approach that calculates mass balance as the net difference between mass gain (through surface mass balance) and mass loss (through solid ice discharge at the grounding line, basal melting, and ice-ocean interface melting) (Rignot et al., 2011; Broeke et al., 2009).

Surface mass balance represents precipitation minus meltwater runoff, sublimation, evaporation, and erosion, typically derived from atmospheric reanalyses and regional climate models rather than direct observations. Ice discharge is calculated using satellite-derived ice velocities and ice thickness measurements at flux gates along the grounding line. This method provides valuable process-level insights by separating the climatic (surface) and dynamic (discharge) components of mass change, enabling attribution of losses to specific mechanisms (Broeke et al., 2009; Mouginot et al., 2019). However, it requires detailed observations of ice velocity and thickness along ice margins (Mouginot et al., 2019) and relies heavily on the accuracy of surface mass balance models (Van den Broeke et al., 2011; Lenaerts et al., 2019).

This thesis focuses on the altimetric approach. By examining the capabilities and limitations of radar altimetry over ice sheets, it aims to evaluate current retrieval performance and contribute to the development of improved measurement frameworks. The following section therefore introduces the theoretical background of radar altimetry and its application to polar ice sheet monitoring.

Chapter 2

Satellite Radar Altimetry

This chapter presents the theoretical and technical foundations of satellite radar altimetry with respect to its application to the cryosphere. It first reviews the physical principles that govern the acquisition and processing of radar echoes over ice sheet surfaces, before outlining recent instrumental developments that have expanded the capability of altimeters over ice sheets. It then describes conventional, non-interferometric processing pipelines, before introducing the satellite missions central to this thesis, which together provide the observational basis for the analyses presented in the following chapters.

2.1 Overview

Whilst a range of techniques exist for remote observation of the cryosphere, our understanding of how ice sheets are changing is largely informed by satellite observations, with the longest continuous record coming from the technique of satellite radar altimetry. The principal use of altimetry over ice sheets is to derive estimates of ice sheet elevation and elevation change (Shepherd et al., 2019; Wingham et al., 1998), which, ultimately, can be used to determine ice sheet mass imbalance. Radar altimeters have also been used to investigate a range of other glaciological processes, including grounding line location and migration (Dawson

et al., 2017; Hogg et al., 2018; Konrad et al., 2018), subglacial hydrology (Gourmelen et al., 2017; McMillan et al., 2013; Siegfried et al., 2018; Wingham et al., 2006b), ice shelf processes (Chuter et al., 2015; Griggs et al., 2011), and surface mass balance (Slater et al., 2021; IMBIE Team, 2018; IMBIE Team, 2020; Otosaka et al., 2023).

The era of systematic polar monitoring by satellite radar altimeters began in the early 1990s with the launch of ERS-1 on July 17$^{\text{th}}$ 1991, followed shortly by ERS-2 on 21$^{\text{st}}$ April 1995. Although designed primarily for oceanographic applications, these missions provided the first continuous records of ice sheet elevation, establishing the foundations for long-term altimetric monitoring. Over the following decades, successive missions progressively improved the resolution and reliability of measurements, leading to today's operational systems that deliver near-continuous polar coverage. This historical progression reflects a clear trajectory toward increasingly sophisticated observations of the cryosphere, supported both by advances in instrument design and by the growing emphasis on sustained polar monitoring.

2.2 Fundamentals of Satellite Radar Altimetry

2.2.1 Principles

Radar altimeters work by transmitting a short radio-wave pulse towards Earth's surface and recording the returned echo in the form of a discretised waveform (Figure 2.1). Each recorded waveform, comprising the sum of incident surface reflections from within the altimeter footprint ordered by arrival time (Brown, 1977), encodes information pertaining to the surface illuminated by the satellite, such as topography, electromagnetic scattering characteristics, and in the case of returns over ocean, wind speed and significant wave height (Quartly et al., 2019).

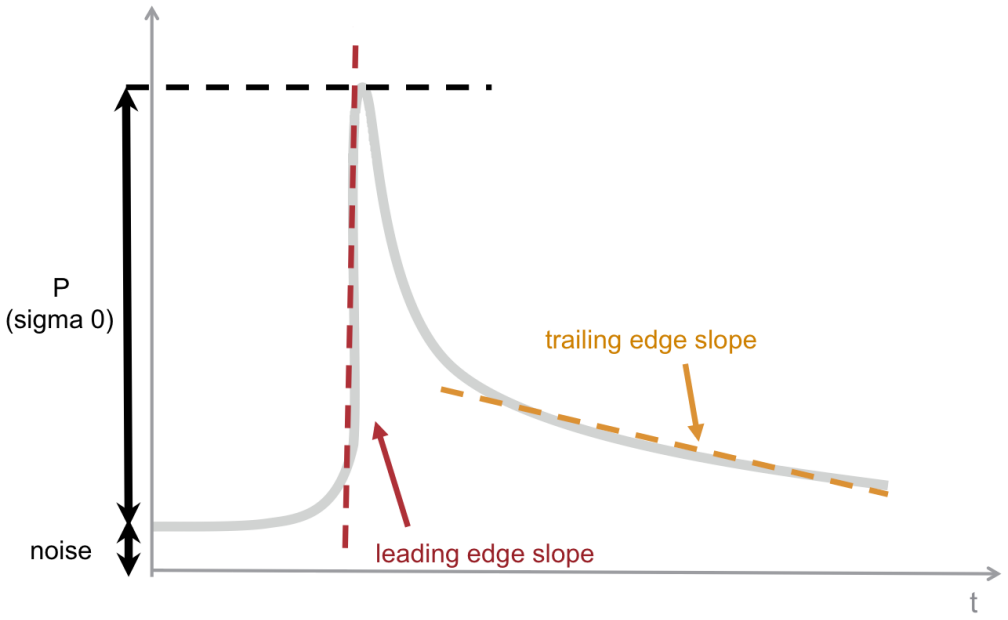

Figure 2.1: The typical geometric characteristics of an altimeter waveform over a uniform surface. Adapted from isardSAT (n.d.), via EUMETSAT.

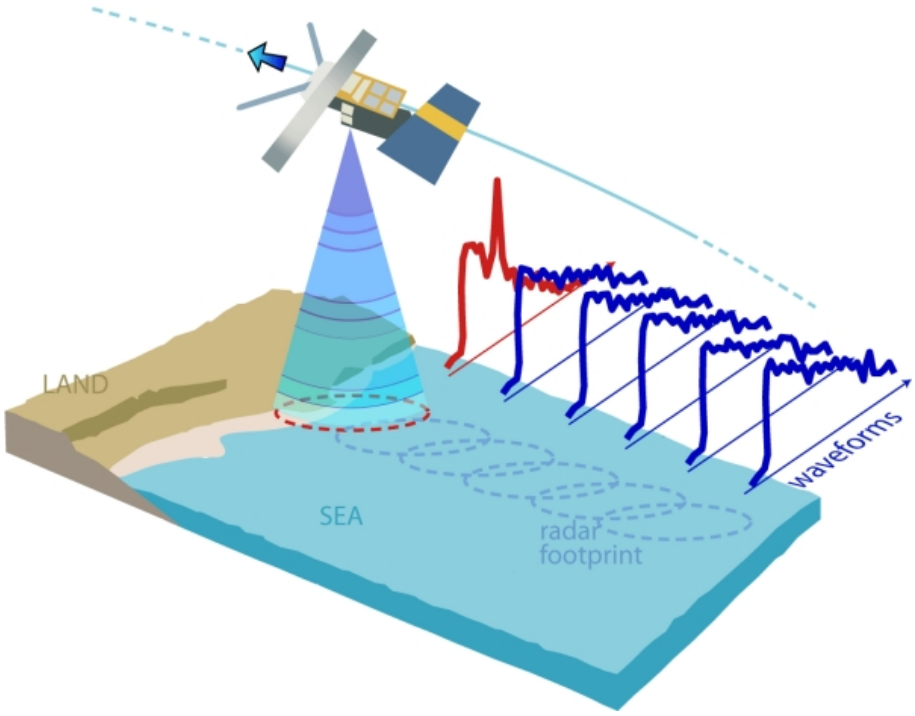

Figure 2.2: Diagram of a satellite altimeter approaching the coastline, outlining the illuminated footprint for each pulse and the resulting captured waveform. From Noor (2018).

Typically, the altimeter antenna footprint width on the surface is defined as the point at which the beam's intensity drops to half-power (-3 dB), which is referred to as the beam-limited footprint. For conventional low-resolution instruments, such as those onboard the ERS-1, ERS-2 and Envisat satellites, this circular footprint is typically tens of kilometres wide, and is determined by the satellite's altitude, attitude, and the shape of the antenna gain pattern, which describes the variation of transmitted power with angle away from the boresight direction. However, in practice, achieving beam-limited footprints from orbit would require an antenna several metres across (e.g. $\sim$5 m at Ku-band to obtain a 5 km footprint) and would consequently be highly vulnerable to mispointing errors - small deviations in the antenna pointing direction that would cause the narrow beam to miss the target surface entirely. Therefore, while the beam-limited footprint defines the full area illuminated by the antenna, both early and current spaceborne altimeters operate in pulse-limited mode, in which the radar pulse length defines a finer resolution cell within that illuminated area. This pulse-limited footprint - typically 3-5 km in diameter depending on surface roughness and slope - corresponds at any given instant to a concentric annulus of the surface, such that each range bin in the returned waveform can be attributed to a specific portion of the surface rather than to the beam-limited footprint as a whole (Figure 2.3). This bin-to-annulus correspondence also helps makes waveform signal attribution to the surface tractable, which would otherwise be more ambiguous across the broad beam-limited footprint. Flat surfaces yield narrower annuli and sharper pulse-limited footprints than rough or sloping terrain.

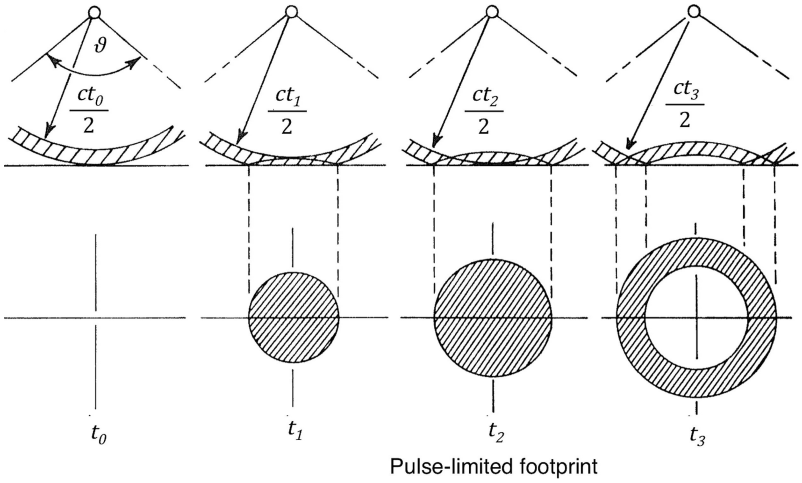

Figure 2.3: An illustration of a pulse-limited footprint through time, for an antenna with beam width ϑ, at times t_0-t_3. The one-way distance to the surface is given by $\frac{c}{t_n}$, where c is the speed of light, and t the time elapsed since emission. From Tilling et al. (2018) and Rapley et al. (1983).

Over uniform surfaces, altimeter waveforms take a distinctive shape, with a clear peak in power, corresponding to the first return from the surface at the point of closest approach (POCA) to the satellite, followed by a gradual decay in the received power - due to the combined influence of the antenna gain pattern and the orientation of the surface relative to the incident radar wave (Figure 2.1). However, for echoes acquired over ice sheet surfaces - which are typically non-uniform, rough, and irregular - waveform shape can be more complex, particularly in regions close to the ice margins. This complexity arises from the presence of numerous discrete scattering regions within the illuminated beam-limited footprint, each contributing differently to the overall returned signal.

In order to capture the backscattered radar pulse that forms each waveform, the altimeter records the incoming energy for a discrete amount of time, corresponding to when it expects the reflection to be received from the surface. Radar altimeters can have a recording window as short as several microseconds, and when this fails to coincide with the surface return then the instrument temporarily loses track of Earth's surface (Figure 2.4; further details in Section 2.2.2).

To illustrate satellite radar altimetry behavior more intuitively, I developed an

interactive simulator (`https://github.com/Joe-Phillips/Satellite-Radar-Altimetry-Tool/`) that allows users to input 2D surface topographies and observe how radar pulses interact with the surface to form corresponding waveforms. An updated version of this simulator is applied in Chapter 6.

2.2.2 Practical Application

From each recorded echo, an estimate of surface elevation can be derived by measuring the range between the satellite and a fixed point on the leading edge of the waveform, which is assumed (in the case of conventional altimetry) to correspond to the POCA return. This range measurement can then be converted to an estimate of the surface elevation at the POCA, given knowledge of the satellite altitude. Additionally, several other useful descriptive parameters can also be derived from the waveform (Figure 2.1), including estimates of (1) the ice sheet's backscatter coefficient (sigma-0), which can be used to characterise ice sheet properties such as snow depth, grain size, and density (Blarel et al., 2015; Mertikas et al., 2020); (2) the leading edge width, which contains information related to surface roughness and radar wave penetration into the near-surface snow layer; and (3) the trailing edge slope, which is sensitive to both footprint scale topography and subsurface backscattering (Legresy et al., 2005; Rémy et al., 2009)

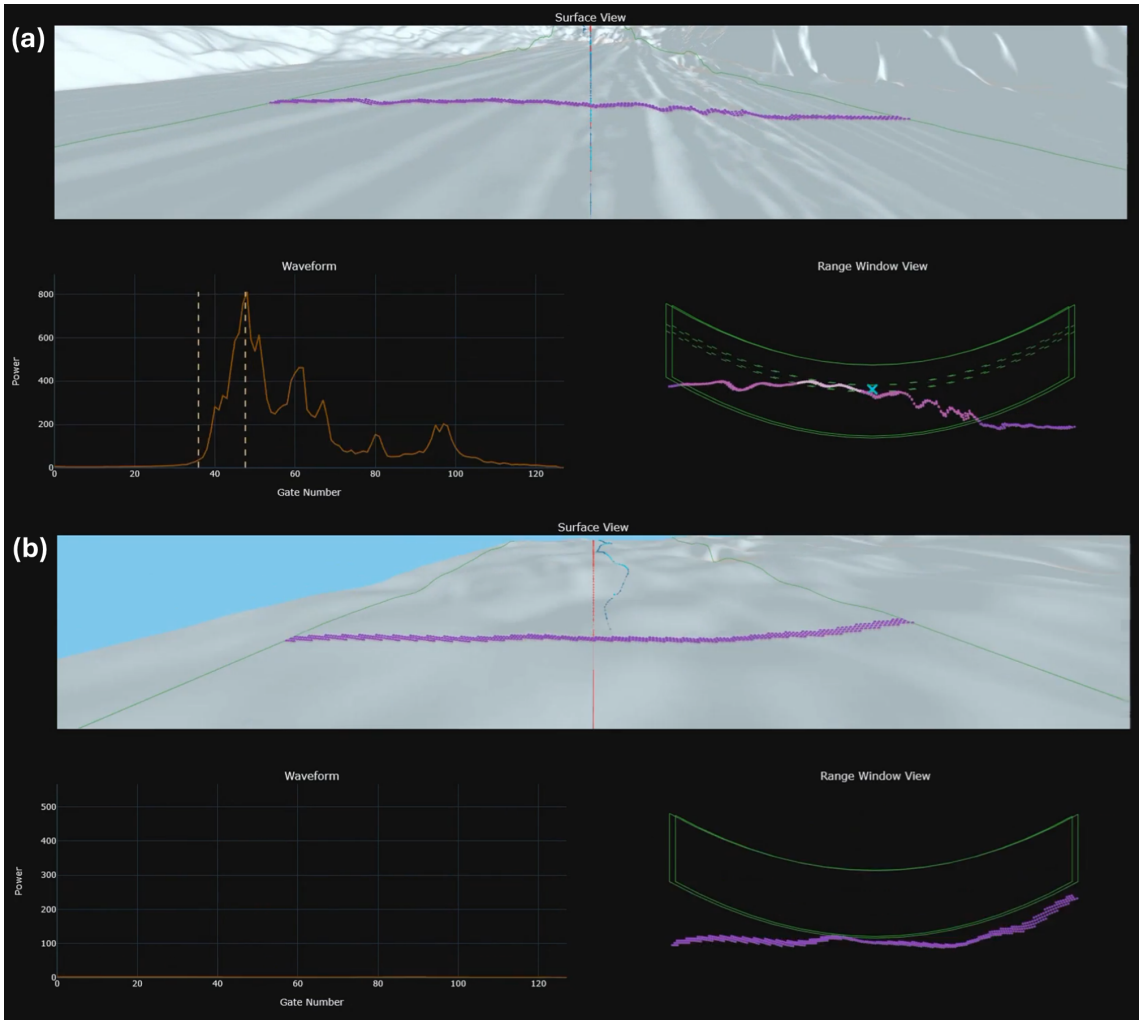

Figure 2.4: Visualisation of Sentinel-3 tracker performance over the margin of the Antarctic Ice Sheet. The figures show frames from a custom tool I developed that interactively plots and animates Sentinel-3 and CryoSat-2 altimetry data (`https://github.com/Joe-Phillips/SAR-Altimetry-Flyover-Visualiser`). Example shown uses Sentinel-3 (BC-005 land ice product) over Mertz Glacier in Antarctica, January 2020, which operated in closed-loop tracking mode. The surface view shows the REMA DEM (100 m). Green outlines mark the -3 dB footprint, orange the nadir, blue the POCA, and purple the REMA points within the illuminated footprint. The waveform (orange) leading edge is given by bounding dashed light orange lines. The range window view shows the range window in green (viewed into the page, along-track), dashed green as the leading edge (in space). REMA points are shaded to show their progressive capture, first within the range window and then within the leading edge. The footprint appears non-circular due to SAR processing (Section 2.2.3.1). Panel (a) shows successful surface tracking; panel (b) shows tracking failure.

To place the range window correctly, and so capture returns effectively, an altimeter requires an estimate of the predicted range to the approaching surface. To provide such information, two methods are utilised: "closed-loop" and "open-loop" tracking. Closed-loop tracking predicts the next range based on the history of the last few seconds. Whilst this works well when the range has linear variation, such as over the ocean or the low slope interiors of the ice sheets, its performance degrades over more highly variable terrain, because the range history is a less reliable predictor of the future range evolution. In contrast, open-loop tracking uses information from an a priori Digital Elevation Model (DEM) to provide the satellite with an estimate of the expected range to the ice surface (Donlon et al., 2012a). This method has the potential to more precisely track the complex topography found across many parts of the ice sheet margin.

2.2.3 Advances in Satellite Radar Altimetry Instrumentation

All conventional radar altimeters operating prior to 2010 functioned in Low Resolution Mode (LRM). In this mode, pulses are transmitted at a relatively low pulse repetition frequency (PRF), such that the backscattered signal has returned to the satellite prior to the next radar pulse being sent. Reflected echoes are then averaged onboard to suppress speckle noise, which further reduces the effective sampling rate. While effective over open oceans, LRM has significant drawbacks when applied to the cryosphere (McMillan et al., 2018). The large, pulse-limited footprint - spanning several kilometres in diameter - results in relatively poor spatial resolution and a limited capability to resolve small-scale variations in ice sheet topography. Moreover, LRM provides no information about where within the beam footprint the echo originates, meaning that reflections from different locations within the footprint cannot be distinguished. Over regions with complex surface geometry, such as steep slopes, crevassed areas, or mountain glaciers the resulting measurements can be highly contaminated or entirely unusable. These limitations motivated the development of new instruments and processing strategies capable

of narrowing the effective footprint and improving geophysical retrievals over ice-covered surfaces. To address these issues, CryoSat-2 (CS2), which was launched in 2010, introduced two key innovations in instrument design: (1) the use of synthetic aperture radar (SAR) processing, and (2) radar interferometry (SARIn when used alongside SAR processing) (Wingham et al., 2006a). The following sections provide a brief summary of each innovation.

2.2.3.1 Synthetic Aperture Radar

In conventional LRM operation, the altimeter must wait for each radar pulse to complete its round trip before transmitting the next. This constraint enforces a low PRF and limits the achievable along-track resolution to the kilometre scale, meaning that fine-scale variations in topography cannot be resolved. Synthetic Aperture Radar (SAR; also referred to as Delay-Doppler) altimetry was developed to overcome this limitation by transmitting bursts of pulses at much higher PRF, allowing overlapping emissions and enabling higher along-track resolution.

In essence, SAR altimetry processing uses the Doppler effect to discriminate multiple overlapping pulses based on their frequency shift caused by the motion of the satellite relative to the target surface, thus enhancing along-track resolution (Figure 2.5). To enable this, the instrument transmits bursts of 64 closely spaced pulses at a higher PRF, increased from ∼2 kHz in conventional low-resolution mode altimetry to ∼18 kHz for SAR altimetry. This allows a fast Fourier transform to partition the beam-limited footprint into 64 along-track "Doppler strips," enabling CryoSat-2 to achieve a four-fold improvement in along-track resolution (∼380 m) compared to earlier missions. Additionally, multiple looks of the same Doppler strip can be "stacked", thereby reducing radar speckle - a process known as multi-looking. As such, instead of a large circular footprint at the surface, SAR processing allows CryoSat-2 to sample the surface in ∼15 km across-track by ∼380 m along-track strips (∼300 m for Sentinel-3), where the across-track resolution is pulse-limited and dependent on surface roughness and slope, and the along-track resolution

is Doppler-limited (Figure 2.5). While the concept of SAR processing has been applied within the context of radar imagers since the 1950s (Love, 1985) and has seen extensive application in previous satellite missions (Curlander et al., 1991; Kovaly, 1976; Rosen et al., 2000), SAR processing has only recently been utilised by Earth-orbiting altimeters with the launch of CryoSat-2, and has since been applied to the Sentinel-3 mission. This innovation has represented an important advance in polar Earth Observation, allowing smaller-scale features across ice sheets and sea ice to be resolved (Raney, 1998; Tilling et al., 2015).

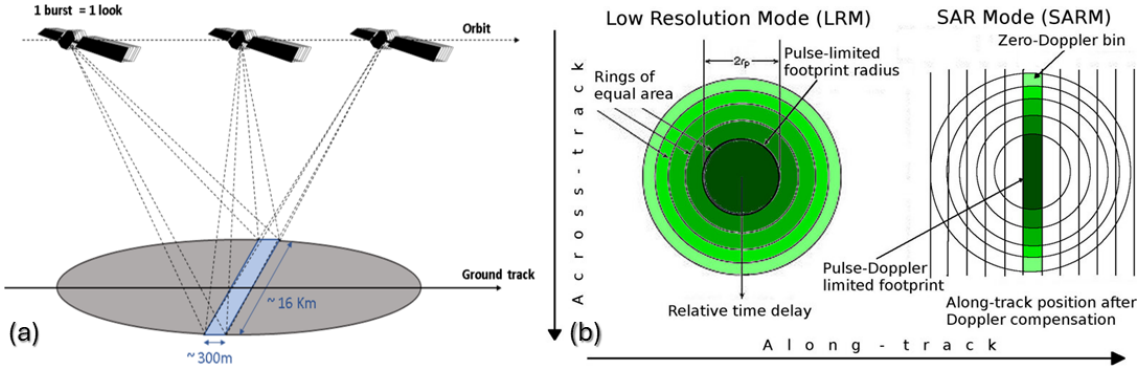

Figure 2.5: (a) Schematic of the SAR altimetry process, from AVISO (2019) and Raney (2021). (b) Top-down illustration of footprint differences between LRM and SAR processing, from Martins et al. (2023) and Raney (1998).

2.2.3.2 Interferometry

Traditional radar altimeters equipped with a single antenna are inherently limited to one-dimensional measurements: they record only the time delay of the returning echoes. This provides information about the range to reflecting surfaces but not their position within the antenna's beam footprint. As a result, the geographic origin of each reflection is ambiguous, lying anywhere along an arc defined by the antenna gain pattern and the echo return time. While SAR processing improves along-track resolution by distinguishing echoes based on their Doppler frequency shifts, it cannot resolve the across-track ambiguity - reflections from different across-track locations within the footprint remain indistinguishable. Over relatively flat

surfaces such as the open ocean, this remaining ambiguity is less problematic, but in regions with complex surface geometry - such as ice sheet margins, and outlet glaciers - the inability to distinguish across-track echo location significantly degrades measurement accuracy. Consequently, single-antenna altimetry cannot reliably resolve fine-scale topography or identify the true source of off-nadir reflections.

Interferometric techniques address this limitation by recording the phase difference between two spatially separated antennas, enabling the angle of echo arrival to be determined. This additional dimension of information allows interferometric altimeters to geolocate surface echoes more precisely, and also provides the capacity to generate detailed elevation profiles consisting of numerous georeferenced points across-track (so called "swath processing"). Compared to conventional POCA measurements, this process has been found to provide up to a tenfold increase in spatial sampling, and at 70° latitude, a three, five, ten and 35-fold increase in temporal sampling at 5, 3, 1 and 0.5 km posting, respectively (Gourmelen et al., 2018). The SARIn capabilities onboard CS2 have demonstrated significant improvements when compared to LRM, especially over complex surfaces (Gray et al., 2017; Wang et al., 2019; Aublanc et al., 2021; Wang et al., 2015; McMillan et al., 2018; Chuter et al., 2015). In addition to improved ice sheet mass balance estimates, interferometry has also benefited other applications including grounding line detection and retreat monitoring (Dawson et al., 2017; Hogg et al., 2018), as well as the resolution of small water bodies such as supraglacial lakes (Gray et al., 2017; Ignéczi et al., 2016).

2.3 Conventional Non-Interferometric Level-2 Processing Pipelines

The discussion so far has outlined the principles of satellite radar altimetry and the major instrumental advances that have enabled improved measurements over the cryosphere. Yet, as introduced in the previous section, conventional non-

interferometric waveforms (both LRM and SAR) remain fundamentally limited by range ambiguity: each waveform can, in principle, correspond to a wide variety of possible topographic surfaces within the satellite footprint. Without additional information, it is therefore impossible to unambiguously determine the true origin of reflections.

To address this, current ice sheet SAR and LRM altimetry Level-2 processing chains - which have followed broadly the same logic for the past three decades - reduce the dimensionality of the problem through two key steps: (1) retracking and (2) slope correction. Together, these aim to reduce the dimensionality of the problem such that a singular measurement of elevation, placed at the correct location on the ice sheet surface, is obtained. Specifically, the assumption is made that the leading edge of a captured echo - the initial steep rise of the pulse from the noise floor to the inflection point at peak power - corresponds to the first surface return at the point of closest approach (POCA) to the satellite. Using this assumption, retracking works to first obtain a single range measurement from the leading edge, corresponding to the distance between the satellite and the surface, with the slope correction step then used to identify the POCA location on the surface to attribute this to.

These steps, together with the application of a range of geophysical and instrumental corrections, form the conventional Level-2 pipeline that transforms radar waveforms into estimates of surface elevation. The following sections describe each stage in turn.

2.3.1 Retracking

During Level-2 (L2) processing, the position of the waveform leading edge relative to a specific onboard tracking gate is estimated. This process is known as "retracking", and can involve adjusting the tracker range by several meters (Martin et al., 1983; Ridley et al., 1988). In general, retracking algorithms can be categorised as "physical" or "empirical". Physical retrackers fit an analytical model to the received waveform in order to represent the underlying physics of the radar wave's interaction

with the scattering surface, whereas empirical retrackers solely consider the geometry of the recorded waveform (Passaro et al., 2022; Villadsen et al., 2016). In practice, physical retrackers are rarely used over ice sheets, because of their inability to fit irregular waveform shapes caused by complex terrain, and their sensitivity to changes in the scattering properties of the surface (Landy et al., 2019; Martin et al., 1983; Quartly et al., 2019; Quartly et al., 2020; Slater et al., 2019). Conversely, empirical retrackers do not assume waveform shape, and as such, are far more robust to cases where the echo deviates from its theoretical shape. Examples of physical and semi-analytical retrackers include ICE-2 (Legresy et al., 2005) which was applied to low resolution mode ENVISAT data, SAMOSA (Ray et al., 2015), which has been applied to delay-doppler Cryosat-2 and Sentinel-3 data, and the UCL ice sheet retracker (Wingham et al., 2006b), which is deployed in the Sentinel-3 land ice ground segment. Examples of empirical retrackers include ICE-1, which applies a threshold based upon the OCOG (Offset Centre of Gravity) amplitude (Bamber, 1994; Wingham et al., 1986), and has been used widely since its implementation within the ground segment of ERS-1; the threshold retracker developed by (Davis, 1997); TFMRA (Threshold First Maximum Re-tracker Algorithm) (Helm et al., 2014); and MultiPeak Ice (MPI), a recent approach designed to handle complex, multi-peaked waveforms (Huang et al., 2024).

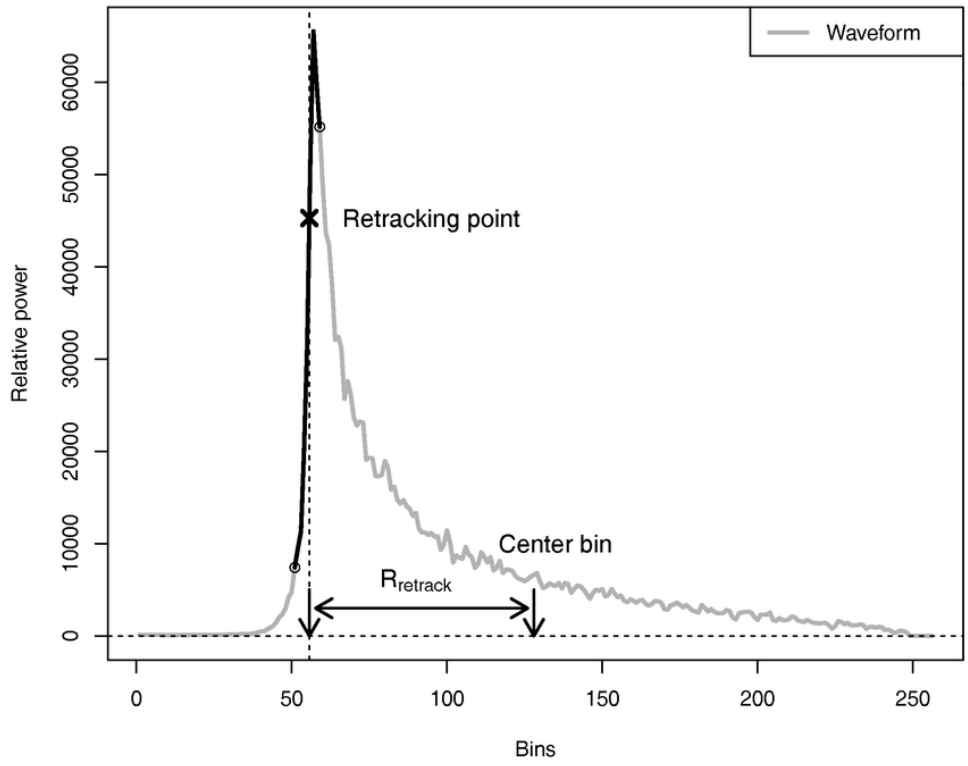

Figure 2.6: Illustration of the retracking process, with the known, onboard tracking gate given at the center of the range window. The leading edge is highlighted in black, and the retracked distance R_{retrack} is defined as the offset between the center bin and the algorithm-selected point on the leading edge. From Nielsen et al. (2017).

2.3.2 Slope Correction

As the measured range to the POCA is 1-dimensional, it is insufficient to determine the location of POCA in 3-dimensional space; rather it constrains POCA to lie upon an iso-range surface, a constant distance from the satellite. Over flat surfaces such as the open ocean, the POCA can be assumed to be at nadir directly beneath the satellite. However, over sloped and rough ice sheet topography, this assumption cannot be made and the ambiguity must be resolved. Solving for this ambiguity - to determine the location of POCA - is a process commonly referred to as slope correction. Specifically, the aim is to identify the origin on the surface corresponding to the measured elevation, based upon the assumption that the point closest to

the satellite corresponds to the backscattered energy from the leading edge of the waveform. Over rugged terrain, the magnitude of this slope correction can reach tens of meters vertically, and several kilometers across-track. As such, uncertainties in applying this correction are commonly the largest source of uncertainty in non-interferometrically derived elevation measurements (Brenner et al., 2007; Li et al., 2022).

Historically, to identify the POCA location, "slope-based" methods were used (Bamber, 1994; Brenner et al., 1983; Cooper, 1989; Remy et al., 1989). These assumed a constant slope within the beam-limited altimeter footprint, which was commonly derived using a priori slope data at nadir (Levinsen et al., 2016; Remy et al., 1989). By assuming an orthogonal reflection from the surface, the approximate location of the POCA can then be determined using trigonometry (Brenner et al., 1983; Roemer et al., 2007). Whilst the assumption of constant slope is reasonable over simpler, homogeneous terrain, the slope method neglects the finer-scale topography within the beam-limited footprint, which may lead to inaccuracies over even moderately undulating areas of the ice sheet (Levinsen et al., 2016). To address this, "point-based" methods were developed which locate the POCA by using an auxiliary DEM to search for points with minimum range within the beam-limited satellite footprint (Levinsen et al., 2016; Li et al., 2022; Roemer et al., 2007). Such an approach was first introduced by Roemer et al. (2007) using a DEM derived from ERS-1 measurements. Specifically, this approach moves a fixed window the approximate size of the satellite's pulse-limited footprint through DEM data within the beam-limited footprint. The POCA location is then identified as the centroid of the window that minimises the mean range between the satellite and the DEM points within that window, with interpolation between DEM nodes used to further improve accuracy and precision. In their initial study, Roemer et al. (2007) found that this new approach outperformed slope-based methods over Lake Vostok within the East Antarctic interior, reducing the error standard deviation from 1.1 m to $\sim$0.5 m when compared to ICESat-1 measurements. This was subsequently

confirmed over Greenland, both for Envisat (Levinsen et al., 2016) and CryoSat-2 Low Resolution Mode data (Li et al., 2022). In both cases the dispersion of the elevation differences relative to either airborne or satellite laser altimetry was 2-6 times greater for a slope-based approach in comparison to a point-based approach.

More recently, LEPTA (Leading Edge Point Based), a formulation of a point-based method was proposed by Li et al. (2022), which utilised the ranges spanned by the waveform leading edge to identify the DEM points corresponding to the POCA (Figure 2.7). The POCA, and consequently the slope correction, was then derived from the mean of the points intersecting the leading edge. This approach has also been generalised to allow retrieval of multiple elevations from a single multi-peaked waveform, which was demonstrated for Sentinel-3 SAR acquisitions (Huang et al., 2024). When benchmarked against ICESat-2, LEPTA achieved a median height difference of <0.01 m and a median absolute deviation of 0.09 m. However, it remains sensitive to DEM biases and temporal elevation changes due to its reliance on a static DEM (Huang et al., 2024; Li et al., 2022). Finally, AMPLI (Altimeter data Modelling and Processing for Land Ice) (Aublanc et al., 2025a) represents a new facet-based waveform simulation approach to slope correction. When applied to Sentinel-3 UF-SAR (Unfocused Synthetic Aperture Radar) data over the Antarctic ice sheet, the AMPLI method reduced the median elevation bias and median absolute deviation relative to ICESat-2 ATL06 by 18% and 51%, respectively, compared to the ESA Sentinel-3 Land Ice Thematic Product, which uses a slope-based approach. In areas with sloped terrain ($>0.5°$), improvements reached up to 83% and 90%. Although this approach constitutes a complete pipeline from waveform to elevation extraction, including a retracking step, Aublanc et al. (2025a) employ the established Leading Edge Detection (LED) algorithm - originally developed for SARIn CryoSat-2 measurements - with their primary contribution focused on improved slope correction.

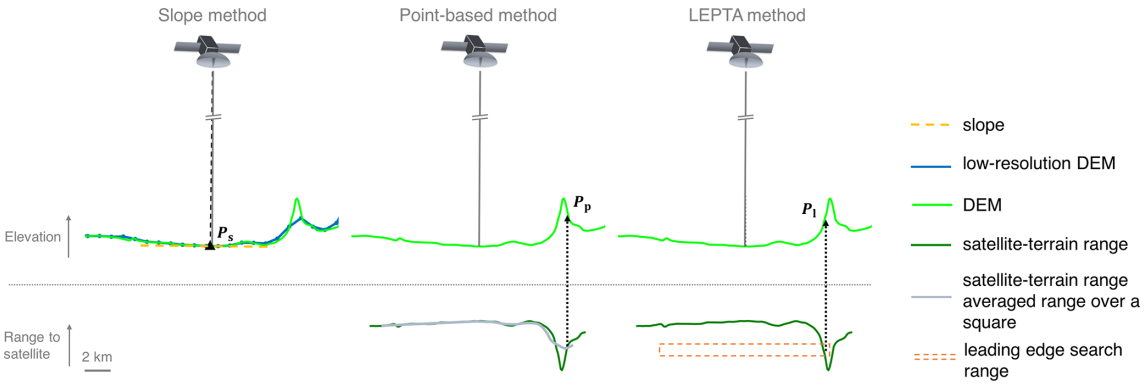

Figure 2.7: Illustration of the LEPTA slope-correction approach compared with standard slope- and point-based methods. The slope method estimates the POCA (P_s) by taking a planar slope derived from a DEM and selecting the point on this surface closest to the satellite. The point-based method determines the POCA (P_p) by minimising the average satellite-DEM range within a predefined, fixed-size rectangular footprint (e.g. 1.65 km × 1.65 km), typically corresponding to the pulse-limited footprint. The LEPTA approach further refines this by selecting only DEM points that contribute to the leading-edge signal and averaging them to locate the POCA (P_l), thereby improving accuracy by excluding non-contributing areas. From Li et al. (2022).

2.3.3 Geophysical and Instrumental Corrections

Before use, measurements need to be corrected for both geophysical and instrumental factors. These include accounting for the distance between the antenna and satellite centre of mass, dry and wet troposphere delays, ionospheric delays, and variations in the solid Earth, ocean loading, and polar tide (Quartly et al., 2020). Additionally, ocean tide and inverse barometer corrections need to be applied over floating ice such as ice shelves, the latter of which accounts for changes in atmospheric pressure. Measurements of the satellite range are also impacted by the on-board clock (ultra-stable-oscillator; USO) responsible for measuring the round-trip of the echo (Quartly et al., 2020). Approaches for accounting for these various corrections differ from mission to mission but commonly amount to adjustments of the order of several meters in total.

This concludes the summary of the principal processing steps in Level-2 ice sheet altimetry chains, along with the key innovations introduced during the transition to

SAR and SARIn altimetry. With this technical foundation established, I now turn to the satellite missions that implement these technologies and provide the data used in this work.

2.4 Satellite Missions used in this Work

Here I present overview descriptions of the main satellite altimetry missions used in this thesis, providing the observational foundation for the analyses and results in the following chapters. The subsections below summarise the design, operation, and key characteristics of each mission to contextualise their use in this work.

2.4.1 CryoSat-2

CryoSat-2 (CS2) is a satellite mission developed and operated by the European Space Agency (ESA), with the primary objective of monitoring variations in the thickness of Earth's continental ice sheets and marine ice cover. The mission was conceived to address the growing need for precise and comprehensive data on the most dynamic and climatically sensitive regions of Earth's cryosphere. For more than a decade, CryoSat-2 has played a crucial role in understanding how changes in ice thickness contribute to global sea-level rise and the overall dynamics of Earth's ice masses (Slater et al., 2021; McMillan et al., 2014; Landy et al., 2022; Hogg et al., 2018; Aublanc et al., 2021; Helm et al., 2014; Gourmelen et al., 2018). CryoSat-2 was launched on 8 April 2010, and forms part of ESA's Earth Explorer program. The mission had an initial design lifetime of 3.5 years, but is still operating successfully, now providing over 15 years of invaluable data. Operating in a non-sun-synchronous low Earth orbit (LEO) at an altitude of 720 km and with an inclination of 92°, CryoSat-2 has a repeat cycle of 369 days with a 30-day sub-cycle (Parrinello et al., 2018). This orbit allows the satellite to achieve near-complete coverage of polar regions, reaching up to 88° latitude, which is necessary for comprehensive monitoring of Antarctica and the central Arctic sea ice pack.

The primary instrument on CryoSat-2 is the Synthetic Aperture Radar/ Interferometric Radar Altimeter (SIRAL), which is designed to measure surface elevation. The instrument operates in three distinct modes: LRM, SAR, and SARIn (Parrinello et al., 2018). Over ice sheet interiors and oceans, the instrument operates in LRM, functioning as a conventional altimeter. In sea ice and some oceanographic regions, SAR mode is used to enhance along-track resolution for improved ice thickness estimates. Across the ice sheet margins and mountain glacier regions, SARIn mode is deployed, whereby interferometry is used to more precisely locate returns in areas of complex terrain, enabling more accurate measurements in these high-relief regions. SIRAL operates at a central frequency of 13.575 GHz (Ku-band radar) and has a pulse bandwidth of 320 MHz, with specific settings for different operational modes (Parrinello et al., 2018). The radar altimeter's data rate varies across modes, with 60 kbit/s for LRM, 12 Mbit/s for SAR, and 2 x 12 Mbit/s for SARIn (European Space Agency, 2021). This high data rate in SAR and SARIn modes supports the high PRF and, in turn, the delay-doppler processing required to resolve fine-scale surface features.

CryoSat-2's measurement accuracy is influenced by the satellite's footprint, which varies depending on the operational mode. The antenna beam-limited elliptical footprint spans approximately 13.2-14.0 km along-track and 14.9-15.8 km across-track, depending on the satellite's altitude (European Space Agency, 2021). In contrast, the pulse-limited footprint in LRM is approximately 2.15 km^2, with a diameter of 1.65 km in both directions, depending upon surface topography. For SAR and SARIn modes, the pulse-Doppler footprint area is reduced to about 0.5 km^2, due to the significantly narrower along-track Doppler beam width of approximately 380 m (European Space Agency, 2021).

The CryoSat-2 ground segment generates several levels of data products, which are broadly categorised according to the degree of processing applied to the telemetered data. Level-0 (L0) consists of raw data, which serves as the foundation for generating higher-level products. Level-1B (L1B) contains waveforms along the

satellite's ground track, including essential instrument and geophysical corrections, and is available in all operational modes (LRM, SAR, SARIn). Level-2 (L2) is derived from L1B data and provides surface elevation and other surface parameters, such as radar backscattering coefficients. These products include slope-corrected heights over land ice, ice thickness over sea ice, and various geophysical corrections. CryoSat-2 also produces thematic products under the Cryo-TEMPO initiative, which are tailored to specific scientific domains, such as land ice, sea ice, and polar oceans (European Space Agency, 2021). These products are processed using advanced altimetry techniques and are designed to be user-friendly with traceable uncertainties. This initiative produces singular POCA measurements, available in the standard Cryo-TEMPO product, and also interferometric swath measurements, available in the Cryo-TEMPO EOLIS product. The latter are made available as a monthly point product of dense elevation measurements, and as a monthly 2 km resolution gridded DEM with the data sourced from a rolling 3-month window.

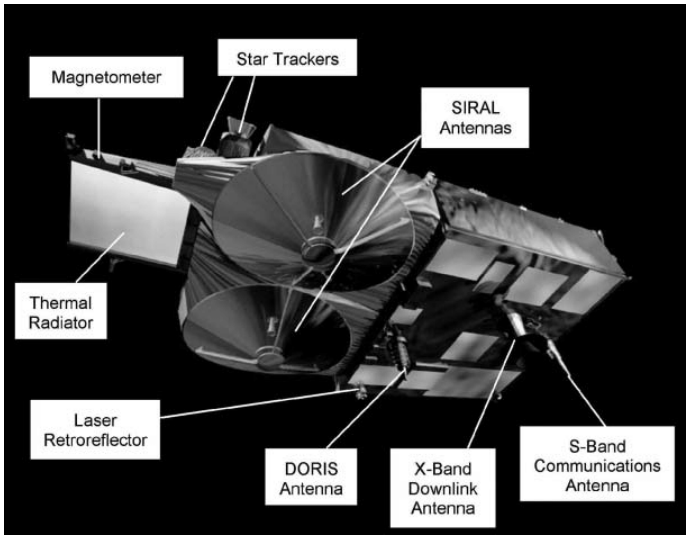

Figure 2.8: A diagram of the CryoSat-2 satellite and payloads. From Drinkwater et al. (2004).

Over ice sheet surfaces, CryoSat-2 has been used extensively to generate estimates of surface elevation and elevation change (Gourmelen et al., 2018; Helm et al., 2014; McMillan et al., 2014; Slater et al., 2021). For instance, Slater et al. (2021) demonstrated the mission's capability to monitor both seasonal variations

and long-term trends across the Greenland Ice Sheet, including detailed mapping of thinning patterns in coastal regions and mass balance changes over nearly a decade of observations (Figure 2.9).

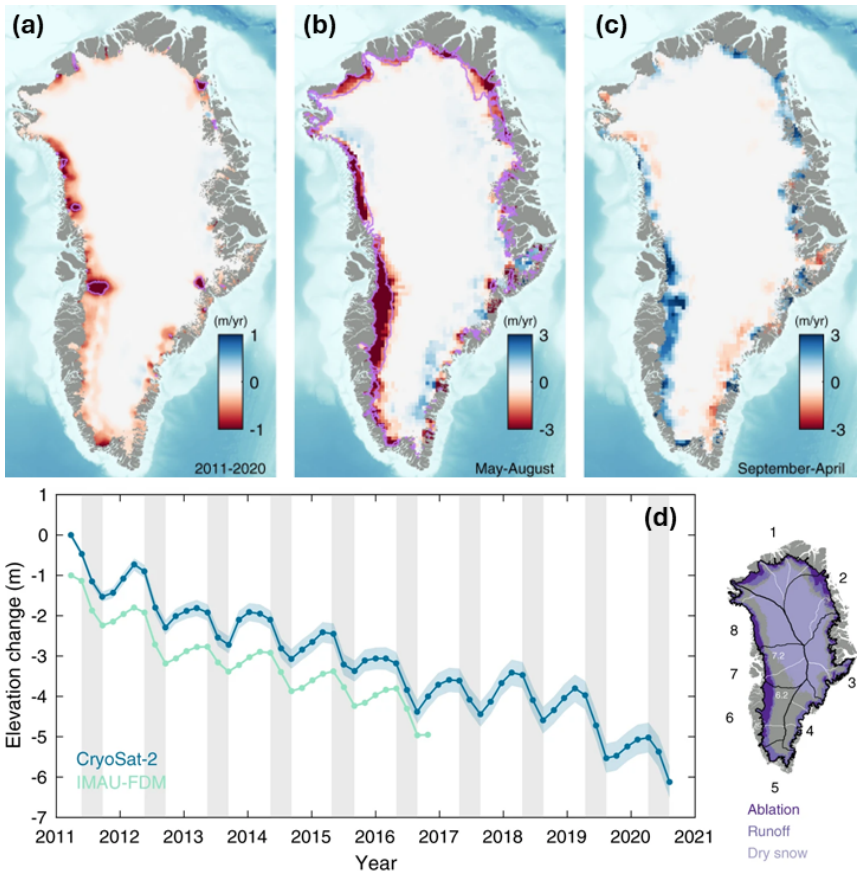

Figure 2.9: Surface elevation changes across the Greenland Ice Sheet derived from CryoSat-2 altimetry. (a) Rate of surface elevation change between 2011 and 2020. Purple contours indicate areas of long-term dynamical imbalance from repeat optical imagery (1985-2018) and dynamic elevation trends from satellite altimetry and the IMAU Firn Densification Model (2011-2017). (b) Mean elevation change rates during May-August 2011-2020, with purple contours showing the ablation zone extent. (c) Mean elevation change rates during September-April 2011-2020. (d) Left: Elevation change time series for the ablation zone from CryoSat-2 altimetry (dark blue; shading shows 1σ uncertainty) and the IMAU Firn Densification Model (light blue; offset by 1 m) from January 2011 to October 2020. Right: Ice sheet facies and principal drainage basins (black) and sub-basins (white). Seasonal contrasts between May-August (more negative) and September-April (more positive) reflect surface mass balance variability, with summer melt-driven lowering and winter accumulation-driven thickening. From Slater et al. (2021).

2.4.2 Sentinel-3

The EU's Copernicus programme has overseen the design and production of a significant number of new satellites during the 2010's and 2020's. One of these missions, Sentinel-3, has so far launched two satellites, the first of which, Sentinel-3A, launched on 6th February 2016, followed by Sentinel-3B on 25th April 2018. Sentinel-3 was designed, in part, as a successor to the European Space Agency's previous altimetry satellites, ERS-1, ERS-2, and Envisat, with the intention of transitioning to an operational programme, and securing a near-continuous ∼30 year record of global altimetry data. During the lifetimes of Sentinel-3A and 3B, two further satellite launches are planned, Sentinel-3C and D, in order to sustain an unbroken observational record into the next decade and, ultimately, pave the way for the Sentinel-3 Next Generation (S3NG) satellites (Egido, 2023). Sentinel-3 has an orbital period of 101 minutes such that after 385 complete revolutions (27 days), the ground-track is repeated to within 1 km. This orbital configuration yields a track-to-track spacing of 0.94° longitude at the Equator (Quartly et al., 2020). Both the 3A and 3B satellites orbit with a 98.65° inclination, thus providing coverage between 81.35° S and 81.35° N. The orbit is also configured to maintain a short-repeat sub-cycle, such that after 4 days the mission obtains quasi-global coverage, with a wider ∼7° spacing of ground tracks at the equator (Quartly et al., 2020). Driven by its mission objectives, Sentinel-3 operates in a lower inclination orbit and with a much smaller ∼60 m range window than CryoSat-2 (which operates at ∼92.03° inclination with a ∼240 m range window in SARIn mode, respectively). This reflects the satellite's global monitoring objectives, with a lower inclination favouring more uniform and frequent sampling at mid- to low-latitudes, and optimisation for low-relief surfaces, for which a ∼60 m range window is sufficient. This lower inclination orbit means that significantly more of the interior of the AIS is missed. However, as Sentinel-3 operates in SAR mode across the entire AIS (unlike CryoSat-2), much of the interior of the ice sheet is for the first time mapped with a higher along-track resolution. Due to the smaller range window of ∼60 m (EUMETSAT, 2017),

however, it is not uncommon for Sentinel-3 to fail to capture the surface echo, especially across coastal regions of the ice sheet that exhibit complex topography.

During routine operations, Sentinel-3 operates nominally in closed-loop tracking mode across the ice sheets, thus mirroring the approach implemented on many previous altimeters, such as Envisat and CryoSat-2. Although the satellite was equipped with open-loop tracking capabilities from launch, and indeed this was activated during the commissioning phase of Sentinel-3A, subsequent analysis indicated that the Open-Loop Tracking Command (OLTC) tables were not sufficiently precise over the steep Greenland and Antarctica ice sheet margins to satisfactorily track the complex ice surface (SentiWiki, n.d.). Recently, however, several target areas have again been acquired with open-loop tracking, for the purposes of re-evaluating and improving open-loop tracking over ice sheets. Notably, the updated OLTC tables are now informed by the high resolution REMA and ArcticDEM DEMs, with the aim to overcome the limitations encountered during commissioning.

The primary altimetry instrument on board Sentinel-3, SRAL (Synthetic Aperture Radar Altimeter), is a Ku-band SAR altimeter, which operates with a delay-Doppler beam-limited footprint of ∼300 m along-track by ∼18.2 km across-track. The across-track resolution of measurements is further constrained by the duration of the radar pulse, to ∼1.6-2 km, depending upon surface topography (Donlon et al., 2012a). Specifically, SRAL emits patterns of 64 coherent Ku-band pulses in a burst (i.e. with a high Pulse Repetition Frequency of 17.825 kHz), surrounded by 2 C-Band pulses (Donlon et al., 2012a). Accompanying SRAL, Sentinel-3 also carries a microwave radiometer (MWR), Sea and Land Surface Temperature Radiometer (SLSTR) and an Ocean and Land Colour Imager (OLCI) (Clerc et al., 2020).

Sentinel-3 data products are distributed under the Baseline Collection (BC), where each baseline represents a consolidated release of the processing chain. Earlier baselines, such as BC-004, provided a unified "Land Product" that applied a single processing convention across surface types under the Level-1/Level-2 convention seen

with CryoSat-2. With the release of BC-005, this was replaced by three distinct thematic product chains: hydrology, sea ice, and land ice, thus mimicking the approach taken by the CryoSat-2 Cryo-TEMPO products. Each chain implements dedicated retracking strategies, tailored geophysical corrections, and specific quality screening suited to the target surface. The hydrology chain is optimised for inland waters, the sea ice chain provides parameters such as freeboard and sea-ice thickness, and the land ice chain delivers elevation estimates over ice sheets and glaciers. More specifically, the BC-005 land ice product builds on BC-004 by introducing a number of ice-sheet-specific algorithms, most notably an artificial extension of the range dimension of the delay-Doppler stack during Level-1 range migration and multilooking, to better preserve backscattered energy in the final waveform, particularly over the ice-sheet margins (Aublanc et al., 2018; Landy et al., 2019; Aublanc et al., 2025b)

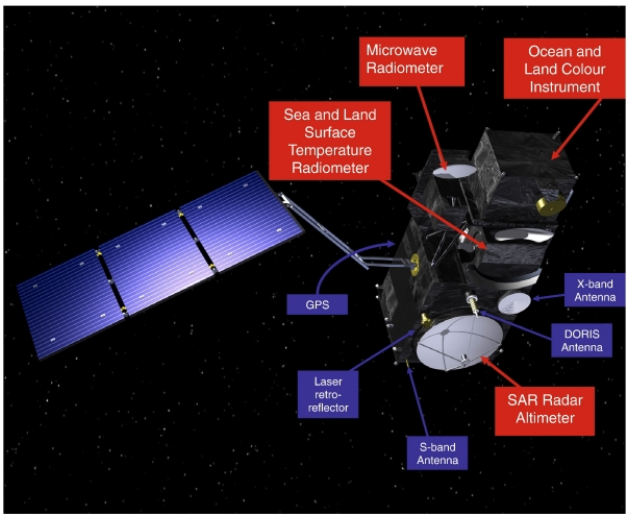

Figure 2.10: A diagram of the Sentinel-3 satellites and payloads. From European Space Agency (2024).

Similar to CryoSat-2, Sentinel-3 has also seen application over the polar ice sheets, though not as extensively, and with less emphasis on tracking changes in surface elevation (Kokhanovsky et al., 2019; Mugunthan et al., 2023; Lawrence et al., 2021; Niehaus et al., 2024). Specifically, McMillan et al. (2019) demonstrated

the mission's capability to monitor elevation trends across Antarctica using SAR altimetry, providing continent-wide coverage of ice sheet dynamics (Figure 2.11).

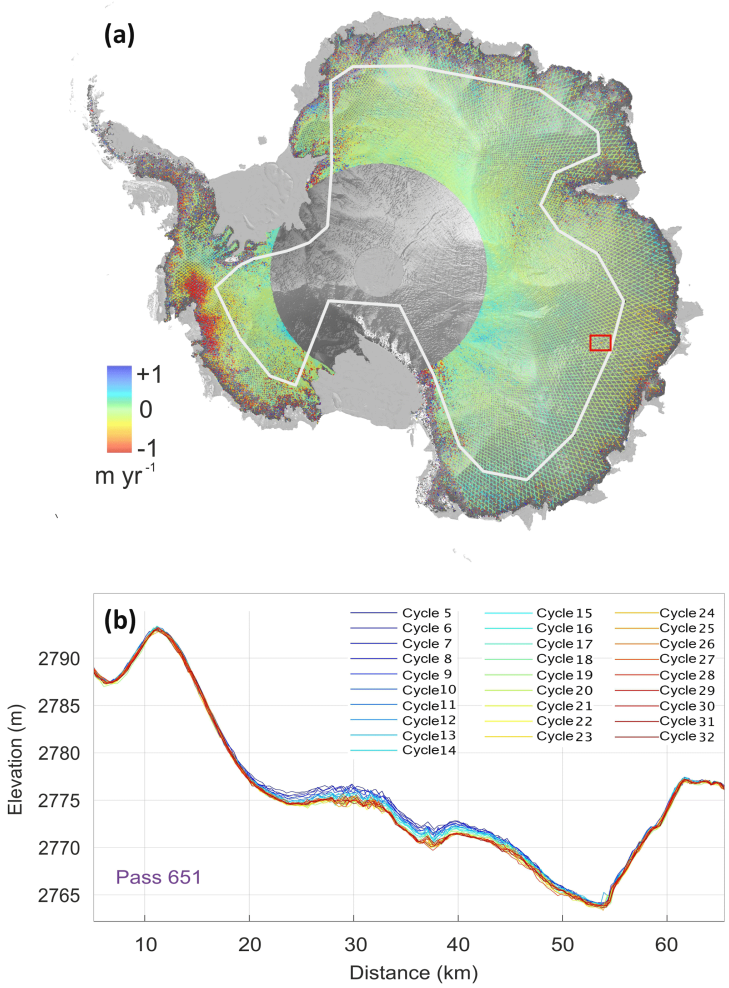

Figure 2.11: (a) Antarctic surface elevation change rates from Sentinel-3A Delay-Doppler altimetry (May 2016-June 2018) overlaid on shaded relief DEM and MODIS Mosaic of Antarctica. White line shows CryoSat-2 mode mask boundary separating low-resolution (interior) and SAR interferometric (coastal) regions. Data gap around South Pole reflects 81.35° orbital limit. (b) Localised surface lowering across 30 km track segment in East Antarctica (red box in panel a) indicating possible subglacial lake drainage. From McMillan et al. (2019).

2.4.3 ICESat-2

The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) is a NASA mission launched on 15th September 2018 to measure changes in ice sheet and sea ice

elevation. Unlike the radar altimeters described above, ICESat-2 uses laser (lidar) technology rather than radar to directly measure surface height (Markus et al., 2017). It succeeds the original ICESat mission (2003-2009), providing higher-resolution data for quantifying ice mass balance, monitoring sea-level contributions, and improving understanding of cryosphere dynamics (Magruder et al., 2021; Markus et al., 2017).

ICESat-2 carries a single instrument, the Advanced Topographic Laser Altimeter System (ATLAS), a photon-counting laser altimeter transmitting green laser pulses (532 nm) at 10 kHz (Markus et al., 2017). ATLAS splits the beam into six ground tracks arranged in three pairs, with ∼3 km spacing between pairs and ∼90 m separation within pairs, providing dense coverage and redundancy. By recording the travel time of individual photons, and aggregating these to generate elevation measurements, ATLAS retrieves surface elevations with centimeter-level precision and absolute accuracy of ∼0.1 m (Magruder et al., 2021). The satellite operates in a near-polar, sun-synchronous orbit at 92° inclination with a 91-day repeat cycle, giving near-complete coverage of Antarctica every three months, with much denser sampling toward the poles due to orbital convergence (Magruder et al., 2021; Markus et al., 2017).

While ICESat-2 provides some of the most accurate elevation data over the polar regions and is widely used to calibrate and validate radar altimetry missions (e.g., CryoSat-2, Sentinel-3) (Aublanc et al., 2025b; Helm et al., 2014; Crétaux et al., 2018; Ravinder et al., 2024; Aublanc et al., 2025a; Dawson et al., 2023), it differs fundamentally from the radar instruments considered previously. Its measurements are limited to narrow ground tracks and require clear-sky conditions, resulting in lower spatial and temporal coverage compared with the broader, more continuous coverage offered by radar altimeters (Pronk et al., 2024). Nonetheless, ICESat-2 is included here because it provides the reference measurements against which radar altimetry products are routinely validated in cryospheric research, and it is used in the same capacity in this thesis.

ICESat-2 has also been used to generate estimates of surface elevation and elevation change (Figure 2.13). For instance, Ravinder et al. (2024) demonstrated how laser altimetry can be combined with radar measurements to provide complementary observations of ice sheet dynamics and improve the accuracy of elevation change assessments over Greenland.

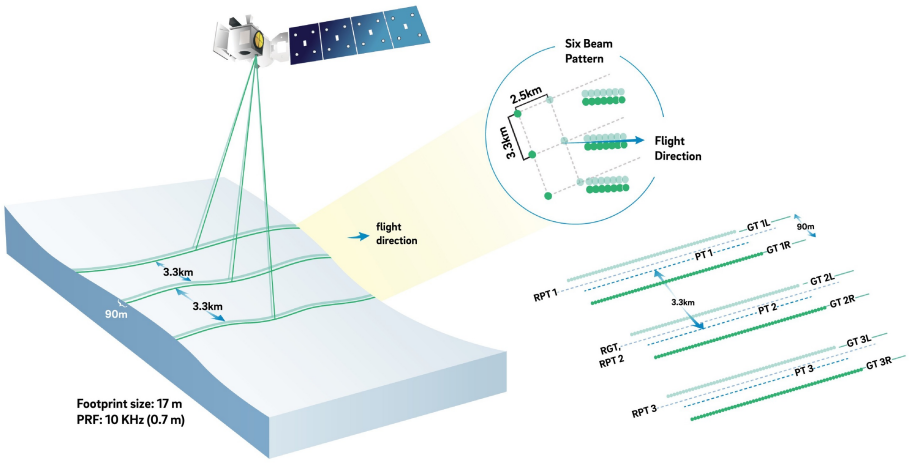

Figure 2.12: A diagram of the six-beam pattern of the Advanced Topographic Altimeter System (ATLAS) instrument onboard ICESat-2, and a demonstration of its application. Adapated from Smith et al. (2019).

2.5 Radar Altimetry: Summary and Outlook

Over the past three decades, satellite radar altimetry has developed from an oceanographic tool with limited capability over ice sheets into a cornerstone technique for monitoring the cryosphere. Early missions such as ERS-1 and ERS-2 established the first continuous records, but their reliance on pulse-limited, low-resolution systems restricted performance over regions of complex topography (Joughin et al., 1996; Rémy et al., 2009). Subsequent innovations - including synthetic aperture radar (SAR) processing and interferometry (SARIn) - have greatly enhanced the precision, resolution, and spatial sampling of altimetric measurements, enabling detailed mapping of ice sheet elevation and dynamical processes (McMillan et al., 2019; Slater et al., 2021). CryoSat-2 pioneered the use

of SAR and interferometric altimetry over ice and continues to provide invaluable data, while the Sentinel-3 constellation has secured the long-term continuity of SAR observations. Looking ahead, the planned CRISTAL mission, which is intended to succeed CS2, will continue to extend this capability with multi-frequency SAR and interferometric measurements (Kern et al., 2020).

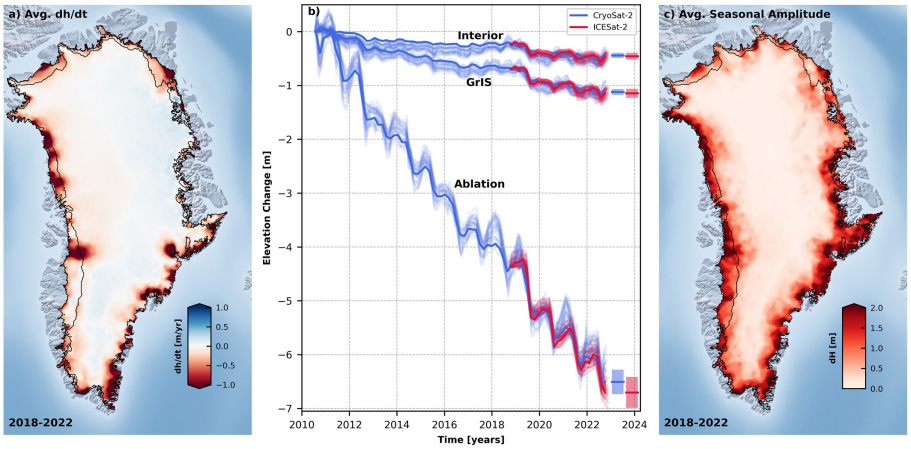

Figure 2.13: Greenland Ice Sheet surface elevation changes from combined CryoSat-2 and ICESat-2 altimetry (2018-2022). (a) Rates of elevation change averaged from both missions. (b) Time series of bi-monthly elevation changes from CryoSat-2 (blue) and ICESat-2 (red); dark lines show preferred scenarios with shaded bands indicating uncertainties, light lines show ensemble scenarios. (c) Amplitudes of seasonal cycle averaged from both missions. A 25 × 25 km median filter is applied to panels (a) and (c) for visualisation. From Ravinder et al. (2024).

Together, these developments mark a clear trajectory toward higher-resolution, more frequent, and more accurate monitoring of the polar regions. Yet important limitations remain. The pre-2010 record - from ERS, Envisat, and Jason missions - consists solely of non-interferometric, LRM altimetry, leaving a gap in the ability to trace the long-term evolution of ice sheet mass balance with the same level of spatial precision and accuracy that is now possible. Current (and future planned) Sentinel-3 satellites, while providing operational SAR coverage, lack interferometric capability (Quartly et al., 2020), and even CRISTAL will only apply SARIn at Ku-band, not at Ka-band (Kern et al., 2020). As a result, the full potential of radar altimetry over complex ice surfaces has yet to be realised.

With altimetric datasets set to expand in both volume and temporal coverage, the challenge now lies less in the application of new instruments and more in overcoming the inherent limits of existing and historical systems. This requires rethinking how radar waveforms are processed and how geophysical information is extracted. By developing approaches that more fully exploit the information content of the altimetric signal, there is considerable potential to enhance the value of both historical and future measurements.

Chapter 3

Deep Learning Prerequisites

The preceding sections highlighted both the instrumental advances and the persistent limitations of satellite radar altimetry. While SAR and interferometric processing have markedly improved measurement capability, the extraction of reliable geophysical information still depends heavily on simplifying assumptions and processing strategies designed to compensate for non-interferometric ambiguity (Andersen, 2025; Aublanc et al., 2025b). This creates a natural role for data-driven approaches: methods that can learn directly from the complex, nonlinear structure of altimetric waveforms and related geophysical datasets. Deep learning, in particular, offers a powerful framework for addressing these challenges, given the abundance of data now available, yet to date its application to the field of ice sheet altimetry is extremely limited. The following sections provide an overview of deep learning methods, offering essential context for Chapters 6 and 7, where their development and application to ice sheet altimetry are explored in detail.

3.1 Overview

Deep learning is a subset of machine learning, which broadly refers to algorithms that learn from data without relying on hand-crafted rules. Deep learning is distinguished by its layered architecture, which automatically extracts and combines features at

progressively higher levels of abstraction. These models are particularly effective for capturing complex, nonlinear relationships in high-dimensional data, especially when patterns are difficult to define explicitly but can be learned from large datasets (Goodfellow et al., 2016). Over the past decade, deep learning has been applied to a wide range of problems in environmental science, often outperforming traditional methods in both accuracy and generalisability. Applications include land cover classification from satellite imagery (Zhang et al., 2020; Campos-Taberner et al., 2020), nowcasting extreme weather events (Ravuri et al., 2021; Weyn et al., 2020; Pan et al., 2021), and monitoring air and water quality (Agbehadji et al., 2024; Zhi et al., 2024; Zheng et al., 2025). In cryospheric research specifically, deep learning has shown success in a variety of use cases (Liu, 2021), including identifying glaciers and tracking their evolution (Maslov et al., 2025; Chu et al., 2022; Li et al., 2025), measuring sea ice concentration (Kvanum et al., 2024; Chen et al., 2024; Andersson et al., 2021), and mapping snow cover and depth (Breen et al., 2024; Wang et al., 2020; Yu et al., 2022).

In this work, I focus on Convolutional Neural Networks (CNNs) - a class of models well suited to structured, spatial data. The following sections provide an overview of the fundamental components underlying deep learning architectures. I begin with fully-connected networks, then expand to CNNs and specific CNN architectures, and finally discuss probabilistic networks and methods for quantifying prediction uncertainty.

3.2 Fully-Connected Neural Networks

In their general form, neural networks consist of interconnected artificial neurons arranged in layers. Each neuron takes in values as input, computes their weighted sum, adds a bias, and applies an activation function to generate a single-valued output (Figure 3.1). Here, activation functions work to incorporate non-linearity into the model and, in some cases, to transform neuron outputs into specific forms

such as probabilities (Goodfellow et al., 2016; Bishop, 1995). At the input layer to a network, each element in some input vector is fed to all of the neurons in the subsequent layer, which compute outputs, passing them to the following layer. This process is repeated layer-by-layer until the output layer is reached, where predictions are made. The layers between the input and output layers are known as hidden layers and are responsible for learning complex patterns in the data (Figure 3.1). This arrangement, where every neuron in one layer connects to every neuron in the next, is referred to as a fully-connected neural network (Goodfellow et al., 2016).

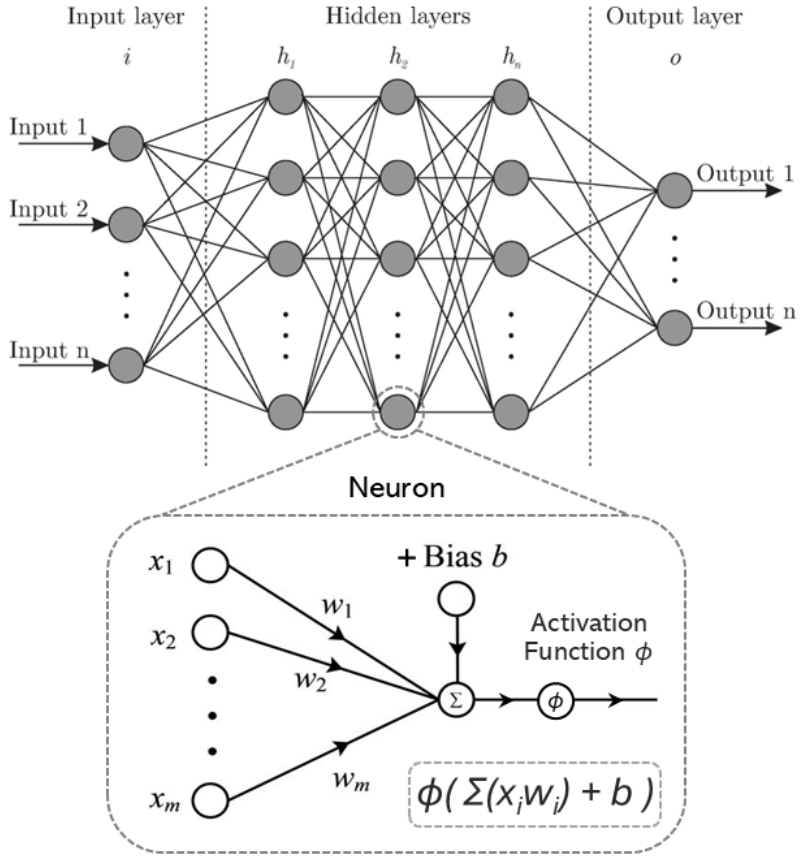

Figure 3.1: Example diagram of a fully-connected, feed-forward network. Adapted from Shao (2018) and Bre et al. (2018).

To perform a task effectively, a neural network must learn appropriate values for its internal parameters, which are initially randomly selected. This process - known as training - involves adjusting the network's weights and biases to improve

its predictions based on examples (LeCun et al., 2015). The goal is to minimise a loss function, which quantifies the error between the network's prediction Y, given some input X, and the known target output $\hat{Y}$ (Goodfellow et al., 2016). During training, data is fed through the network, and the loss with respect to the expected output is computed. This loss is then propagated back through the network, where the weights and bias in the neurons are updated in the direction of the steepest descent of the loss landscape; a multi-dimensional space comprising the network loss with respect to the model parameters (weights and biases) (Figure 3.2). This update occurs with distance e, referred to as the learning rate (Rumelhart et al., 1986). The algorithm commonly used in calculating the gradients of the loss function with respect to the parameters, and incrementally updating them, is referred to as gradient descent (Rumelhart et al., 1986; LeCun et al., 2015). The process of doing this with a full set of data once is called an epoch and is repeated until convergence is achieved, ensuring the network optimally captures underlying patterns in the data (Goodfellow et al., 2016).

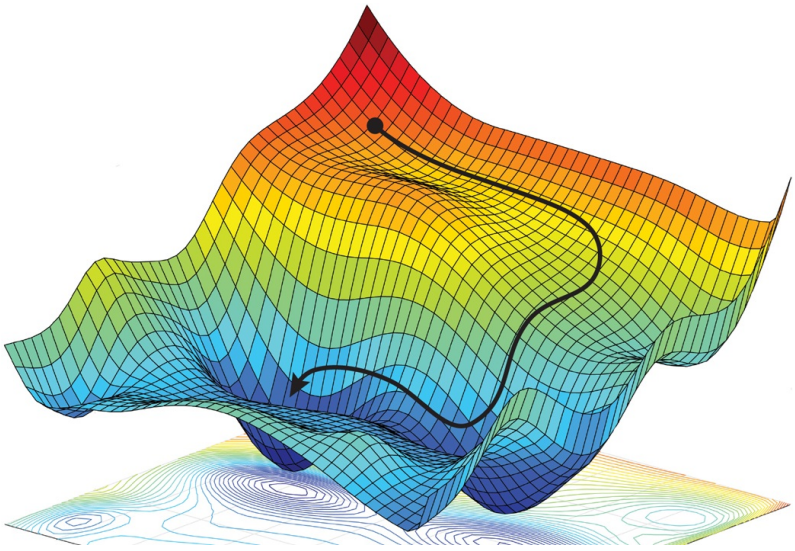

Figure 3.2: Abstract representation of the "loss landscape". The z-axis denotes the loss, while the x- and y-axes provide a two-dimensional projection of the network's weights and biases. The black arrow illustrates a possible trajectory taken during training via backpropagation. In reality, this landscape exists in a space with far more dimensions. From Amini et al. (2019).

Whilst neural networks are adept at solving various tasks, including regression and classification, achieving optimal performance requires careful tuning of various hyperparameters such as layer depth, neuron count, activation functions, and learning rates (Goodfellow et al., 2016). This hyperparameter tuning is crucial for ensuring the effectiveness of neural network models, and in practice is typically carried out through systematic search strategies (e.g., grid or random search), optimising for model performance (Bergstra et al., 2012).

3.3 Convolutional Neural Networks

Fully-connected layers in neural networks, although effective, have limitations when handling multidimensional data, such as images. The need to flatten input data leads to a loss of spatial information, which is crucial in structured data (Goodfellow et al., 2016; LeCun et al., 2015). Moreover, these networks lack translation invariance, learning features in-place (Goodfellow et al., 2016). This means that knowledge of a given feature in an image is location-dependant. Additionally, these networks require an unmanageable number of weights for large structured inputs (such as images) due to their fully-connected nature (Bishop, 1995; Goodfellow et al., 2016).

Convolutional Neural Networks (CNNs) address these issues through the use of convolutional layers, which use learnable filters (kernels) to extract specific features from small, overlapping regions of the input data through element-wise multiplications and summations as they slide (convolve) across the input (Figure 3.3) (Lecun et al., 1998; Goodfellow et al., 2016; Rawat et al., 2017). The filters in CNNs - which can be applied in varying, odd-numbered sizes (e.g. 1×1, 3×3, 5×5) - function similarly to the weights and biases in fully-connected neural networks (being technically a multi-dimensional extension), which are likewise updated via backpropagation (Rumelhart et al., 1986; Lecun et al., 1998). Each convolutional layer applies one or more filters to capture different features, followed by a nonlinear activation function which is applied element-wise to the output (Goodfellow et al.,

2016). This local processing, where each neuron considers only a small portion of the input, is essential for capturing local patterns and features (Lecun et al., 1998).

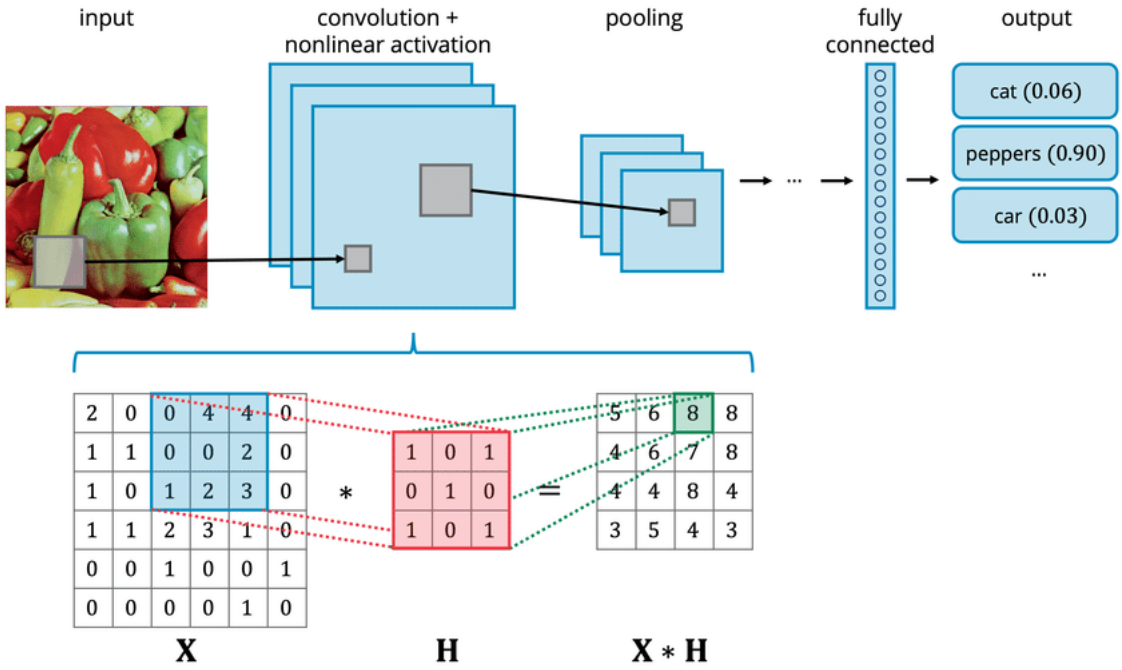

Figure 3.3: Diagram illustrating the process of a convolutional neural network. An input image (a pepper, in this case), represented as a 3-channel RGB image, is passed through the network. At each convolutional layer, learnable kernels (e.g., a 3×3×1 kernel shown here) convolve over the image to generate feature maps. In this case, the resulting feature map is smaller than the input image, but padding can be applied to the image (e.g. zeros at the edges) to result in a same-size output. These feature maps are then pooled to reduce spatial dimensions, thereby increasing the effective receptive field of subsequent kernels. After multiple such layers, the feature maps are flattened and passed to a fully connected layer, followed by a final activation function, which outputs class probabilities between 0 and 1 for N possible categories. From Cheung et al. (2020).

The use of convolution reduces redundancy and inefficiency seen in fully-connected layers through sparse interaction, where each neuron connects to only a local subset of inputs rather than all inputs, significantly reducing parameters and computation (Goodfellow et al., 2016; Lecun et al., 1998; Rawat et al., 2017). Filters, being smaller than the input data, extract meaningful information across the entire data space, reducing memory requirements and improving learning efficiency, as fewer parameters need to be estimated from the available training data (Goodfellow

et al., 2016; LeCun et al., 2015). This also facilitates translation invariance, as filters can detect specific features throughout the image via weight sharing (Fukushima, 1980; Lecun et al., 1998). Consequently, CNNs are equivariant with respect to translation (Goodfellow et al., 2016).

Following convolutional layers, CNNs commonly implement pooling layers, which reduce the spatial dimensions of feature maps while retaining essential information (Lecun et al., 1998; Goodfellow et al., 2016). This down-sampling, typically performed via max pooling or average pooling, allows the network to process data in gradually larger spatial contexts (Goodfellow et al., 2016). After several convolutional and pooling layers, the network commonly ends with one or more fully-connected layers, which handle the flattened output from previous convolutional and pooling layers to produce the network's final output (Figure 3.3) (Lecun et al., 1998).

Although CNNs are built for multi-dimensional data, they have also seen extensive success in one-dimensional tasks such as time series analysis (Kashiparekh et al., 2019; Tang et al., 2022), speech recognition (Han et al., 2020; Abdel-Hamid et al., 2014), and natural language processing (Kim, 2014; Jacovi et al., 2020). Their ability to capture local patterns and hierarchies, combined with reduced parameter numbers due to convolutional layers and weight sharing, makes CNNs more robust and efficient than fully-connected networks for these pattern recognition tasks. In contrast, fully-connected networks often suffer from over-parameterisation and become impractical due to excessive connections that require prohibitively large amounts of memory and training data. These factors have made CNNs indispensable in various environmental application, including land-cover classification and semantic segmentation of remote sensing imagery (Li et al., 2021; Sumbul et al., 2019); forecasting forest-loss and crop-loss risk (Ball et al., 2022; Thapa et al., 2022); and climate-variable downscaling and weather/climate forecasting (Fallah et al., 2025; Elabd et al., 2025).

3.4 Deep Learning Architectures Underpinning this Work

To provide more specific context for the models applied in this thesis, it is useful to introduce the ResNet family of architectures. The following sections first outline the design principles of the original ResNet, which has become a foundational deep learning architecture, and then describe ResNet-RS, the specific variant employed in Chapters 6 and 7. In addition to introducing these models, this overview also highlights broader deep learning concepts that are relevant throughout the thesis.

3.4.1 ResNet

This section provides a general overview of ResNet as background for Section 3.4.2, which describes the specific variant used in this thesis.

In deep neural networks comprised of many layers, as gradients are back-propagated they can become exceedingly small, resulting in what is known as the vanishing gradient problem (Hochreiter, 1991; Bengio et al., 1994). This occurs because repeated multiplication of small derivatives during backpropagation causes updates to earlier layers to diminish towards zero, preventing those layers from learning effectively (Goodfellow et al., 2016). As a result, the network struggles to converge and its performance degrades (Bishop, 1995). To address this, He et al. (2015) introduced ResNet (Residual Network), a CNN architecture designed to maintain effective gradient flow through very deep networks (Figure 3.4).

The core innovation of ResNet is the residual block - a repeatable building block composed of convolutional layers and a skip (or shortcut) connection that can be stacked together to create very deep networks. This skip connection enables the input to bypass the block and be added directly to its output, facilitating more effective gradient propagation during training (He et al., 2015). By addressing the vanishing gradient problem, the authors demonstrated that this architecture not only stabilises deep network training but also improves convergence speed and overall

model robustness (He et al., 2015).

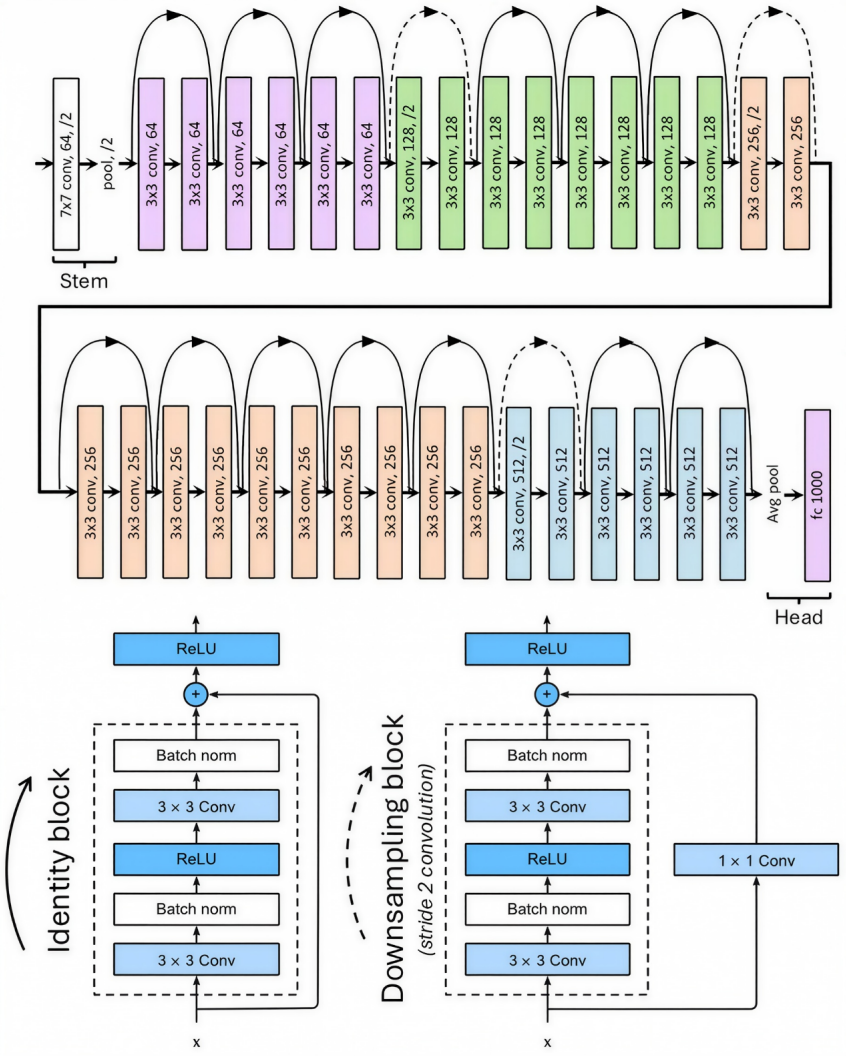

Figure 3.4: Diagram of the ResNet-34 architecture. Colours indicate the four residual stages, each consisting of multiple residual blocks. Within each stage, identity blocks are bracketed with solid skip connections, while downsampling blocks are bracketed with dashed skip connections. Adapted from Thakkar (2024) and Zhang (2020).

The ResNet architecture is organised into four main residual stages, bracketed by a stem block at the beginning and a prediction head at the end (Figure 3.4). The stem block performs an initial 7×7 convolution with stride 2, followed by a 3×3 max pooling operation, reducing spatial resolution while increasing channel depth. This effectively shifts computationally expensive spatial processing into the channel

dimension, improving efficiency and feature representation (He et al., 2015). Here, channels represent the depth dimension of the data (for example, RGB images have 3 channels, but internal network layers may have hundreds) (Goodfellow et al., 2016). The prediction head applies a global average pooling layer (first introduced by Lin et al. (2014) and later popularised by the Inception architecture (Szegedy et al., 2015)), which reduces each feature map to a single value, followed by a final, fully connected layer which outputs the predictions.

Residual blocks in ResNet are typically divided into two functional types: (1) downsampling blocks and (2) identity blocks (Figure 3.4). Downsampling blocks reduce the spatial resolution of feature maps (individual channels of the data) and usually appear at the start of each new stage in the network - where a stage is a group of layers or blocks (groups of layers), commonly operating at the same spatial resolution. Instead of using traditional pooling operations for spatial downsampling like max or average pooling, ResNet uses convolutions with a stride of 2 (meaning the learnable filter moves 2 pixels at a time instead of 1), effectively halving the image dimensions. The shortcut path in these blocks is also adjusted using a $1\times\times1$ convolution with a stride of 2 to match both the spatial and channel dimensions of the main path. Identity blocks, on the other hand, maintain the spatial dimensions and are used when the input and output shapes are the same. Each residual stage begins with one downsampling block, followed by a sequence of identity (bottleneck) blocks.

In ResNet, the activation function used is ReLU (Rectified Linear Unit), a popular choice due to its simplicity and effectiveness in introducing non-linearity into the model (Nair et al., 2010; Goodfellow et al., 2016). ReLU outputs the input directly if it is positive, and zero if it is negative (Goodfellow et al., 2016). This functionality helps accelerate the convergence of the training process by mitigating the vanishing gradient problem commonly associated with other activation functions (Glorot et al., 2011). ResNet also incorporates batch normalisation after each convolutional layer in the stem, bottleneck, and downsampling blocks. Batch

normalisation is a technique that standardises the data flowing through the network during training, and is applied to groups of training samples (called batches) that are processed together (Ioffe et al., 2015). This standardisation reduces a problem called internal covariate shift, where the statistical properties of data change as it passes through different layers of the network during training (Ioffe et al., 2015). By stabilising these statistical properties, batch normalisation allows the network to train more quickly and reliably, leading to improved model performance (Ioffe et al., 2015).

In ResNet, the number of blocks and the number of output channels at each stage determine the overall depth and capacity of the network. For example, ResNet-50 consists of 50 layers, with 3, 4, 6, and 3 bottleneck blocks in its four stages, respectively. In deeper ResNet variants like ResNet-50 and beyond, both residual block types adopt a bottleneck architecture (Figure 3.5). Each bottleneck block consists of three convolutions: a 1×1 convolution that reduces the number of channels (creating the "bottleneck" - layers of data compression where the dimensions are lower than the neighbouring layers), a 3×3 convolution at the reduced dimensionality, and a final 1×1 convolution that expands the feature (channel) dimension, typically by a factor of 4. This structure reduces the computational cost, as the more expensive 3×3 convolution operates on a narrower set of channels (He et al., 2015).

ResNet-50 is one of the smallest of the commonly used ResNet variants and typically uses a width factor of 1, where the width factor controls how many channels (feature maps) the network uses at each layer. This results in output channel dimensions of 256, 512, 1024, and 2048 at each stage, which are the expanded dimensions after the final 1×1 convolution in each bottleneck block. The internal 3×3 convolution layers operate on reduced channel dimensions of 64, 128, 256, and 512, respectively, before being expanded fourfold.

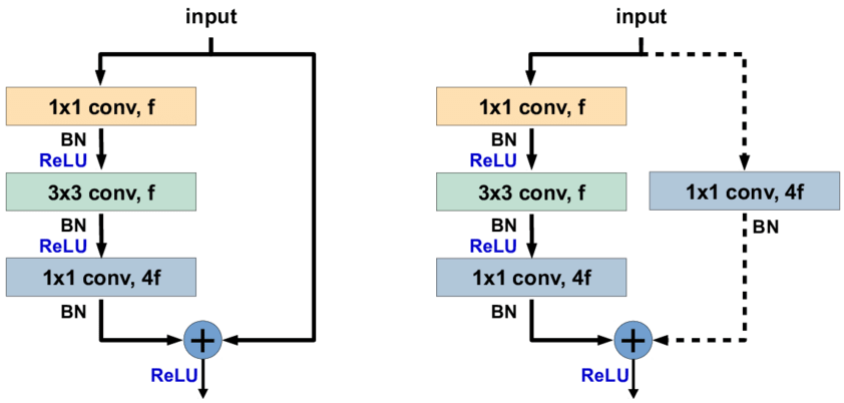

Figure 3.5: Diagrams of the identity block (left) and downsampling block (right) in ResNet-50 and beyond. From de Rezende et al. (2018).

These architectural hyperparameters - depth and width - are tuneable. Deeper networks tend to perform better on tasks requiring hierarchical feature extraction, while wider networks may capture a richer variety of features at each level (Zagoruyko et al., 2017). However, increasing depth or width does not guarantee improved performance (Goodfellow et al., 2016). Overparameterised models - those with more parameters than needed - can lead to overfitting, where the model excels on training data but fails to generalise to unseen inputs (Goodfellow et al., 2016; Bishop, 1995).

ResNet in particular has seen significant successes within the field of environmental science - from medium-range weather forecasting and operational air-quality (ozone) prediction (Rasp et al., 2021) to flood inundation detection (Mandawade et al., 2025), hyperspectral crop monitoring and classification (Bu et al., 2024), and automated photovoltaic defect inspection (Abdelsattar et al., 2025), consistently yielding state-of-the-art performance in high-impact studies.

3.4.2 ResNet-RS

This section describes ResNet-RS, the specific variant employed in Chapters 6 and 7.

Since ResNet in 2015, many different CNN architectures have been developed,

each offering improved performances on standardised datasets such as ImageNet (Deng et al., 2009), with notable models including Xception (Chollet, 2017), DenseNet (Huang et al., 2017), ResNeXt (Xie et al., 2017), MobileNet V2 (Sandler et al., 2018), and EfficientNet (Tan et al., 2020). However, Bello et al. (2021) at Google demonstrated that ResNet, with modern training methods and architectural tweaks, can outperform these newer models, achieving faster speeds and comparable accuracies on the ImageNet dataset. This model is referred to as ResNet-RS.

One of the architectural innovations introduced in ResNet-RS is the Squeeze-and-Excitation (SE) block, first introduced by Hu et al. (2018). This block, which is placed after every residual block, enhances the network's ability to focus on important features by adaptively weighting feature channels. This is done by compressing each channel into a single representative value via global pooling, whereafter convolutional layers are then used to weight the importance of each channel according to these global descriptors (Hu et al., 2019). This concept is generally referred to as attention (Figure 3.6).

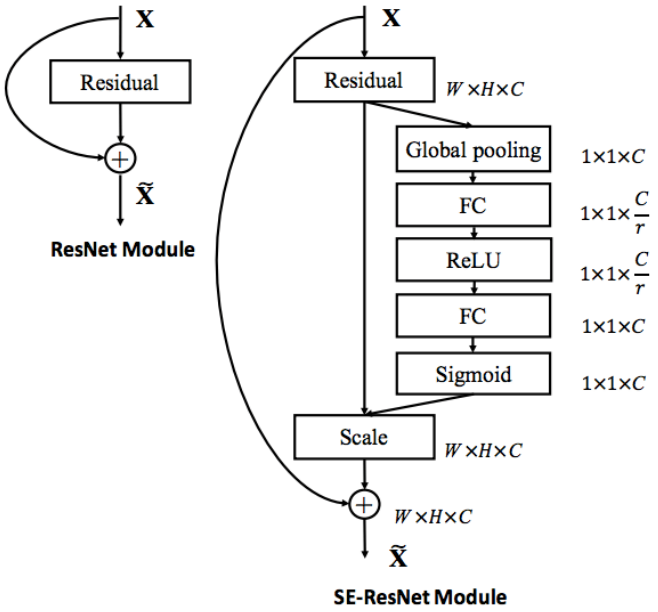

Figure 3.6: The original ResNet residual block (left) and the residual block with Squeeze-and-Excitation implemented (right). From Hu et al. (2019).

As well as SE blocks, ResNet-RS also incorporates several architectural changes found in ResNet-D, an improved version of ResNet introduced by He et al. (2018) (Figure 3.7). Firstly, the 7×7 convolution in the stem block is replaced by three smaller 3×3 convolutions, which reduces the number of trainable parameters whilst keeping the same receptive field (the region of the input image that each neuron can "see"). This is an approach referred to as factoring, popularised by the Visual Geometry Group (VGG) in 2014 in their network VGG-16 (Simonyan et al., 2015). Secondly, in the residual path of the downsampling blocks, the stride sizes are switched for the first two convolutions, as the original ordering leads to three-quarters of the input feature map being ignored. For the same reason, the stride-2 1×1 convolution in the skip connection path of the downsampling blocks is also replaced by a stride-2 2×2 average pooling layer and then a non-strided 1×1 convolution. Finally, global average pooling is applied prior to the final fully-connected layers (Lin et al., 2014; Szegedy et al., 2015). This serves to avoid overfitting since there are no parameters to optimise and is also more robust to spatial translations of the input.

When training ResNet-RS, Bello et al. (2021) employ a range of strategies to enhance performance and robustness, closely following approaches taken in training EfficientNet by Tan et al. (2020), a more complex model outperformed by ResNet-RS. To improve generalisation, dropout is applied after the final global average pooling layer (Srivastava et al., 2014), which randomly deactivates neurons during training to prevent co-adaptation (where neurons become too reliant on specific other neurons rather than learning robust, independent features), and so overfitting. This is complemented by the use of stochastic depth (Huang et al., 2016), which extends the dropout concept by randomly dropping entire blocks of layers, forcing the network to learn multiple redundant pathways and further improving its ability to generalise.

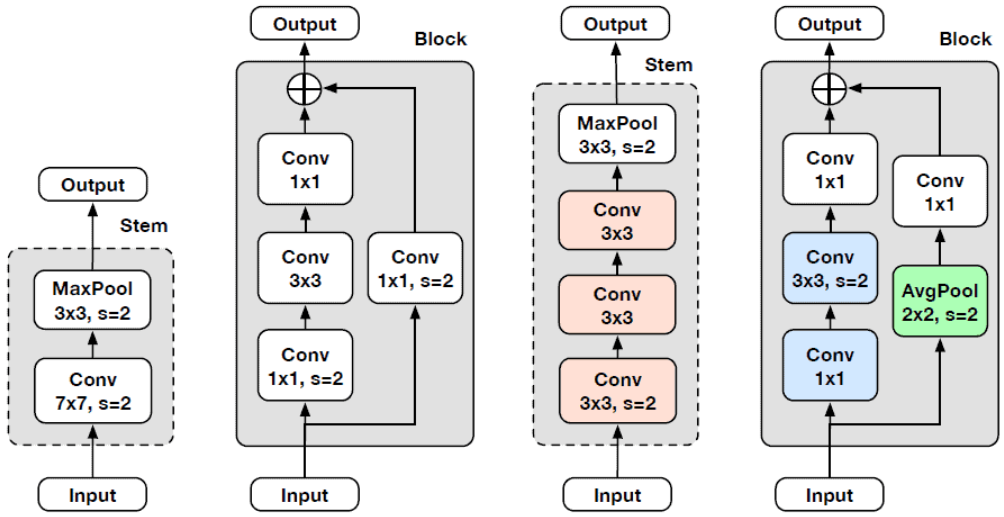

Figure 3.7: Comparison of changes for the stem and downsampling blocks between ResNet and ResNet-D. From Lee et al. (2020).

To facilitate smoother convergence, a cosine learning rate scheduler with linear warmup is implemented (Loshchilov et al., 2017; Goyal et al., 2018). In practice, this means that the learning rate starts at a very small value (to prevent unstable updates at the beginning of training), then increases linearly over a short warmup phase, before gradually decaying according to a cosine curve. This schedule allows the optimiser to explore the parameter space more aggressively early on, while ensuring progressively smaller updates as training approaches convergence, reducing the risk of overshooting the optimum.

Additionally, to combat overfitting, weight decay (L2 regularisation) is employed (Krogh et al., 1991), which penalises large weights by adding their squared magnitude to the loss function. This discourages overly complex solutions and constrains the model's capacity, thereby promoting generalisation. Importantly, the regularisation is applied only to convolutional and fully connected weights, leaving biases and batch normalisation parameters unaffected, as penalising these terms can

interfere with stable training (Ioffe et al., 2015).

As an extension to the gradient descent algorithm discussed in Section 3.2, ResNet-RS also utilises a momentum optimiser (Polyak, 1964), which accelerates convergence by incorporating the direction and magnitude of previous updates. By maintaining momentum in the direction of consistent gradients, it helps smooth oscillations and aids in escaping local minima, leading to better overall solutions. Many different optimisers exist, with notable alternatives to the momentum optimiser including RMSprop (Hinton et al., 2012), which was used by Tan et al. (2020) when training EfficientNet. This optimiser works by adapting the learning rate for each parameter individually based on a moving average of recent gradient magnitudes, allowing for more effective training, especially in models with large parameter spaces. The momentum optimiser was, however, chosen by Bello et al. (2021) for simplicity. They also departed from random initialisation by using a pre-trained model to initialise the parameters in the ResNet-RS network shared by the original ResNet architecture, whilst randomly initialising any remaining additional layers.

While ResNet-RS has not yet seen widespread adoption in environmental and remote-sensing research, a number of recent studies have applied it in specific contexts, including land-cover classification (Dastour et al., 2023) and agricultural crop growth stage recognition (Yan et al., 2025).

3.5 Probabilistic Neural Networks and Uncertainty Quantification

This section provides an overview of probabilistic neural networks and uncertainty quantification methods relevant to the work presented in this thesis. It introduces the distinction between aleatoric and epistemic uncertainty, before surveying several approaches to capturing each: Gaussian likelihood regression, quantile regression, Monte Carlo dropout, ensemble methods, and deep evidential regression. The

selection and depth of coverage are motivated by the methods trialled in Chapters 6 and 7, where quantile regression and ensemble-based uncertainty quantification are ultimately employed to extract distributional elevation information from SAR altimetric waveforms. Further details on their practical implementation are provided in Chapter 6.

Neural networks, as described above, are deterministic, meaning that they always produce the same prediction for any specific input. They operate under the assumption that a single optimal set of parameters can best describe the underlying data distribution (Murphy, 2023). While these models can achieve high accuracy, they inherently lack the ability to express uncertainty in their predictions. This limitation becomes apparent in situations where data are noisy or where the model encounters out-of-distribution (OOD) examples. In real-world applications, it is often crucial to quantify the uncertainty associated with predictions (Kendall et al., 2017).

Uncertainty in machine learning predictions can be broadly categorised into two types: aleatoric and epistemic (Kendall et al., 2017). Aleatoric uncertainty refers to the inherent noise in the data, which is irreducible because it arises from the stochastic nature of the environment or measurement errors. Epistemic uncertainty, on the other hand, arises from the model's lack of knowledge due to limited training data or insufficient model capacity. This type of uncertainty can be reduced by incorporating more data or improving the model architecture (Kendall et al., 2017). Epistemic uncertainty is particularly significant when the model encounters new, unseen data or operates in regions of the input space that are sparsely populated with training examples. It is generally more challenging to estimate than aleatoric uncertainty (Smith et al., 2025).

To address the limitations of deterministic neural networks in capturing uncertainty, researchers have developed probabilistic neural networks (Murphy, 2023; Goodfellow et al., 2016). One widely used method involves training a model to learn the parameters of a Gaussian distribution - the mean (μ) and variance (σ^2)

- over the outputs given the inputs. This approach captures aleatoric uncertainty through the learned variance (σ^2). This method, often referred to as Gaussian likelihood regression, is typically implemented by minimising a loss function based on the negative log-likelihood of the Gaussian distribution (Russell et al., 2021). The network thereby converges on the distribution that best explains the observed data. This framework is flexible and is more broadly referred to as maximum likelihood estimation, where the model learns different parameters depending on the chosen likelihood function (Goodfellow et al., 2016). Assuming a Gaussian likelihood, Russell et al. (2021) outlines a negative log-likelihood loss function for the multivariate case (Equation 3.1), where the model learns multiple means (μ) and a covariance matrix (Σ), accounting for the potential correlations between prediction uncertainties. To encourage valid covariance matrices, the diagonal variance elements of Σ are parameterised with an exponential (or softplus) activation to enforce positivity, while the off-diagonal terms are constrained using the tanh function to restrict correlation parameters to (-1, 1). This parameterisation helps produce positive-definite covariance matrices, a requirement for the loss function (Russell et al., 2021).

$$L = \frac{1}{2}\big(y - f(x)\big)^\top \Sigma(x)^{-1}\big(y - f(x)\big) + \frac{1}{2}\ln\big|\Sigma(x)\big| \tag{3.1}$$

Where L is the negative log-likelihood loss function, y is the true target vector, x is the model input, and $f(x)$ represents the predicted mean vector. The diagonal and off-diagonal elements of the covariance matrix $\Sigma(x)$ are given by Equation 3.2.

$$\begin{aligned} \Sigma_{ii} &= \sigma_i^2 = g_v(s_i) = \exp(s_i) \\ \Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j &= g_\rho(r_{ij})\sqrt{g_v(s_i)g_v(s_j)} = \tanh(r_{ij})\sqrt{\exp(s_i)\exp(s_j)} \end{aligned} \tag{3.2}$$

While powerful, Gaussian likelihood regression assumes that the data is Gaussian-distributed. To overcome this assumption, alternative methods like multivariate quantile regression offer a robust approach to capturing the distributional characteristics of data without relying on strong parametric assumptions (Koenker et al.,

1978). Instead of learning μ and σ^2, this method estimates quantiles, which represent specific points in the conditional distribution of the target variable. The most widely used loss function for quantile regression, commonly referred to as Pinball loss (Equation 3.3), was first introduced by Koenker et al. (1978).

$$L_\alpha(y, \hat{y}) = \begin{cases} \alpha\,(y - \hat{y}), & \text{if } y \geq \hat{y}, \\ (1 - \alpha)\,(\hat{y} - y), & \text{if } y < \hat{y}, \end{cases} \tag{3.3}$$

Where $L_\alpha(y, \hat{y})$ is the Pinball loss function, y represents the true target value, $\hat{y}$ is the predicted quantile estimate, and $\alpha \in (0, 1)$ is the quantile level.

The Pinball loss function works by penalising overestimates and underestimates differently depending on the chosen quantile. For instance, if the quantile q = 0.9 is selected, the loss function becomes more sensitive to underestimates than overestimates, aiming for predictions that fall below the true value only 10% of the time. For q = 0.5, which corresponds to the median, the Pinball loss is symmetric and equivalent to the absolute error. More generally, when q $<$ 0.5, the loss penalises overestimations more heavily, whereas for q $>$ 0.5, it penalises underestimations more heavily. This asymmetry is important for accurately estimating different quantiles. Multiple quantiles can be learned separately, or simultaneously, summing their loss during training. Similarly, extending to the multivariate case, multiple features in the input data can be treated independently, with their corresponding losses summed or averaged.

Extending on Pinball loss, Garcia-Cardona et al. (2021) propose a more advanced multivariate quantile regression loss function using a geometric formulation. Here, each quantile is represented by a vector within the unit ball (the set of all vectors with length ≤ 1 in Euclidean space), capturing both the direction and magnitude of data points relative to the centre of their distribution. By minimising a tailored loss function (Equation 3.4), this method determines geometric quantiles that adhere to the non-crossing property, meaning that the quantile estimates maintain a logical order across multiple dimensions without intersecting. This is not accounted for

in the standard Pinball loss formulation. This approach effectively finds a specific quantile by selecting the vector within the unit ball that minimises the discrepancy between observed values and predicted quantiles.

$$L_\alpha(\xi_i) = \Phi\Big(\tfrac{1}{\sqrt{d}}(2\alpha - 1),\, \xi_i\Big) \tag{3.4}$$

Where $L_\alpha(\xi_i)$ represents the geometric multivariate quantile loss for the ith data point ξ_i, $\alpha \in (0, 1)$ is the quantile level, and $\Phi(u, t) = |t| + \langle u, t\rangle$ is the geometric quantile function. The loss is computed using the normalised direction vector $u = \frac{1}{\sqrt{d}}(2\alpha - 1)$, which indexes the desired quantile.

Although robust in not assuming the form of the underlying distribution, quantile approaches are not without limitations. Unlike parametric methods that provide a complete distributional characterisation, quantile regression only offers point estimates at specific quantiles, missing important distributional density information. As such, the distribution of probabilities within the quantiles is unknown. This approach also faces challenges in model comparison, as conventional metrics like likelihood ratios or information criteria are not directly applicable. Lastly, while parametric coefficients often have clear interpretations, the meaning of model parameters in quantile regression can be less intuitive, potentially obscuring insights into variable relationships.

While the approaches discussed allow for the prediction of aleatoric uncertainty, predicting epistemic uncertainty requires additional consideration. Often, the best performing method is to utilise an ensemble approach, wherein a model is trained N times with different parameter initialisations (Lakshminarayanan et al., 2017; Tan et al., 2023). Here, the expected prediction is given by the mean over the ensemble models, and the epistemic uncertainty by the standard deviation in these predictions (Tan et al., 2023). However, while simple and robust, this approach is often intractable due to extreme requirements in computational time during both training and inference (Thuy et al., 2024). As such, a common alternative is to employ a method known as Monte Carlo dropout, which works by applying dropout to layers

in the network at inference time (Gal et al., 2016). By performing N forward passes on the same input during inference, each time with randomly dropped out neurons, an ensemble of models is effectively approximated. As with an ensemble, the mean of these predictions provides the final output, while the variance across the predictions serves as an estimate of the model's epistemic uncertainty. Monte Carlo dropout offers significant computational efficiency compared to full ensembles, as it requires training only one model. However, this efficiency comes at the cost of less reliable uncertainty estimates, which are highly sensitive to the chosen dropout rate (Huang et al., 2025a). The approach has been found to underestimate uncertainty in out-of-distribution regions, where the model's predictions are less reliable (Liu et al., 2021; Sedlmeier et al., 2020; Ovadia et al., 2019; Detlefsen et al., 2019), and also relies on the assumption that dropout is a valid approximation of Bayesian inference, which is not always accurate (Folgoc et al., 2021).

Recently, another innovative approach to uncertainty quantification was proposed, known as deep evidential regression, as outlined by Amini et al. (2020). Unlike previous approaches outlined in estimating epistemic uncertainty, deep evidential regression bypasses the need for sampling by directly learning the parameters of the higher-order "distribution-of-distributions". This approach conceptualises training as an evidence acquisition process, where the accumulation of more evidence translates into greater predictive confidence.

In this framework, the network is trained to output the parameters of an evidential distribution, which is characterised by priors over each of the distribution parameters - such as mean (μ) and variance (σ^2) - drawn from a Normal-Inverse-Gamma (NIG) distribution. This evidential distribution encodes both aleatoric and epistemic uncertainties. The loss function in this method is designed to maximise model evidence while simultaneously penalising incorrect evidence, thereby regularising the training process (Equation 3.5). The result is a model that can directly infer prediction uncertainty without the computational burden of sampling, offering a significant efficiency advantage over methods like Monte Carlo

dropout.

$$L_i(\mathbf{w}) = L_{\text{NLL},i}(\mathbf{w}) + \lambda L_{R,i}(\mathbf{w}) \tag{3.5}$$

Where $L_{\text{NLL},i}(\mathbf{w})$ is the negative log-likelihood loss for the i-th data point (Equation 3.6), $L_{R,i}(\mathbf{w})$ is the evidence regulariser (Equation 3.7), and Ω is given by Equation 3.8. Here, $\mathbf{w}$ represents the model parameters, ν is the degrees of freedom of the Student's t-distribution, α is a shape parameter controlling the weight of the prior, β is a scale parameter of the NIG prior distribution that helps control the spread of the uncertainty estimates, y_i is the true target value for the i-th data point, γ is the location parameter (predicted mean for the i-th data point), Γ is the gamma function, and λ is the regularisation coefficient that determines the influence of the regularisation term on the total loss.

$$L_{\text{NLL},i}(\mathbf{w}) = \tfrac{1}{2}\log\left(\tfrac{\pi}{\nu}\right) - \alpha\log(\Omega) + \left(\alpha + \tfrac{1}{2}\right)\log\left((y_i-\gamma)^2\nu + \Omega\right) + \log\left(\frac{\Gamma(\alpha)}{\Gamma\left(\alpha+\frac{1}{2}\right)}\right) \tag{3.6}$$

$$L_{R,i}(\mathbf{w}) = |y_i - \gamma| \cdot (2\nu + \alpha) \tag{3.7}$$

$$\Omega = 2\beta(1+\nu). \tag{3.8}$$

Meinert et al. (2022) extends this to the multivariate case by adopting a Normal-Inverse-Wishart (NIW) prior, which generalises the NIG prior to account for correlations between multiple regression targets (Equation 3.9). Here, they also address several issues identified in the original methodology, such as degeneracy in the loss function and difficulties in accurately estimating epistemic uncertainty, especially with small data samples. However, in a later work, Meinert et al. (2023) caution that, despite the method's effectiveness, the approach should be viewed as a heuristic rather than an exact uncertainty quantification method, due to its reliance on regularisation techniques that may not yield fully reliable uncertainty estimates.

$$L_i(\mathbf{w}) = -\log L_{\mathrm{NIG},i}(\mathbf{w}) + \lambda \left(\frac{y_i - \gamma}{w_{\mathrm{St}}} \right) \cdot \sqrt{\Phi} \tag{3.9}$$

Where w_{St} is the width of the t-distribution (Equation 3.10). Here, L_{NIG} is the NIG likelihood, and Φ represents the total evidence of the inferred posterior $(2\nu + \alpha)$. This formulation normalises the residual with w_{st}, which inhibits convergence for large but insignificant residuals in terms of their associated aleatoric uncertainty, particularly when the fit has reached its optimal state.

$$w_{\mathrm{St}} = \sqrt{\frac{\beta(1+\nu)}{\alpha\nu}} \tag{3.10}$$

While this overview covers probabilistic neural networks and uncertainty quantification in general terms, the selection and level of detail provided here (e.g. Gaussian likelihood regression, quantile regression, Monte Carlo dropout, ensemble methods, and deep evidential regression) are tailored to the approaches trialled in Chapters 6 and 7, within the ResNet-RS architecture described in Section 3.4.2, where quantile regression and ensemble-based uncertainty quantification are ultimately employed to extract distributional elevation information from SAR altimetric waveforms. Chapter 6 presents further details relating to their practical implementation.

Chapter 4

Thesis Aims and Objectives

The preceding chapters have established the scientific context for ice sheet monitoring, the central role of satellite-based observations in contemporary assessments, the technical foundations of satellite radar altimetry, and the potential of data-driven methods to address persistent limitations in conventional processing. Having outlined both the capabilities and constraints of current altimetric systems, and the emerging opportunities offered by high-resolution geodetic datasets and deep learning approaches, this section now sets out the specific aims and objectives of this thesis.

4.1 Overview

The polar ice sheets are undergoing rapid change in response to contemporary climate warming, with implications for global sea level rise and regional climate dynamics (Fox-Kemper et al., 2021). Satellite radar altimetry has emerged as a cornerstone technique for monitoring ice sheet elevation and elevation change over multi-decadal timescales, providing the spatial coverage and temporal consistency necessary to track ice sheet evolution (IMBIE Team, 2018; IMBIE Team, 2020; Otosaka et al., 2023). However, significant challenges remain in extracting reliable geophysical information from radar altimetry, especially for non-interferometric in-

struments, and particularly over regions of complex topography where conventional processing assumptions break down (McMillan et al., 2018; Bamber, 1994; Simonsen et al., 2017). Ultimately, this results in reduced certainty in tracking ice sheet change, and reduced confidence in determining future climate projections.

Alongside these challenges, two key developments offer new opportunities to improve our ability to monitor ice sheets with radar altimetry. First, the emergence of new high-resolution geodetic datasets, such as the Reference Elevation Model of Antarctica (Howat et al., 2022) and ArcticDEM (Porter et al., 2023), provide unprecedented means to characterise ice sheet surfaces and evaluate instrument performance across a wider range of conditions than was previously possible. These datasets enable more systematic assessments of how current systems operate and where their limitations lie. Second, advances in data-driven methodologies, particularly deep learning, offer powerful tools for extracting more information from radar waveforms. By learning complex, nonlinear relationships directly from the data, these methods open up new possibilities for improving altimetric retrievals beyond the constraints of conventional processing. However, the application of such approaches to satellite altimetry is still in its early stages, and their relative strengths and limitations remain to be fully understood.

4.2 Aims and Objectives

This thesis aims to advance the capability of past, present and future satellite radar altimeters to monitor polar ice sheet elevation change, through the development and utilisation of new datasets and processing approaches.

To achieve this aim, three main objectives have been defined:

1. **To use new high resolution geodetic datasets to assess and understand the performance of current operational SAR altimetry systems over the Antarctic ice sheet.** This objective addresses the need to understand the fundamental constraints of existing SAR altimetry systems by

conducting a comprehensive evaluation of Sentinel-3 altimeter performance across the full range of Antarctic surface conditions. By developing new ice-sheet-wide surface characterisation datasets and applying them to assess altimeter tracking, waveform acquisition, and processing chain performance, this work helps to better establish the baseline capabilities and limitations that motivate the drive towards improved processing approaches. As such, this analysis provides important context for the subsequent development of alternative methods and identifies the topographic regimes where conventional approaches are most likely to fail.

2. **To develop, implement and validate a new probabilistic deep learning framework for extracting across-track distributional elevation information directly from non-interferometric SAR radar waveforms.** This objective directly addresses the core limitation of conventional, non-interferometric SAR altimetry processing by conceptualising, developing and testing a fundamentally different approach, which embraces rather than eliminates ambiguity in the origin of surface reflections. Using an ensemble of deep neural networks trained with quantile regression, this work demonstrates that radar waveforms contain sufficient information to predict full distributions of plausible across-track surface profiles without requiring interferometric phase data. By providing explicit quantification of both aleatoric and epistemic uncertainty, this framework offers a pathway towards robust and informative full-swath altimetric retrievals, with the potential to unlock decades of underutilised non-interferometric observations.

3. **To generate the first estimates of ice sheet elevation change using the newly-developed deep learning framework, in order to evaluate the generalisability and practical utility of this novel approach for large scale operational monitoring.** This objective aims to generalise and test the downstream application of the methodology developed in Chapter 6,

through the generation and evaluation of the first deep learning based estimates of ice sheet elevation change. Specifically, I use my deep learning model to predict surface elevations over 5 years of unseen data across parts of West Antarctica and the Greenland Ice Sheet and, in turn, to derive estimates of ice sheet elevation change over this period. By evaluating these novel elevation change retrievals against multiple independent datasets (ICESat-2, CryoSat-2 conventional processing, CryoSat-2 interferometric swath processing), this work establishes the practical utility of this approach, with a view to assessing its potential future deployment within more operational scenarios.

Ultimately, the work presented here contributes to the broader goal of improving our understanding of ice sheet evolution in a changing climate, by establishing methodological foundations that can enable more effective utilisation of radar altimetry missions in the future.

4.3 Thesis Structure

The remainder of this thesis is organised as follows. Chapter 5 evaluates the performance of Sentinel-3 SAR altimetry over the Antarctic Ice Sheet using new high-resolution geodetic datasets. Chapter 6 develops and validates a probabilistic deep learning framework for extracting distributional elevation information directly from non-interferometric radar waveforms. Chapter 7 applies this framework to generate and assess new estimates of ice sheet elevation change across Pine Island Glacier (West Antarctica) and the Greenland Ice Sheet. Finally, Chapter 8 synthesises the thesis findings, situates them within the wider literature, and outlines directions for future research.

Chapter 5

Assessment of Sentinel-3 Altimeter Performance over Antarctica using High Resolution Digital Elevation Models

Joe Phillips[1], Malcolm McMillan[1]

[1]UK Centre for Polar Observation and Modelling, Centre of Excellence in Environmental Data Science, Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence: Joe Phillips (j.phillips5@lancaster.ac.uk)

A manuscript based on the following work was submitted prior to the viva and has since been published in *The Cryosphere* (Phillips et al., 2026). The original thesis noted that this work was being prepared for submission. A reference to the published version is now included here as part of the corrections.

JP performed all code development, analysis and manuscript preparation, with supervision from MM.

Abstract

Since 2016, the Sentinel-3 satellites have provided a continuous record of ice sheet elevation and elevation change. Given the unique, operational nature of the mission, and the planned launch of two additional satellites before the end of this decade, it is important to determine the performance of the altimeter across a range of ice sheet topographic surfaces. Whilst previous studies have assessed elevation accuracy, more detailed investigations of the underlying instrument and processor performance are lacking. This study therefore examines the performance of the Sentinel-3 Synthetic Aperture Radar (SAR) altimeter over the Antarctic Ice Sheet (AIS), utilising new detailed topographic information from the Reference Elevation Model of Antarctica (REMA). Applying Singular Value Decomposition (SVD) to REMA, I firstly develop new self-consistent Antarctic surface slope and roughness datasets. I then use these datasets to assess altimeter performance across different topographic regimes, targeting a number of key steps in the altimeter processing chain. I also evaluate the impact of topography upon waveform decorrelation. For the new Sentinel-3 Thematic Product, I find that, for 94.1% of acquisitions, the point of closest approach to the satellite is successfully captured within the range window - an improvement of ∼5% compared to the previous non-thematic (BC-004) product. For both products, performance declines with increasing topographic complexity, which also limits the ability to record all backscattered energy within the beam footprint. I estimate that 57.4% of the ice sheet exhibits greater topographic variance within the footprint than can be captured by the range window, and that the current window placement captures a median of 89.2% of the total possible topography that could be recorded. These findings provide a better understanding of the performance of the Sentinel-3 altimeters over ice sheets, and can guide the design and optimisation of future satellite missions such as the Copernicus Polar Ice and Snow Topography Altimeter (CRISTAL) and the Sentinel-3 Next Generation Topography (S3NG) mission.

5.1 Introduction

From 1992 to 2020, the Antarctic Ice Sheet (AIS) lost a total of 2671 ± 530 billion tonnes of ice (Otosaka et al., 2023), with current rates of ice loss now ∼6 times greater than those measured in 1979 (Rignot et al., 2019). Recent studies suggest that this accelerating loss of ice, which has occurred mostly across sectors of West Antarctica and the Antarctic Peninsula, increased global sea levels by 7.4 ± 1.5 mm from 1992 to 2020 (Otosaka et al., 2023). As Earth's climate continues to warm, current projections suggest that the sea level contribution from the AIS will reach multiple decimetres by 2100, with high-end projections exceeding 1 m (Frederikse et al., 2020). This large range in projections, whilst partly due to a variety of forcing pathways, is also linked to uncertainty in the physical processes governing the ice sheet's response to climate change. Within this context, it is imperative to establish and maintain long-term monitoring programmes to better understand the physical processes that control ongoing ice sheet imbalance (Davis et al., 2005; Price et al., 2011; Shepherd et al., 2004). Whilst a range of techniques exist for remote observation of the cryosphere, understanding of how ice sheets are changing is largely informed by satellite observations, with the longest continuous record coming from the technique of satellite radar altimetry. This monitoring began in the early 1990s with ERS-1 and ERS-2, before progressing through dedicated missions including CryoSat-2 (2010) with ice-specific innovations, followed by Sentinel-3 (2016), which provided global SAR coverage (Abdalla et al. (2021); Chapter 2). The principal use of altimetry over ice sheets is to derive estimates of ice sheet elevation and elevation change (Wingham et al., 1998; Shepherd et al., 2019), which, ultimately, can be used to determine ice sheet mass imbalance. Radar altimeters have also been used to investigate a range of other glaciological processes, including grounding line location and migration (Dawson et al., 2017; Konrad et al., 2018; Hogg et al., 2018), subglacial hydrology (Siegfried et al., 2018; Wingham et al., 2006b; McMillan et al., 2013; Gourmelen et al., 2017), ice shelf processes (Griggs et al., 2011; Chuter et al., 2015), and surface mass balance (Slater et al., 2021).

The measurements made by radar altimeters over the past 30 years have provided important new insight into the changing nature of Earth's ice sheets. However, like all geodetic techniques, the fidelity of measurements varies, as a consequence of both the instrument design and the complexity of the illuminated ice surface. In the case of radar altimeters, measurement quality typically degrades with increasing topographic complexity, due to the difficulty of keeping track of the ice surface, the failure of assumptions used in Level-2 processing under such conditions, and the limitations of the instrument's spatial resolution. Whilst a number of previous studies (McMillan et al., 2019; Clerc et al., 2020; Quartly et al., 2020; McMillan et al., 2021; Aublanc et al., 2025b) have quantified the overall accuracy of elevation measurements derived from the Sentinel-3 altimeter, less attention has been given to investigating the underlying performance of the instrument and, ultimately, how characteristics relating to the instrument and algorithm design have affected the ability of the altimeter to make reliable elevation retrievals. With the recent availability of high resolution Digital Elevation Models, such as the Reference Elevation Model of Antarctica (REMA) (Howat et al., 2019), comes the opportunity to evaluate a number of these design choices, and therefore to advance understanding of altimeter performance. This, in turn, can benefit the design and operation of future altimeter missions. In this study, I therefore use REMA to undertake a detailed evaluation of Sentinel-3 elevation retrievals across the Antarctic Ice Sheet, in order to better understand the impact of topographic characteristics on altimeter performance.

5.2 Principles of Radar Altimetry

The broader theoretical framework of radar altimetry is introduced in Chapter 2. In this section I provide a focused overview of the elements most relevant to this study, with specific focus on the processing of non-interferometric, conventional SAR measurements.

Satellite radar altimeters transmit short microwave pulses towards the surface and record the returned echoes as discretised waveforms, representing the aggregate of reflections from all scattering points within the footprint, ordered by arrival time (Brown, 1977). Over flat surfaces, the waveform exhibits a sharp leading edge associated with the point of closest approach (POCA) to the satellite, followed by a trailing decay shaped by the antenna gain pattern and surface geometry. In complex Antarctic terrain, however, echoes often deviate from this simple form, complicating the retrieval of elevation.

To ensure that these echoes are successfully captured, the altimeter must record the incoming energy within a finite sampling window centred on the predicted surface return. Since this window is only a few microseconds wide, it must coincide with the expected return time from the surface. If it does not, the altimeter can temporarily lose track of the Earth's surface. To avoid this, the instrument relies on an estimate of the predicted range. In closed-loop tracking, this estimate is derived from the recent history of ranges, which performs well when the surface evolves smoothly, such as over the ocean or the low-slope interior of the ice sheet. However, performance degrades in more variable terrain, where past behaviour is a poor predictor of future returns. Open-loop tracking provides an alternative by using an a priori digital elevation model to guide the placement of the window (Donlon et al., 2012b). This approach has been shown to improve tracking over regions of complex topography, including ice sheet margins. Open loop tracking was first applied onboard the Jason-2 mission, which over land utilised the ESA ACE-1 DEM at 1 km resolution (Birkett et al., 2010). It has subsequently been used by Sentinel-3 during commissioning and ad hoc tasking over land ice surfaces, and is planned to be implemented operationally onboard CRISTAL. However, a key practical consideration for open-loop tracking is the temporal validity of the a priori DEM used to predict the surface range, which over certain areas may need occasional updates. In stable regions of the ice sheet interior, elevation changes are sufficiently slow (typically millimetres to centimetres per year) that a DEM acquired years prior remains adequate. However,

in dynamically active regions - such as fast-flowing outlet glaciers where thinning rates can exceed several metres per year - the DEM may become outdated, leading to systematic range prediction errors.

Once a waveform has been successfully acquired, conventional, non-interferometric processing works to reduce it to a geophysically meaningful measurement of surface elevation. The central challenge is that when using a single antenna, a single waveform can, in principle, be explained by many different surface configurations within the footprint, and additional assumptions are therefore required to identify the reflection point. The conventional approach, which has remained broadly consistent over the past three decades, addresses this through two core operations: retracking, to derive a one-dimensional range estimate from the waveform, and slope correction, to locate the corresponding point of closest approach (POCA) on the surface. These steps are followed by the application of geophysical and instrumental corrections, which remove signal delays and biases introduced by the atmosphere, tides, and the satellite system itself.

Retracking determines the position of the leading edge relative to the initial tracking gate - corresponding to the measured satellite-to-surface range - and can involve adjustments on the order of several meters in height (Martin et al., 1983; Ridley et al., 1988). Retrackers are broadly divided into physical methods, which fit analytical models to the waveform (Villadsen et al., 2016; Passaro et al., 2022) (e.g. Ice-2 for ENVISAT (Legresy et al., 2005), SAMOSA for CryoSat-2 and Sentinel-3 (Ray et al., 2015), and the UCL ice-sheet retracker in the Sentinel-3 land-ice ground segment (Wingham et al., 2006b), and empirical methods, which instead use waveform geometry (e.g. ICE-1, which applies a threshold based upon the OCOG (Offset Centre of Gravity) amplitude (Wingham et al., 1986; Bamber, 1994), threshold schemes such as TFMRA (Helm et al., 2014; Davis, 1997), and recent multi-peaked waveform approaches like Multi-Peak Ice (Huang et al., 2024). Because they are less sensitive to irregular, multi-peaked echoes, empirical methods are generally more robust over ice sheets (Quartly et al., 2020; Martin et al.,

1983; Quartly et al., 2019; Landy et al., 2019; Slater et al., 2019). In operational processing, Sentinel-3 employs both the UCL model-fit retracker and a threshold OCOG retracker.

Slope correction locates the POCA, to which the re-tracked elevation measurement is attributed. Early slope-based methods assumed a constant surface gradient across the footprint and solved the geometry trigonometrically, but these simplifications introduced large errors in complex terrain (Remy et al., 1989; Cooper, 1989; Bamber, 1994; Brenner et al., 1983; Levinsen et al., 2016; Roemer et al., 2007). Modern, point-based methods improved on this by searching a DEM for the location of minimum range within the footprint and have been found to reduce bias compared with slope-based approaches (Roemer et al., 2007; Levinsen et al., 2016; Li et al., 2022). More recent point-based refinements include the LEPTA algorithm (Li et al., 2022), which exploits the range spanned by the waveform leading edge to identify DEM points contributing to the echo, and has been extended to retrieve multiple elevations from multi-peaked Sentinel-3 SAR waveforms (Huang et al., 2024). Another development, AMPLI, applies a facet-based waveform simulation and has shown substantial improvements over slope-based corrections in sloped Antarctic terrain (Aublanc et al., 2025a). In steep terrain, slope corrections can reach tens of meters vertically and kilometers across-track, making them commonly the largest source of uncertainty in non-interferometric elevation estimates (Li et al., 2022; Brenner et al., 2007).

Following measurement extraction (retracking) and attribution to the POCA (slope correction), the measurements are adjusted for geophysical and instrumental effects. These include ionospheric and tropospheric delays, solid Earth and ocean tides, and polar loading, along with instrument-specific effects such as oscillator drift and antenna-centre of mass offsets (Quartly et al., 2020). Although individually small, these corrections can sum to several metres and are therefore essential to isolate the geophysical signal of interest.

Together, retracking, slope correction, and geophysical adjustment form the

core of the conventional altimetric framework. While more advanced techniques, such as interferometry, have expanded capability in recent years, this conventional processing chain remains the foundation of operational ice-sheet monitoring.

5.3 The Sentinel-3 Mission

This study focuses on the performance of the Sentinel-3 radar altimeter. Here, I summarise the principal characteristics of the mission. More comprehensive details are provided in Chapter 2.

5.3.1 Mission and Instrument Overview

The Copernicus Sentinel-3 constellation forms part of the EU's operational programme to extend Europe's multi-decadal altimetry record. Two satellites, Sentinel-3A (launched 2016) and 3B (2018), are currently in orbit, with 3C and 3D planned to ensure continuity into the next decade. Sentinel-3 follows a 27-day repeat orbit at 98.65° inclination, providing global coverage to 81.35° latitude, with quasi-global coverage achieved every four days (Quartly et al., 2020).

The mission was designed as a successor to ERS-1/2 and Envisat, but with significant enhancements. Most notably, the primary altimetry instrument SRAL (Synthetic Aperture Radar Altimeter) - which is a Ku-band SAR altimeter - operates in unfocused SAR mode across the entire ice sheets, providing ∼300 m along-track resolution - a four-fold improvement over earlier pulse-limited instruments (Quartly et al., 2020). However, the smaller ∼60 m range window (EUMETSAT, 2017) means echoes are sometimes missed in regions of steep or complex topography.

During routine operations, SRAL runs in closed-loop tracking, mirroring previous ESA missions. Although Sentinel-3 was designed with open-loop capability, initial tests showed the on-board OLTC tables were insufficient over the ice sheet margins (SentiWiki, n.d.). These tables have since been updated using high-resolution REMA and ArcticDEM data, and new acquisitions over specific areas are now being

evaluated to improve performance.

Accompanying SRAL, Sentinel-3 carries several additional sensors - the MWR, SLSTR, and OLCI - enabling synergistic measurements of surface properties (Clerc et al., 2020). Together, these instruments make Sentinel-3 well-suited for operational monitoring of the cryosphere, while also providing continuity with, and improvements upon, CryoSat-2 and earlier ESA altimetry missions.

5.4 Data

The objective of this study is to provide a detailed assessment of the impact of surface topography on the performance of SRAL across the AIS. For the purposes of this study, I therefore analysed one complete cycle of data acquired by both Sentinel-3A and Sentinel-3B. Specifically, I selected cycle 54 for S3A and cycle 35 for S3B, which were acquired over the periods 15$^{\text{th}}$ January to 11$^{\text{th}}$ February 2020 and 25$^{\text{th}}$ January to 22$^{\text{ned}}$ February 2020, respectively. As my primary dataset, I used the Sentinel-3 Hydro-Cryo Altimetry Thematic Baseline Collection 005 Land Ice product (S3A_SR_2_LAN_LI and S3B_SR_2_LAN_LI; available at `https://dataspace.copernicus.eu/`). This product builds on the previous, unified BC-004 "Land Product", with a specific land ice chain that makes up one of three new families of "Thematic Products" (Aublanc et al., 2025b). The BC-005 land ice product includes a number of dedicated ice sheet processing algorithms, the most notable of which involves artificially extending the range dimension of the delay-Doppler stack during Level-1 range migration and multilooking, in order to better preserve backscattered energy in the final waveform, particularly over the ice sheet margins (Aublanc et al., 2025b; Aublanc et al., 2018). Further details relating to the thematic product can be found in the Sentinel-3 SRAL Land User Handbook (Aublanc et al., 2024). Additionally, for comparison purposes, I also utilise the BC-004 version product over the same cycles (also available at `https://dataspace.copernicus.eu/`).

In order to investigate the performance of Sentinel-3, I used the 100 m resolution, version-2 REMA (Reference Elevation Model of Antarctica) mosaic (Howat et al., 2022). REMA is a high-resolution DEM covering nearly 98% of Antarctica, which has been used for a range of cryospheric applications (Chartrand et al., 2020; Zinck et al., 2023; Liu et al., 2023). REMA was created by applying stereophotogrammetric techniques to submeter resolution commercial optical satellite imagery, including data from Maxar, WorldView-1, WorldView-2, and WorldView-3, as well as a small number of acquisitions made by GeoEye-1 (Howat et al., 2019). To form the mosaiced product, individual DEMs were registered to satellite altimetry measurements from CryoSat-2 and ICESat, leading to an estimated absolute uncertainty of less than 1 m, and relative uncertainties of the order of decimeters (Howat et al., 2019). For the purposes of this study, I selected REMA over other DEMs, such as the ICESat-2-derived gridded ATL14 product (Smith et al., 2021), as it is generated from stereoscopic techniques, and thereby provides a high-resolution, inherently two-dimensional product with minimal gaps. In contrast, the ATL14 product includes interpolation between ICESat-2 tracks, particularly as track-to-track spacing increases away from the poles (Smith, 2021), which makes it less well suited to constraining small-scale topographic characteristics, such as footprint-scale surface roughness.

5.5 Methodology

In this study, my aim is to assess and better understand the performance of the S3 altimeter over a variety of topographic regimes. To achieve this, I firstly use REMA to generate new ice sheet wide estimates of surface slope and roughness. I then performed three different sets of analysis (1) an evaluation of the extent to which the onboard tracker's placement of the range window and subsequent processing maximises the proportion of the illuminated topography that contributes to each waveform (i.e. that is captured by the range window), (2) an assessment of the

extent to which the placement of the range window successfully captures the point of closest approach on the illuminated ice surface, and (3) quantification of the degree of correlation between along-track sequences of echoes. The first analysis is designed to provide a first assessment of the effectiveness of the altimeter in adequately tracking the underlying topographic surface. The second analysis is designed to evaluate the assumption - which is inherent to most L2 processing schemes - that the waveform leading edge originates from the closest point within the beam footprint; and then to investigate how the validity of this assumption varies with increasing topographic complexity. The third analysis evaluates the extent to which the complexity of the topographic surface - characterised by slope and roughness - causes decorrelation in successive delay-Doppler echoes. In the following sections, I outline the methodology used for each of these analyses, in turn.

5.5.1 Ice Sheet Surface Slope and Roughness from REMA

To understand the impact of ice sheet surface topography on the performance of the Sentinel-3 altimeter, I used the 100 m resolution REMA dataset to compute 100 m resolution, continent-wide estimates of ice sheet surface slope and roughness. Within the context of this study, these parameters are important because they can affect (1) the ability of the altimeter to track the ice sheet surface, (2) the effectiveness of approaches to derive the waveform, and (3) the degree of correlation between sequences of waveforms acquired along the satellite track.

I generated my own slope and roughness maps using REMA in order to tailor the algorithms used to fit with the objectives of this study. Here, I am primarily interested in how surface topography varies within the beam-limited footprint, as this directly governs the complexity of the returned waveform and the instrument's ability to capture backscattered energy. Given that the pulse-limited footprint - which broadly defines the altimeter's effective surface resolution - is on the kilometre scale, roughness in this context refers to surface variability at this same macro scale. Although there exist several easy-to-apply methodologies for calculating slope

from DEMs (Horn et al., 1981; Zevenbergen et al., 1987), common approaches for calculating roughness such as the Terrain Ruggedness Index (TRI) (Riley et al., 1999), the Topographic Position Index (TPI) and Roughness (Wilson et al., 2007), do not fit my use-case, despite extensive application in GIS (Geographic Information System) programs and packages such as GRASS, ArcGIS, and GDAL (used by QGIS). This is because these methods use the local variance in elevation as a measure of roughness, but neglect to account for topographic slope. Instead, they calculate the differences between a pixel and its immediate neighbours, which would, for example, return a non-zero roughness value over a perfectly smooth plane inclined at an angle. Consequently, the roughness values produced by these methods are dependent on slope, making it difficult to separately assess their individual contributions to topographic variation. More advanced metrics such as the Vector Ruggedness Measure (VRM) (Sappington et al., 2007) attempt to address this through calculating the dispersion of local surface normals, but (1) do not explicitly decouple slope from the roughness calculation, and (2) express roughness as a unitless 0-1 value, which neglects recording useful elevation amplitude information. To address these prior limitations in slope-roughness coupling explicitly, I therefore calculate roughness by quantifying the dispersion of orthogonal residuals from a best-fit plane, implemented using Singular Value Decomposition (SVD), and applied over a sliding-window. Specifically, after fitting a best-fit plane to the point cloud within each window via SVD, the orthogonal (perpendicular) distance of each elevation point from that plane is computed by taking the dot product of each mean-centred coordinate vector with the unit normal of the fitted plane. The roughness is then defined as the peak-to-peak range of these residuals - that is, the difference between the maximum and minimum orthogonal distances. This captures the amplitude of surface undulation relative to the local slope, making it geometrically meaningful: a perfectly planar inclined surface will return a roughness of zero, whereas a surface with metre-scale undulations superimposed on a slope will return a roughness value reflecting those undulations alone. In addition to

eliminating any artificial correlation between slope and roughness that is introduced by other methods, this approach also provides a more accurate measure of slope than commonly used approximations such as Horn's method (Horn et al., 1981) by explicitly fitting a plane, whilst also allowing for flexible window sizes beyond the 3×3 pixel constraint inherent to those approaches. To fit the plane, SVD was chosen, rather than a least-squares (LS) approach, as SVD fits a plane by minimising orthogonal residuals, whereas LS minimises residuals in the vertical (z) direction. This, I believe, makes SVD more suitable for fitting surfaces over complex terrain, where minimising the true geometric (orthogonal) distance is expected to yield a more representative estimate of the surface slope. Furthermore, SVD directly returns the normal vector to the fitted plane during the fitting process, with LS otherwise requiring a separate, additional calculation. Alternative approaches, such as wavelet decomposition, could in principle provide a multi-scale characterisation of surface variability and may offer advantages in separating topographic signal at different spatial frequencies; however, this was beyond the scope of the present study, which required a computationally efficient, spatially consistent approach applicable at continental scale. For the window size, I used 9×9 pixels (900×900 m), with slope and roughness parameters only computed where more than half of the pixels within a window contained a valid elevation measurement. This window size was chosen to be broadly representative of the surface geometry seen by the instrument: whilst the SAR processing achieves an along-track resolution of ∼300 m, the pulse-limited footprint - which more closely reflects the spatial scale over which the altimeter integrates surface backscatter - is approximately 1600 m in the across-track direction. A 900×900 m window therefore provides a practical approximation at a scale consistent with this effective sensing resolution, whilst also serving to minimise the impact of missing data in the REMA mosaic. It should be noted that this algorithmic decoupling assumes a meaningful scale separation between the planar trend (slope) and the residual variability (roughness) within the chosen window. Over surfaces with fractal or self-affine properties, no such

natural separation exists, and the distinction between slope and roughness becomes window-size dependent - an inherent limitation of any fixed-window approach. This window-size dependency is explored briefly in Section 5.6.1, where results from 3×3 (e.g. Horn et al. (1981)) and 9×9 pixel windows are compared.

To obtain slope and roughness values, I first construct a matrix $\mathbf{M} \in \mathbb{R}^{n\times 3}$, containing the n mean-centered 100 m resolution REMA coordinates (x, y, z) within each window (Equation 5.1). I then apply SVD to this matrix, as defined in Equation 5.2, which decomposes $\mathbf{M}$ into the product of three matrices: an orthogonal matrix $\mathbf{U}$ containing the left singular vectors, a diagonal matrix $\mathbf{\Sigma}$ containing the singular values in decreasing order, and an orthogonal matrix $\mathbf{V}$ whose columns are the right singular vectors. In this decomposition, $\mathbf{V}$ encodes the principal directions of variation in the point cloud, with its columns ordered by decreasing variance, whilst $\mathbf{U}$ and $\mathbf{\Sigma}$ together describe how the data projects onto these directions. The direction of least variation - given by the last row of $\mathbf{V}^T$, corresponding to the smallest singular value in $\mathbf{\Sigma}$ - defines the unit normal vector $\mathbf{n} = [n_x, n_y, n_z]$ of the best-fit plane through the points in the window. The surface slope is then derived from the gradient of this fitted plane, given by the partial derivatives $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ (Equation 5.3), with a final, singular slope value s computed by taking the Euclidean norm of the two directional gradients and converting to degrees. This yields the local surface slope angle, independent of orientation (Equation 5.4). Roughness r is calculated as the minimum-to-maximum range of the orthogonal residuals, obtained by taking the dot product of each centered point with the plane normal, where $\mathbf{m}_i$ is the i-th row of $\mathbf{M}$ (Equation 5.5). To obtain single slope and roughness values for each record, I then take the mean of the pixel values contained within the corresponding beam-limited footprints.

$$\mathbf{M} = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{bmatrix} \tag{5.1}$$

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{5.2}$$

$$a = -\frac{n_x}{n_z}, \quad b = -\frac{n_y}{n_z} \tag{5.3}$$

$$s = \arctan\left(\sqrt{a^2 + b^2}\right) \cdot \left(\frac{180}{\pi}\right) \tag{5.4}$$

$$r = \max_i(\mathbf{m}_i \cdot \mathbf{n}) - \min_i(\mathbf{m}_i \cdot \mathbf{n}) \tag{5.5}$$

To facilitate comparison with established methods, I also generate a slope map using Horn's method (Horn et al., 1981), applied to the same 100 m resolution REMA dataset. Since Horn's method uses a 3×3 pixel window (300x300 m), I re-compute my slope map using the same window size for a fair comparison. Similarly, I compute roughness maps using TRI and my method, again using a 3×3 pixel window to ensure consistency with TRI. Here, TRI was chosen over alternatives such as TPI and Roughness due to its more extensive prior application in terrain analysis (Trevisani et al., 2023).

The code for generating slope and roughness maps via the SVD approach outlined above can be found at `https://github.com/Joe-Phillips/Slope-and-Roughness-from-DEMs`.

5.5.2 Window Placement Optimisation

Prior to performing a detailed assessment of the range window placement relative to the point of closest approach, I first investigated more broadly the extent to which the placement of the range window maximises the capture of the topographic surface illuminated by the beam-limited footprint. My motivation for this analysis is grounded on the principle that maximising the amount of information captured is beneficial for a range of present and future use cases. Specifically, whilst it is clear that current non-interferometric processing methods largely rely on information

extracted from the leading edge of the waveform, it is likely that future advances in retrieval algorithms may allow full-waveform retrievals. Thus, there is benefit to assessing how the size and placement of the range window can be optimised in order to capture as much of the illuminated surface as possible.

To perform this analysis in practice, for each echo within my chosen cycles (comprising a total of 8,125,872 records; where each record corresponds to a single 20 Hz altimetric echo), I computed the proportion of DEM points (100 m resolution) within the beam-limited footprint that fell above, within and below the ∼60 m range window, where the beam-limited footprint was defined with dimensions of 18,200 and 300 m in the across and along-track directions, respectively. It should be noted that the footprint dimensions used here (18,200 × 300 m) represent a fixed geometric approximation. In practice, the effective footprint can depend on conditions such as surface roughness, slope, and satellite altitude. However, these fixed dimensions are well defined in the literature (Donlon et al., 2012a), and simplify subsequent inter-record comparisons.

Because over complex topography, the range of elevations within the beam-limited footprint exceeds that spanned by the range window, it is, in places, impossible for the range window to entirely capture the illuminated topographic surface. To account for this, I therefore also computed the hypothetical positioning of the range window that would have maximised the capture of topography, together with the associated percentage of the illuminated surface that could be captured within it. This was done by iteratively moving the range window by 1-meter increments in range, to determine the position that maximised topographic capture. Having located the optimal range window placement, I were then able to quantify the possible increase in the proportion of the surface captured, which could be realised with a refined range window placement, whilst respecting the inherent limitations imposed by the 60 m range window size. This was assessed by computing the ratio of actual to maximal topographic capture. Finally, I also computed the optimal range window position - and thus the maximum possible topographic capture - that could

be achieved for varying-sized range windows; specifically: 60 m, 120 m, 180 m, 300 m, and 360 m. This allowed me to explore the extent to which an increased range window size would, in principle, yield greater capture of the illuminated topographic surface, with a view to improving broader understanding within the context of other current and future satellite altimetry missions overflying Antarctica. Notably, these instrument choices (range window dimensions) are fixed for each mission, and therefore findings here should be placed in the context of understanding limitations in current and historic missions, and helping inform choices made for future missions.

5.5.3 Capture of the Point of Closest Approach

Following the methodology presented in Section 5.5.1, I also performed a similar assessment, which focused specifically on evaluating the window position relative to the point of closest approach. This is motivated by the fact that, within conventional L2 altimetry processing chains, an assumption is made that the echoing point (the so-called POCA) lies within the intersection of the range window and the beam-limited footprint (class 2 in Figure 5.1); and thus that the return from the POCA location corresponds to the range to the waveform leading edge. However, around the complex topography of the ice sheet margin, this assumption does not always hold, particularly for Sentinel-3, due to its relatively small ($\sim$60 m) range window. In regions of high slope, the POCA may lie outside the beam-limited footprint, meaning that any backscattered energy will be highly attenuated by the antenna gain pattern, whereas in regions of rugged (highly variable) topography, the POCA may lie at an elevation not captured by the range window. With the availability of new, high-resolution DEMs such as REMA, comes the opportunity to explicitly assess the extent to which the echoing point lies within the intersection of the beam-limited footprint and the range window, and hence the extent to which this common assumption in L2 processing chains is satisfied. Furthermore, by comparing success and failure at the ice sheet scale to covariates such as surface roughness and slope, I aim to assess how performance is affected by topographic complexity. For

all S3A and S3B echoes within my chosen cycles, I therefore classified where the identified POCA location lay relative to the positioning of the S3 range window. The acquisition geometry and related definition of the POCA classes are shown in Figure 5.1.

Within this assessment, I performed two sets of analyses. First, I evaluated the positioning of the range window relative to the POCA location that was defined within the standard L2 product. Here, slope correction - and consequently the determination of the location of the POCA within the SAR footprint - is performed via a slope-based approach using a surface slope model precomputed from DEMs over Antarctica and Greenland (Helm et al., 2014; Aublanc et al., 2024; IPF, 2024). This analysis allows me to determine the extent to which the slope correction applied within the L2 processing chain identifies an echoing point that is consistent with the actual acquisition geometry (i.e. lies within both the beam-limited footprint and the range window); in other words, whether the subsequent elevation measurement derived by retracking the waveform can reasonably be attributed to the assigned POCA location. Secondly, I performed the same evaluation, but for a POCA location that I determined, independently, by minimising the range between the ice surface and the altimeter, using the REMA DEM. This procedure is analogous to the methodology underlying the point-based relocation approach first introduced by Roemer et al., 2007, albeit using an extended search region to include instances where the POCA lies beyond the 3 dB beamwidth (Figure 5.1). I refer to the identified surface location as the REMA-derived POCA. It is important to note that although the underlying algorithm is similar, the purpose of this analysis is not to perform a slope correction (i.e. the intention is not to derive relocated elevation measurements as part of a standard L2 processing chain). Rather it is simply to determine the location of the closest surface point to the satellite (as determined by REMA), thereby allowing me to identify the extent to which the positioned range window captures this POCA location.

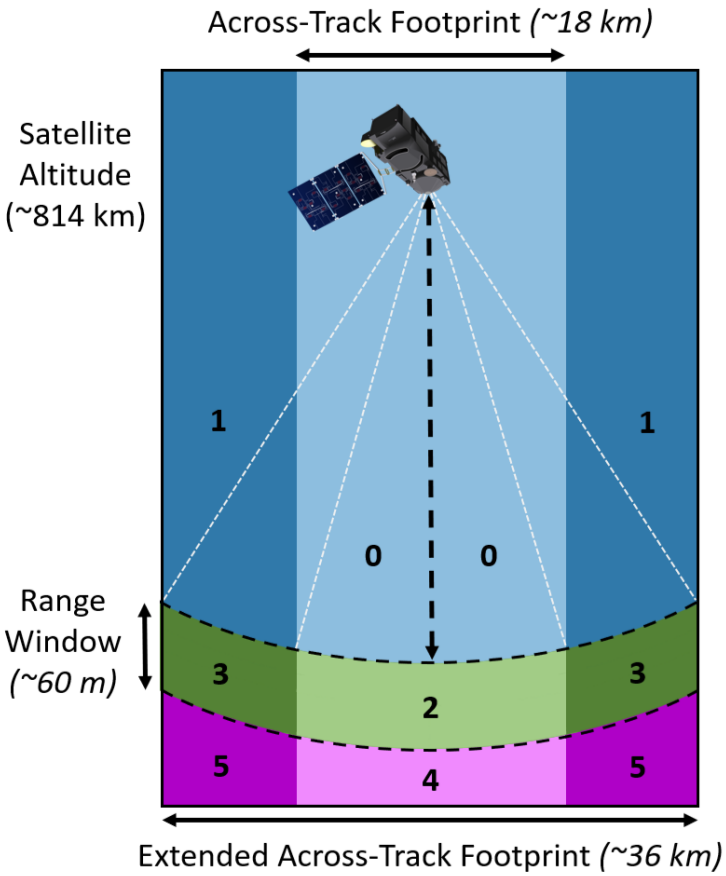

Figure 5.1: Illustration of the Sentinel-3 SAR altimetry acquisition geometry in the across-track plane (satellite flying into the page), showing the classification scheme used to categorise each acquisition according to the location of the POCA relative to the positioning of the range window. Above the range window is shaded in blue (classes 0, 1), below in magenta (classes 3, 4), and within the range window in green (classes 2, 3). Darker shades of each colour represent locations outside of the across-track beamwidth.

To perform these analyses, I first located the range window relative to the ice surface for each of the ∼8 million altimeter records considered in this study, and then compared it to the two POCA locations (the first derived from the L2 product, and the second independently determined by minimising the range between the satellite and the REMA surface). Specifically, I adjusted the range to the nominal tracking point (defined for S3 as the 43rd bin in the range window) according to the geophysical and instrument corrections supplied within the product, to account for path delays due to dry tropospheric, wet tropospheric and ionospheric effects;

tidal and atmospheric pressure induced variations, and the instrument Centre of Gravity (COG) correction. As all of these corrections are supplied at 1 Hz, I used linear interpolation to match these to the native 20 Hz frequency of the range and elevation measurements. Finally, the dimension of the range window (128 bins) was used to locate the range to the start and end of the window for each record. To reference the range window to the DEM elevation, I then subtracted the satellite altitude.

Next, I computed the geographic coordinates and elevation of the POCA for both the L2 and REMA derivations, in order to classify them according to their position relative to the range window (Figure 5.1). For the L2 evaluation, the geographic coordinates (x, y) were simply taken from the product. For the surface elevation (z) at the L2 POCA, I chose to extract the elevation at this location from the closest REMA pixel, rather than use the elevation within the altimeter product, because the purpose was simply to compare the range window to a reference surface, rather than interpret the altimeter elevation itself. This approach therefore avoided missing data in the L2 product. I selected the nearest REMA point, rather than applying a more complex interpolation method, as the 100 m posting is already finer than the altimeter resolution ($\sim$1600 x 300 m). In this regard, my aim is a coarse-scale classification of the POCA location relative to the range window, rather than the derivation of relocated elevation measurements themselves, and so a more detailed interpolation was unnecessary. For the REMA derivation of the POCA, I computed the echoing point location based upon the DEM pixel that minimised the range between the satellite and the ice surface.

5.5.4 Along-track Decorrelation of Echoes

Finally, I explore the extent to which ice sheet surface topography causes decorrelation in sequences of SAR waveforms, as they are acquired along the satellite track. This analysis extends the work of McMillan et al. (2021), to investigate the rate at which waveforms decorrelate in space (along-track), and the extent to

which this is governed by the topographic characteristics (slope and roughness) of the surface. More specifically, I calculate (1) correlations between adjacent pairs of waveforms aligned in range, and (2) correlations between every waveform and the 50 subsequent waveforms along-track (spanning a distance of ∼15 km), also aligned in range. The latter assessment allows me to quantify the rate at which waveforms decorrelate in space, spanning distances that are representative of those over which altimetry measurements are commonly aggregated, for example when computing rates of surface elevation change. The overall intent of this analysis is to assess how the topographic characteristics of present day ice sheets influences the spatial decorrelation of SAR waveforms, which is motivated by the hypothesis that topographically complex areas drive a more rapid decline in correlation between a given waveform and those acquired subsequently along-track. At a practical level, understanding this relationship is of interest, as decorrelation characteristics may influence the effectiveness of combining information from waveforms acquired over different spatial scales. Specifically, if waveforms decorrelate rapidly due to surface complexity, the assumption that nearby echoes sample a similar surface - which underpins many along-track averaging and elevation change estimation strategies - may not hold, with implications for the reliability of elevation retrievals in topographically complex regions. Ultimately, by examining these relationships, I aim to provide initial insights into how surface morphology impacts waveform structure, which can help to inform future work towards improved strategies for processing SAR altimetry data in areas of complex topography.

To determine the correlation between any two waveforms, I first normalise the power of each to be between 0 and 1. To avoid cases where a given waveform was excessively noisy, correlations were not computed when the mean power did not exceed the estimated thermal noise by more than 0.15, with thermal noise computed as the mean of the 6 lowest-value waveform samples (Helm et al., 2014; McMillan et al., 2019). Next, I aligned both waveforms according to their centres of gravity (COG) according to Equation 5.6, where N is the number of waveform bins (128),

and P_n is the power at bin n. Following this, I retained only the overlapping region of the two waveforms, removing any bins that did not intersect.

In order to focus on correlating the principal surface response, rather than peripheral signals and waveform noise, I masked the aligned waveforms by removing peripheral bins from each side of the range window, until a bin with power above 0.15 (chosen for consistency with the previous noise detection) was found in either waveform. If the clipped waveforms had at least 16 bins remaining, I then computed the Pearson correlation coefficient using the pairs of normalised waveform samples. A minimum threshold of 16 bins was chosen as a reasonable lower bound for what constitutes a valid response, based upon an assessment of the waveforms that passed the L2 ice quality checks, which retained a mean COG width of 27.1 bins and standard deviation of 10.2.

$$\mathrm{COG} = \frac{\sum_{n=1}^{N} n P_n^2}{\sum_{n=1}^{N} P_n^2} \tag{5.6}$$

Next, for each waveform, I computed its Pearson correlation with each of the subsequent 50 waveforms along-track, yielding a series of correlation values relative to the initial waveform. If more than 25 correlation values were available, I fitted a linear regression to estimate the rate of change in correlation as a function of along-track distance. To ensure comparability between all regressions, I explicitly included a correlation value of 1 at $d = 0$, representing the waveform's self-correlation, and constrained the regression to pass through this point, such that the fitted line followed Equation 5.7, where d is the number of records along-track, $R(d)$ is the Pearson correlation coefficient, and m is the along-track rate of change in correlation, per 20 Hz measurement. This slope captures the linear decay in correlation from the initial waveform to increasingly distant waveforms along-track, computed over 50 subsequent records ($\sim$15 km).

$$R(d) = 1 + m \cdot d \tag{5.7}$$

Using the fitted slope, I computed the change in correlation at record 50. This provides a more stable estimate of the change in waveform correlation after 15 km, which is not solely dependent upon the initial and 50th records. Based on this approach, a value of 0 indicates that waveforms 15 km away are perfectly correlated, whilst a value of -1 corresponds to waveforms that are fully decorrelated (R=0) after 15 km. Values less than -1 indicate that waveforms are anti-correlated after 15 km; i.e. an increase in power in one waveform corresponds to a decrease in power in the other. This is likely to reflect limitations in the COG alignment of complex waveforms (e.g. correlating a single peak with a double peak waveform) and so I focus my analysis on correlation values between -1 and 0.

To analyse the effect of surface topography on these metrics, I then derived estimates of across-track surface slope and roughness for a 1-D across-track profile approximating the beam footprint of each record, using a 2-D version of the SVD approach described in Section 5.5.1. I intentionally focused this assessment of waveform similarity on across track slope and roughness, because the anisotropic shape of the delay-doppler footprint means that these parameters in the across-track direction are most likely to have the largest influence on waveform shape.

5.6 Results

5.6.1 Ice Sheet Surface Slope and Roughness from REMA

First, I analysed surface slope and roughness over the Antarctic Ice Sheet using REMA-derived datasets (Figure 5.2), computed at 100 m resolution over a 900×900 m sliding window. This analysis examines their ice sheet-wide characteristics, the influence of window size, and consistency with established approaches, without additional filtering or co-location with Sentinel-3 observations. In later sections, where slope and roughness are used as covariates to assess altimeter performance, values are instead extracted at individual Sentinel-3 record locations, representing the specific topographic conditions sampled by the altimeter rather than the ice

sheet as a whole.

My new surface slope map shows well-established patterns of very low ($< 0.1°$) surface slope at the ice divides and the floating ice shelves, with steeper slopes prevailing towards the grounded ice sheet margin. Overall, I find that 76.9% and 89.1% of the ice sheet have a surface slope below 0.5 and 1.0°, respectively. In addition to the large-scale characteristics of slope, my new map also shows detailed signatures of topography and ice dynamics, such as the flow of ice across the Ross and Filchner-Ronne Ice Shelves from their upstream tributaries, and the numerous ice rises that lie close to these ice shelves' grounding lines. Turning to ice sheet surface roughness, I find that 62.4% and 76.1% of the ice sheet have a surface roughness below 0.5 and 1.0 m, respectively, with lower roughness persisting mostly within the interior of the ice sheet. Of particular note are the clear signatures of mega dune fields in East Antarctica, which typically have roughness of the order of $\sim$1 m and the patterns associated with ice flow across the Ross and Filchner-Ronne Ice Shelves. In broad terms, the interior of West Antarctica is rougher than East Antarctica and, as with surface slope, there is a general trend to increasing roughness towards the ice margin, with surface roughness often reaching several meters close to the coast.

I find that the ice sheet as a whole has a median slope and roughness of 0.192° and of 0.297 m, respectively. I characterise the degree of spatial variability of each parameter across the ice sheet by computing the median absolute deviation, which is 0.134° and 0.219 m for slope and roughness, respectively. To explore in more detail the relationship between surface slope and roughness, I plot their joint and marginal distributions (Figure 5.3), which shows a positive relationship between the two variables. Taking the Pearson correlation coefficient, I obtain a value of 0.808, indicating that the two values are significantly linearly correlated. Because I have calculated roughness via the dispersion of orthogonal residuals to the fitted slope, the two variables remain algorithmically independent. Therefore, this correlation reflects a physical relationship between these topographic glaciological parameters, indicating that areas with higher surface slope are also more likely to exhibit higher

surface roughness. Broadly, I find that roughness increases by ∼4.38 m per degree increase in surface slope.

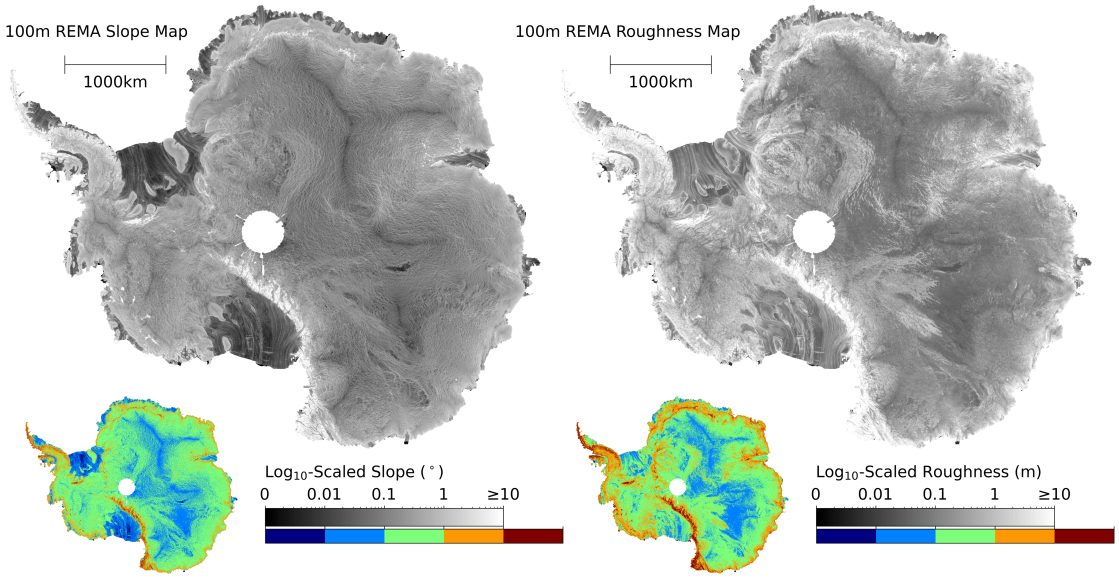

Figure 5.2: Log_{10}-scaled, 100 m resolution slope and roughness maps of Antarctica generated using REMA and the Singular Value Decomposition techniques outlined in Section 5.5.1. The coloured inset maps show discretely binned versions of these datasets, to aid the visualisation of broader scale patterns.

As outlined in Section 5.5.1, I also compare my slope estimates to those calculated using Horn's method (Horn et al., 1981). This comparison includes both the slope map shown in Figure 5.3, which was generated using a 9×9 pixel window, and an alternative version that used a window size of 3×3 pixels (300×300 m) to match the window dimensions of Horn's method (Figure 5.4 a). The three results show very close agreement, all attaining the same Gaussian shape on a $\log_{10}$ scale. Specifically, the slope maps derived using Horn's method and my SVD-based approach (both with 3×3 pixel windows) yield median values of 0.199° and 0.198°, respectively, and identical MADs of 0.138°. The difference between my 3×3 and 9×9 pixel implementations is also negligible (median difference $< 0.01°$). These results indicate that, when assessed at the scale of the entire ice sheet, and at a resolution of 100 m, neither the choice of window size (3×3 versus 9×9) nor the underlying slope

estimation method has a significant impact. It is important to note, however, that the SVD approach is more flexible in offering a generalised solution that can be applied to variable and irregular window shapes, making it better suited for cases where a super-resolution DEM is used, or where slope is estimated over an anisotropic area, such as the 1-D window geometry utilised in Section 5.5.4.

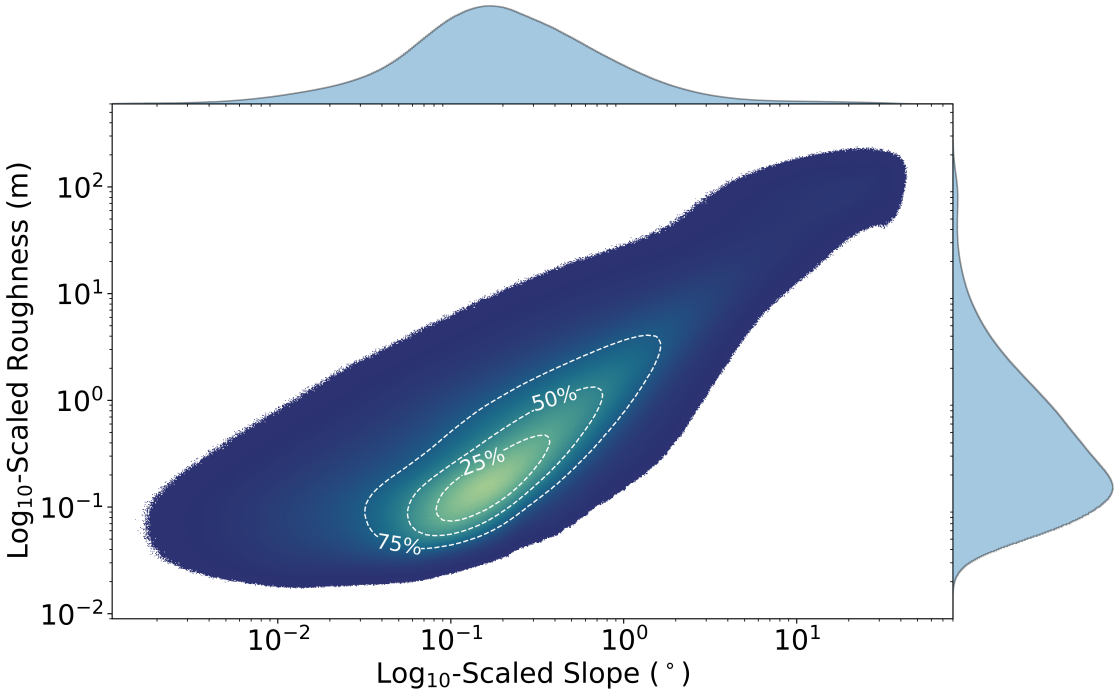

Figure 5.3: Log_{10}-binned joint and marginal distributions of slope and roughness across Antarctica. The joint distribution is displayed as a density plot, with lighter colours indicating a higher density of measurements, and the contours bounding 25%, 50%, and 75% of the data.

Secondly, I performed a similar analysis for my SVD-derived roughness map (calculated using 3×3 pixel windows), comparing it with estimates made using a 9×9 pixel window, and using the Terrain Ruggedness Index (TRI) approach (Riley et al., 1999) (Figure 5.4 b). My method yields a median roughness of 0.03 m and a median absolute deviation of 0.0181 m for a 3×3 pixel window, and a median roughness of 0.297 m and median absolute deviation of 0.219 m for a 9×9 pixel window. For TRI, the corresponding values are 0.864 m and 0.592 m, respectively, indicating that my method produces consistently lower roughness values than TRI for both

window sizes. The higher values produced by TRI likely reflect its dependency between slope and roughness, whereby steeper slopes inflate roughness estimates. To assess this, I computed the Pearson correlation coefficient between slope and roughness for both methods. I find a stronger correlation for TRI (0.947) than for my approach using the same 3×3 window (0.720), suggesting that my method is indeed more effective at reducing algorithmic slope-roughness correlation. Turning next to the choice of window size, I find that unlike for slope, the choice of window size has a substantial effect on roughness estimates, with the 3×3 window producing lower values than the 9×9 window (Figure 5.4). This likely reflects the fact that an increased number of points in a larger window raises the likelihood of higher minimum-to-maximum variance relative to the fitted plane. Nonetheless, even with a 9×9 window, the distribution of roughness values is still lower for my method than for TRI. Finally, within the context of this study, it is important to note that it is primarily the variance in roughness that is analysed (through comparison to altimetry measurements) rather than the absolute roughness values themselves.

5.6.2 Window Placement Optimisation

The S3 product records backscattered energy over a range window of ∼60 m. As such, there are areas of the AIS where the degree of topographic variance within the beam-limited footprint makes capturing the full topography within the range window impossible. Determining the locations and extent of these regions is of interest because it indicates where (1) waveforms are likely to be truncated or highly complex, (2) relocation errors may be prevalent, and (3) positioning of the range window will, ultimately, have a major impact upon the reliability of retrievals of ice sheet elevation. By assessing the proportion of illuminated terrain that could be captured with an optimal placement of the 60 m range window (Section 5.5.2), I found that, for 57.4% of recorded echoes, full (100%) topographic capture was impossible. With optimal window placement, however, 83.5% of acquisitions would be able to acquire at least 80% of the topography illuminated by the beam-limited

footprint. Failure to capture the full topography primarily occurs in coastal regions of the ice sheet with high topographic complexity (Figure 5.5 a), and reflects a fundamental limit imposed by the 60 m size of the Sentinel-3 range window, rather than the onboard tracker implementation.

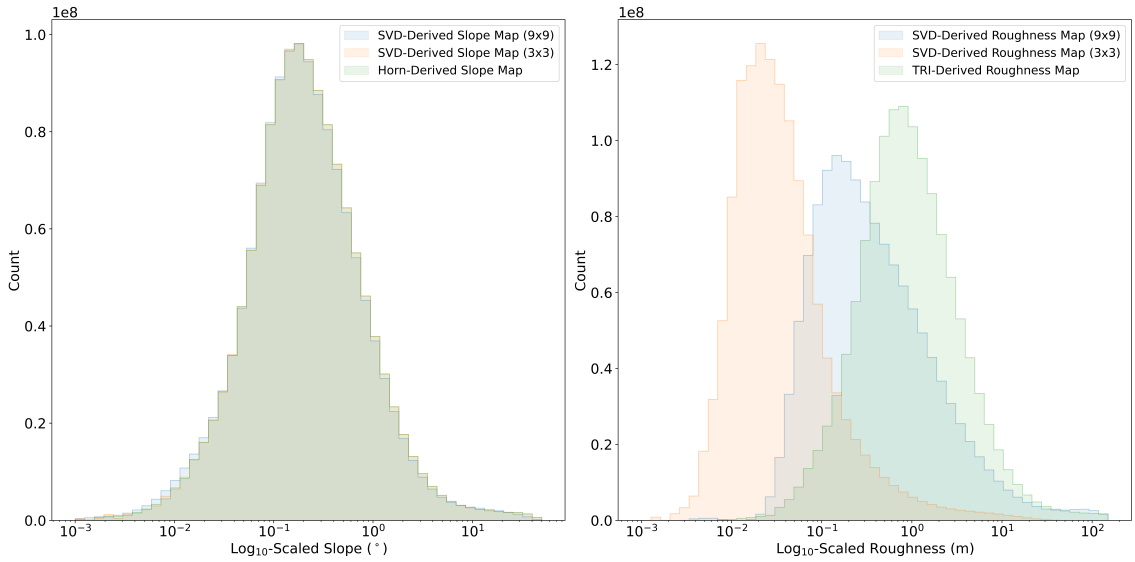

Figure 5.4: Comparison of slope and roughness estimates using different methods applied to the 100 m resolution REMA DEM. Panel a shows slope calculated using (1) Horn's method with a 3×3 pixel (300x300 m) window, (2) my SVD-based approach, calculated with a 3×3 pixel window for consistency with Horn's method, and (3) my SVD-based approach with a 9×9 pixel window. Panel b shows roughness computed using (1) TRI (3×3 pixel window), (2) my SVD-based approach using the same 3×3 pixel window size for consistency with TRI, and (3) my approach using a 9×9 pixel window.

Next, to assess the performance of the current window placement, I then computed the ratio of actual to maximal topographic capture for each record (Figure 5.5 b), where the latter was determined by calculating the maximum percentage of the illuminated surface that could have been captured with a refined placement of the range window (Section 5.5.2). At a practical level, this analysis was designed to assess the extent to which a greater proportion of the illuminated surface could be captured if refinements are made to the positioning of the range window. On average, across all records, the current window placement acquired a median of 89.2% of the

total possible topography that could be captured, with a median absolute deviation of 2.47%. This demonstrates that, even over areas where full topographic capture is impossible, further refinement to the placement of the range window could allow a greater proportion of the surface return to be recorded.

To assess the impact of recent changes made to the Level-1 processing, I also performed a similar analysis for the previous BC-004 product. As outlined in Section 5.4, BC-005 differs from BC-004 primarily through the artificial extension of the delay-Doppler stack range dimension during Level-1 processing, which allows for better placement of the range window around the surface return, preserving backscattered energy in topographically complex terrain (Aublanc et al., 2025b; Aublanc et al., 2018). Relative to BC-004, I find that although the percentage of possible capture is is broadly similar ($< 1\%$ difference) over the low slope interior of the ice sheet ($< 1.0°$ slope), the BC-005 product performs substantially better ($+7.6\%$) across the higher sloped margins (1.0-2.5° slope). Broadly similar behaviour is observed with respect to surface roughness, with $< 1\%$ difference and a $+2.2\%$ improvement for the latest product, in regions where the surface roughness is $<$ 5 m and 5-10 m, respectively. As such, this analysis suggests that the extended window processing and waveform centering has indeed improved data capture over high sloped terrain, where the along-track variation in tracker range during the SAR integration time is greatest.

In interpreting, and potentially acting on, this analysis of topographic capture, it is important to note that in some cases where there is high topographic variability within the illuminated footprint, positioning the range window to maximise topographic capture may result in the return from the POCA being missed. In these cases, there is a trade-off between maximising topographic capture and acquiring the reflection from the point of closest approach. For future missions that aim for full waveform retrievals (such as swath processing), this type of analysis can therefore help to inform decisions relating both to the size (i.e. the range dimension) and the positioning (onboard tracking and L1 processing) of the range window. It

should also be noted that to resolve the leading edge clearly, the range window must record a brief period of no backscattered power before the first return. This requires positioning of the upper boundary of the window slightly above the POCA to ensure the reliable detection of the waveform onset.

Comparing these results to my new slope and roughness datasets, I find that both the actual, and theoretical maximum, percentage of the illuminated topography that can be captured within the S3 range window decreases with increasing slope and roughness (Figures 5.5 c and d). Specifically, for $\log_{10}$-binned slope I observe that $\sim$90% of the illuminated surface is typically captured for slopes up to $\sim$0.1°, followed by a steady decline to $\sim$20-30% capture by the time slopes reach $\sim$5°. I find a similar trend in relation to $\log_{10}$-binned surface roughness, with capture rates decreasing with increasing roughness, from $\sim$90% (low roughness) to $\sim$30-40% (once roughness reaches $\sim$10 m). In both cases, I observe a difference of $\sim$10-20% between the maximum possible percentage of topography that could be captured and the actual percentage captured by the current BC-005 processing configuration.

While the nominal 60 m range window is a fixed instrument limitation, the use of "extended window" processing in BC-005 - where the range dimension of the delay-Doppler stack is artificially extended during Level-1 range migration and multilooking to better preserve backscattered energy (Section 5.4) (Aublanc et al., 2025b; Aublanc et al., 2018) - demonstrates that further gains in topographic capture may still be achievable through reprocessing existing Level-0 data, provided the raw telemetry are available.

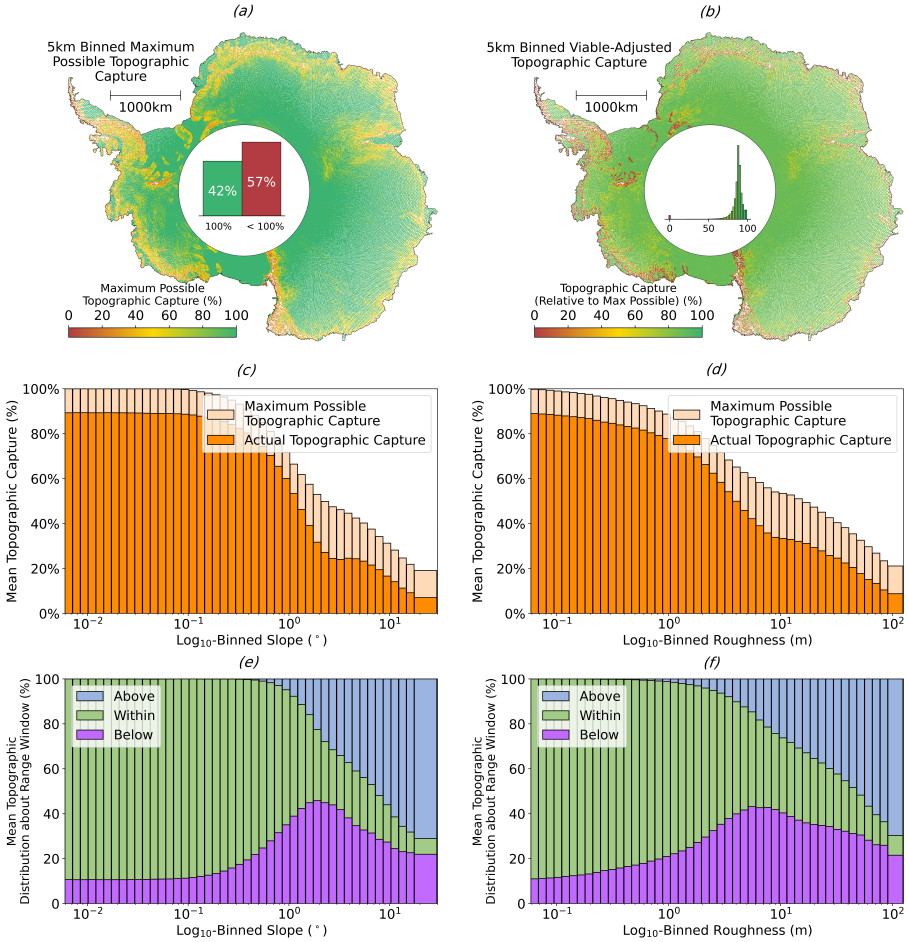

Figure 5.5: Assessment of the Sentinel-3 range window's ability to capture surface topography within the beam-limited footprint. Panel a shows the maximum possible percentage of REMA topography (100 m resolution) capturable by an optimally placed 60 m range window, binned to 5 km. This represents a ceiling imposed by window size relative to topographic variability. The inset bar chart shows the percentage of records where full (100%) capture is and is not achievable. Panel b shows the percentage of topography actually captured by the current range window placement relative to this maximum, binned to 5 km. A value of 100 indicates the current placement matches the theoretical optimum, not necessarily full surface capture. Panels c and d show the maximum possible and actual capture percentages as a function of $\log_{10}$-binned slope and roughness, respectively. Panels e and f show the distribution of REMA points above, within, and below the range window for the maximum possible and actual configurations.

Next, I assessed the mean proportion of the illuminated surface that was above, within, and below the range window. This allowed me to determine the degree

to which the window was centered on the target surface, and how this varies as a function of slope and roughness (Figures 5.5 e and f). This analysis indicates that the ∼10% of the illuminated topography not captured within the range window at lower values of slope and roughness falls predominantly below the range window. As slope and roughness increases, topographic capture decreases, with the proportion of topography above and below the range window increasing in approximately equal measures. Once slope and roughness exceed ∼2° and ∼7 m respectively, the proportion of topography below the range window begins to decrease, and is replaced with surface topography above the range window.

In addition to assessing the dependency on surface slope and roughness, I also evaluated whether the heading of the satellite as it crosses the Antarctic coast had a significant impact on window placement. I found that 50.5% of records within 30 km of the coast were acquired in a configuration where the satellite was flying from land to ocean, and 49.5% from ocean to land. When the satellite is heading from land to ocean, I obtain a median topographic capture (with respect to the theoretical maximum) of 87.0% for records within 30 km of the coast. Conversely, when the satellite flies from ocean to land, I obtain a median topographic capture of 87.4%. With a difference of less than 0.5%, I conclude that there is therefore insufficient evidence to suggest that coastal heading has a significant impact on the adequacy of window placement.

Within the preceding analysis, I have evaluated the extent to which the 60 m range window of Sentinel-3 impacts upon the instrument's capacity to capture the full topographic surface illuminated by the beam footprint. To conclude this component of my analysis, I therefore broadened the scope to consider how the statistics relating to the theoretically possible topographic capture were affected by varying sized range windows (Figure 5.6). For a 60 m range window, the mean percentage of topography captured with optimal placement is 90.6%, with a standard deviation of 15.9%. As the range window increases, the mean topographic capture rises to 96.9% (120 m window), 98.4% (180 m), 99.0% (240 m), and 99.5%

(360 m), with corresponding standard deviations of 10.5%, 7.7%, 6.0%, and 4.0%, respectively. The most substantial improvement occurs between 60 m and 120 m, with diminishing returns observed at larger window sizes. These findings also support the range window size chosen for the upcoming CRISTAL mission, which will operate with a 256 m range window and interferometric (i.e. swath processing) capability over land ice (Kern et al., 2020).

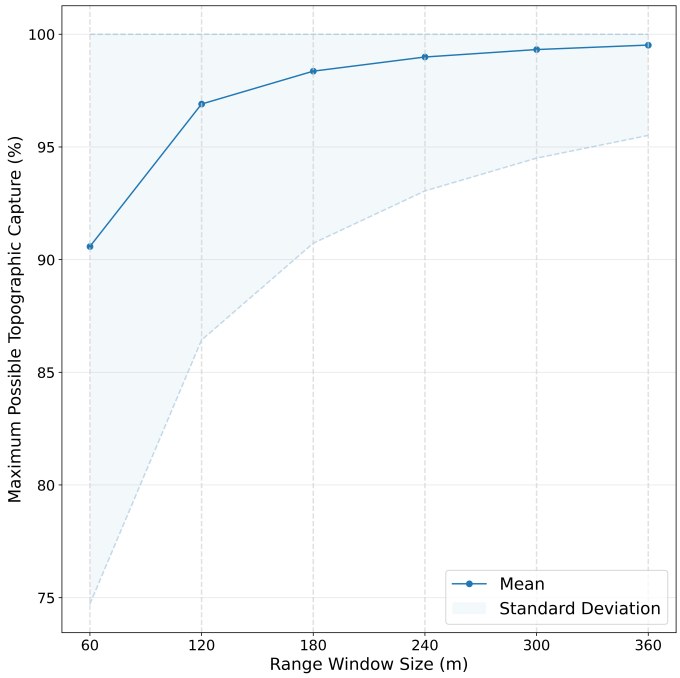

Figure 5.6: Maximum possible percentage of topography captured with optimal placement of the range window as a function of range window size. The shading shows the variability across records in the percentage of topographic capture, characterised by the standard deviations and limited at its maximum to 100%.

5.6.3 Capture of the Point of Closest Approach

Next, I analysed the positioning of the POCA identified in the Level-2 product relative to the range window and beam-limited footprint, and compared this with the geometric point of closest approach determined by minimising the distance between the satellite and the REMA surface. In practice, each measurement was classified according to its position relative to the range window and footprint (Figure 5.7a). As

outlined in Section 5.5.3, the L2 POCA classes (LPCs) are derived by comparing the geographic coordinates of the POCA location prescribed in the Sentinel-3 Level-2 product, assigned the nearest REMA elevation, with the bounds of the range window for each acquisition. The REMA-derived POCA classes (RPCs) are computed analogously, but use an independently determined POCA location - defined as the DEM pixel within an extended search region that minimises the satellite-to-surface range - rather than the Level-2 product value. Both classifications assign each acquisition to one of six classes (Figure 5.1), depending on whether the POCA falls within or outside the beam-limited footprint, and above, within, or below the range window.

To explore spatial variability, both the LPCs and RPCs were binned to 5 km resolution (Figures 5.7b and 5.8b). As these data are discrete, the modal class was assigned to each grid cell to represent the most common class within that area. Using the new slope and roughness derivations (Figure 5.2), I also examined the relationships between these classes and the topographic parameters (Figures 5.7 and 5.8).

As a result of quality control and filtering in the L2 processing, I observe that ∼0.447% of the data is missing for the LPCs within the BC-005 product. This represents a significant improvement compared to the previous BC-004 product, wherein ∼3.20% of the data were missing, primarily over topographically complex areas. The proportion of missing data increases with slope and roughness, comprising 5-10% of samples in regions of very high slope and roughness (Figures 5.7 e and f, respectively). For 94.1% of the full, non-aggregated data, I observe that the L2 POCA lies within both the Sentinel-3 beam-limited footprint and the range window. This occurs almost exclusively in regions of lower slope and roughness, with a median of 0.170° and 0.277 m, and median absolute deviation of 0.105° and 0.189 m respectively (Figures 5.7 c and d). In contrast, the remaining 5.9% of acquisitions where the L2 POCA fails to be located within the range window encompass a far greater range of slope and roughness values, retaining significantly higher medians.

This is evident spatially, with LPCs with a value other than 2 grouped around areas of complex topography such as the Transantarctic mountains, the Antarctic Peninsula, and coastal regions (Figure 5.7 b). Comparing to the performance of the BC-004 product, again I find that the evolutions in BC-005 offer an improved capacity to derive a POCA that is consistent with the range window. Specifically, I find that BC-004 returned a lower (89.5%) percentage of records where the L2 POCA was correctly located within the range window and beam footprint (i.e. LPC class 2). Thus, the switch from BC-004 to BC-005 has approximately halved the proportion of acquisitions where the prescribed POCA location and the range window position are inconsistent.

I find that the proportion of L2 POCA's falling within both the beam-limited footprint and the range window (class 2) decreases with increasing slope and roughness. For both covariates (Figures 5.7 e and f), the relationship follows trends similar to those reported for topographic capture in Section 5.6.2. I find that for surface slopes (roughness) up to $\sim$0.3° ($\sim$0.3 m), the current window placement achieves near to 100% capture of the POCA location. As slope increases beyond this limit, however, the proportion of class 2 decreases, such that by $\sim$1° only $\sim$70% of the L2 POCA's lie within the desired range, and by $\sim$5° this reduces further to $\sim$30% (Figures 5.7 e and f). Likewise for roughness, I observe a reduction in the proportion of class 2 with increasing roughness, wherein by $\sim$10 m the proportion of LPCs of class 2 reduces to $\sim$40%, and further to $\sim$20% by $\sim$50 m (Figures 5.7 e and f). With increasing slope and roughness, the observed reductions in class 2 are replaced in approximately equal parts by classes 0 (POCA above the range window) and 4 (POCA below the range window), up until $\sim$3° slope and $\sim$1 m roughness, wherein the increase in proportion of class 4 plateaus, and classes 0 dominate towards the extremes. This likely reflects the inability of the linear slope correction to adequately identify a point within the range window, in regions of high topographic variability. Notably, the occurrences of classes 1, 3, and 5, which constitute POCA locations outside of the beam-limited footprint, are negligible (0.3%, 0.4%, and

0.0%, respectively), and much lower than found in BC-004 (1.1%, 0.7%, and 0.7% respectively), likely a result of the algorithmic decision to avoid relocations beyond the 3 dB beamwidth.

The RPCs, which are defined using the location that minimises the range between the surface and satellite, rather than the location written in the Level-2 product, follow broadly the same distribution as the LPCs, with the exception of minimal data loss ($\sim$0.00261%), which is simply due to missing data within the REMA dataset (Figure 5.8 b). Overall, 90.6% of the full, non-aggregated data has an RPC of 2 (corresponding to a POCA within both the beam-limited footprint and range window), with median slope and roughness for this class of 0.163° and 0.259 m, respectively (Figures 5.8 c and d). Notably, using BC-004 product values leads to a slightly reduced 89.0% of class 2 RPCs, indicating that the BC-005 product better captures the "true" POCA as defined by REMA, due to its extended window processing. The level of agreement on an RPC classification of 2 between the two product versions varies significantly with surface slope. For slopes of 0.0-1.0°, 1.0-2.5°, and $\geq$2.5°, the percentage of records for which BC-004 and BC-005 agree is 93.6%, 19.0%, and 3.25%, respectively, illustrating the impact that the refined BC-005 processing is having for retrievals in topographically complex areas. In these same slope bins, the percentage of acquisitions where BC-005 correctly returns a class 2 classification, whereas BC-004 does not, is 1.93%, 13.8%, and 1.88%; whereas the converse case where BC-004 returns class 2 and BC-005 does not, occurs for only 0.766%, 3.04%, and 0.832% of records. A similar pattern is observed with respect to surface roughness. For roughness ranges of 0.0-5.0 m, 5.0-10 m, and $\geq$10 m, the percentage of records where BC-005 returns class 2 and BC-004 does not is 2.45%, 8.26%, and 2.10%, whereas the converse occurs for only 0.845%, 2.59%, and 1.23% of acquisitions. Together, these observations illustrate the improvements delivered by the latest baseline, and suggests that the extended window processing implemented in BC-005 is able to more reliably track the POCA in regions with more complex topographic terrain. Nonetheless, the fact that 64.1% and 94.0% of

acquisitions in regions of 1.0-2.5° and ≥2.5° slope, respectively, do not have a REMA-derived POCA within the range window for either BC-004 or BC-005 illustrates the challenges presented by the relatively small 60 m range window in areas of complex topography, and the opportunity for further algorithmic evolution.

In comparison to the LPC's, the RPCs return more cases where the POCA is located above the range window, and less instances where the POCA is below the range window (Figures 5.8 c and d). This is unsurprising given that the RPC approach, by minimising the range, favours the highest point within the footprint, which is more often above rather than below the range window. With respect to slope and roughness, the RPC's exhibit similar trends in class 2 as the LPC's, although there are a greater proportion of measurements that are located outside of the 3 dB beamwidth (1, 3, and 5), likely due to the fact that when I determine the RPC, I do not place any restrictions on it being within the 3 dB beamwidth. This highlights the fact that in areas of complex topography the point of closest approach can often be beyond the 3 dB beamwidth. The proportion of classes 0 and 1 increases with slope and roughness in approximately equal parts, however, unlike the LPC's, classes 4 and 5 experience only a temporary increase with increasing slope and roughness, peaking at ∼3° and ∼9 m, respectively, before being replaced by classes 0 and 1 for more extreme topographic variance (Figures 5.8 e and f). I also find that classes 4 and 5 often occur at the transition between floating and grounded ice (Figure 5.8 b), where there is a rapid change in the surface gradient, and hence may be due to the inertia in the closed-loop tracker, which positions the range window based upon preceding tracker information. As slope and roughness increase further, then classes 0 and 1 again dominate, reflecting the behaviour typical of mountainous terrain where local topographic highs exceed the positioning of the range window.

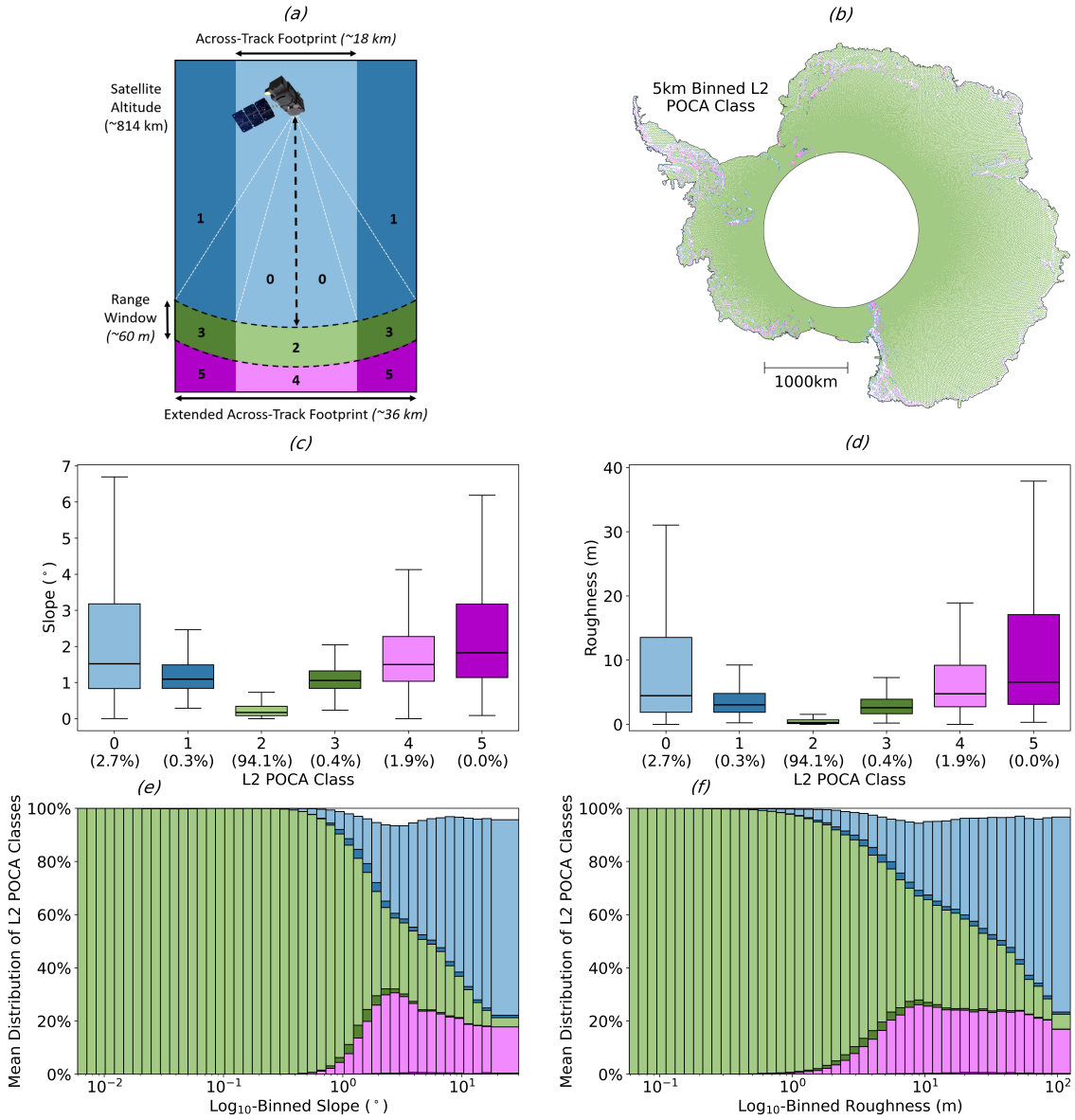

Figure 5.7: Assessment of the ability of Sentinel-3 to capture the surface reflection at the L2-derived POCA. Panel a outlines the POCA class definitions as per Figure 5.1 for ease of cross-referencing. Panel b shows the L2-derived POCA classes, spatially binned to 5 km. A 5 km grid was chosen as it provides sufficient spatial resolution to resolve regional patterns whilst ensuring an adequate number of measurements per grid cell for statistically meaningful aggregation. Panels c and d show box plots of the L2-derived POCA per-class distribution with respect to slope and roughness, with the percentage of data within each class given in parentheses. Panels e and f show stacked bar charts representing the percentage distribution of the L2-derived POCA classes within each slope and roughness bin.

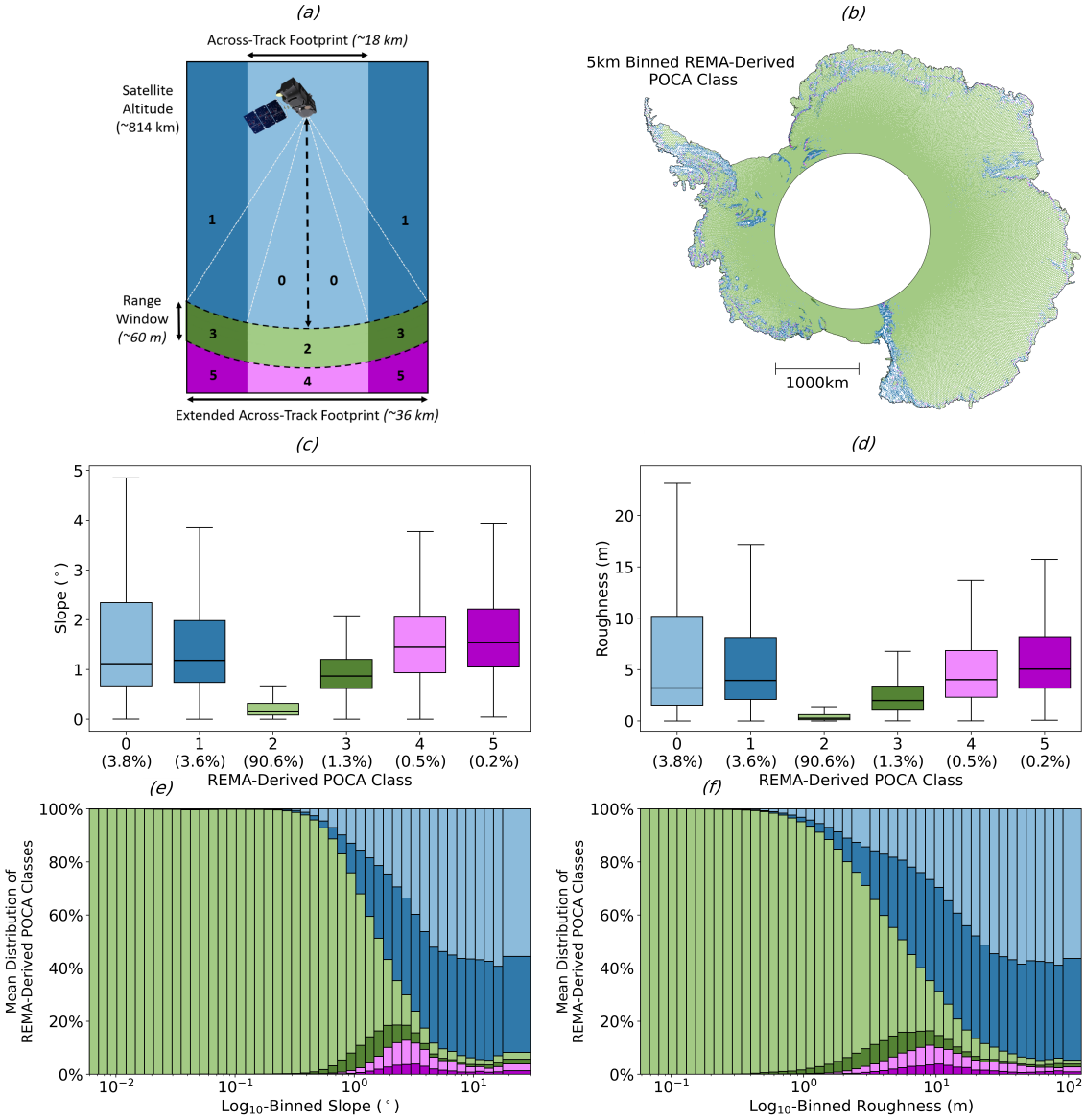

Figure 5.8: Assessment of the ability of Sentinel-3 to capture the surface reflection at the REMA-derived POCA. Panel a outlines the POCA class definitions as per Figure 5.1 for ease of cross-referencing. Panel b shows the REMA-derived POCA classes, spatially binned to 5 km. Panels c and d show box plots of the REMA-derived POCA classes with respect to slope and roughness. Panels e and f show stacked bar charts representing the percentage distribution of the REMA-derived POCA classes with respect to $\log_{10}$-binned slope and roughness.

Lastly, I again assessed whether the per-class distribution was affected by the satellite heading for acquisitions made within close proximity (within 30 km) of the

ice sheet coast. For the RPCs, I observed that the range window was adequately placed (i.e. class 2) for 50.4% of the measurements that were acquired when heading from land to ocean, and 51.6% when heading from ocean to land. A difference of only ∼1.2% suggests that there is no definitive evidence that coastal heading affects the per-class distribution close to the ice sheet coast, similar to the negligible difference in topographic capture rate (with respect to the theoretical maximum) that was observed in Section 5.6.2.

5.6.4 Along-track Decorrelation of Echoes

My final set of analyses focuses on exploiting my new slope and roughness datasets to assess the correlation of series of altimetry echoes acquired along the orbit track, with a view to investigating the impact of surface topography on the degree of waveform correlation achieved. To investigate correlation across the entire ice sheet, I bin both the correlation between consecutive waveforms, and the fitted 50-record correlation change, taking the mean values within each 5×5 km grid cell (Figures 5.9 a and b). For the inter-record correlation, I found a median of 0.944 and a median absolute deviation of 0.00300, with the distribution skewed towards higher correlation values, and lower values grouped around areas of topographic complexity (Figure 5.9 a). For the fitted 50-record correlation change, I found a median reduction in the correlation of -0.418 over 15 km and a median absolute deviation of 0.0250. 12.6% of records had a correlation change of less than -1 and were removed from my analysis (as discussed in Section 5.5.4). By comparing qualitatively to my REMA-derived estimates of across-track surface slope and roughness (Figure 5.2), it is clear that both of these topographic parameters influence the degree of correlation between waveforms. For example, correlation reduces in areas of high roughness, such as megadune fields in East Antarctica, and high surface slope, such as the coastal margin of the ice sheet (Figures 5.9 a and b).

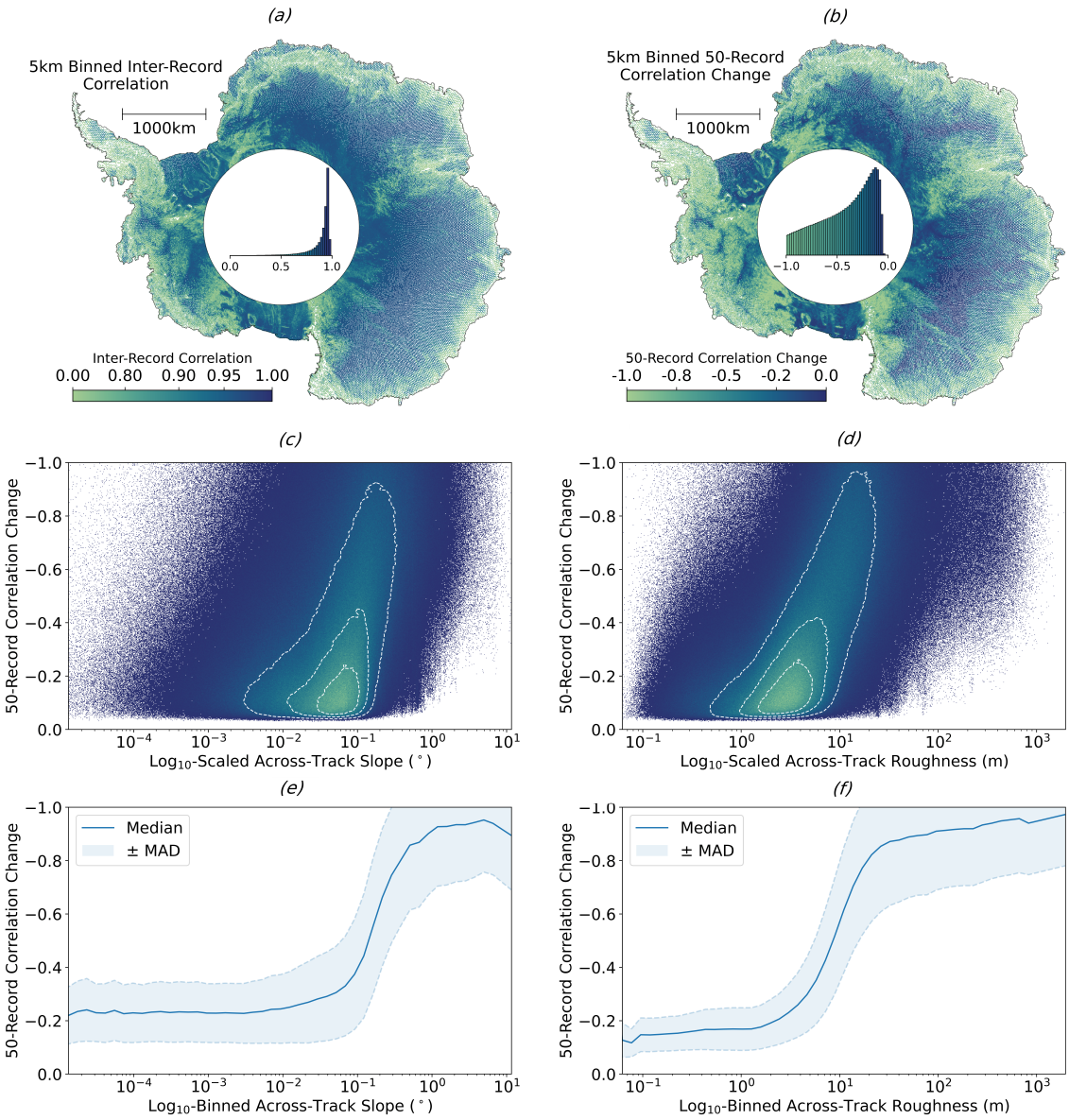

Figure 5.9: Analysis of the along-track correlation of Sentinel-3 SAR altimetry echoes. Panel a shows the correlation between consecutive records, spatially binned to 5 km. Panel b shows the 50-record correlation change, also spatially binned to 5 km. Histograms showing the overall distribution of each parameter are given within the pole holes of the respective plots. Panel c and d show density plots of the 50-record correlation change as a function of $\log_{10}$-binned slope and roughness respectively. Concentric contours are included, which contain 75%, 50%, and 25% of the data from outside to in respectively. Panels e and f show the median and median absolute deviation of the 50-record correlation change with respect to $\log_{10}$-binned slope and roughness, respectively.

Comparing the fitted 50-record correlation change to across-track surface slope and roughness, it can be seen that the 15 km change in correlation remains relatively stable at ∼-0.25, for slopes up to ∼0.01°, before dropping rapidly to ∼-0.95 for slopes above this value (Figure 5.9 e). For low surface roughness, the 50-record correlation change is ∼-0.15, but this quickly drops towards ∼-1 once roughness exceeds ∼2 m (Figure 5.9 f). Taking the Pearson correlation coefficient of the 50-record correlation change to $\log_{10}$-binned surface slope and roughness I find values of -0.618 and -0.571. This implies that, linearly, slope and roughness explain 38.2% and 22.2% of the ice-sheet wide variance in along-track waveform decorrelation. I therefore conclude that 50-record correlation change is moderately, linearly correlated to $\log_{10}$-binned surface slope and $\log_{10}$-binned roughness, with slope providing a stronger control. Whilst conceptually, it is understood that slope and roughness impact the shape of the echo, and associated parameters such as the number of peaks, leading edge width, and trailing edge slope, this analysis demonstrates quantitatively the strength of this correlation, through the comparison of independent altimetry and DEM-derived datasets.

5.7 Conclusions

Since the launch of Sentinel-3A in 2016, and Sentinel-3B in 2018, the Sentinel-3 mission has acquired hundreds of millions of echoes over Antarctica, providing topographic information at a resolution of ∼300 m along-track by ∼1.6-2 km across-track, which are repeated every ∼27 days. To date, however, previous assessments of Sentinel-3 performance over ice sheets have concentrated predominantly on the outputs of L2 processors, namely the high-level validation of the resulting elevation products. Here, I extend previous work by using novel topographic datasets derived from REMA, to more directly explore the underlying factors that influence Sentinel-3's performance. Specifically, I use Singular Value Decomposition to robustly separate local elevation variability into slope and roughness components,

and investigate the impact of each upon Sentinel-3 altimeter performance.

Quantitatively, my new estimates of slope and roughness indicate a strong positive correlation (Pearson correlation coefficient of 0.808) between these two variables; areas with higher surface slopes also exhibit greater kilometre-scale (orthogonal) surface roughness. Overall, I find that 76.9% and 89.1% of the ice sheet exhibits slopes below 0.5 and 1.0°, respectively; and 62.4% and 76.1% has a roughness lower than 0.5 and 1.0 m, respectively (where roughness is defined as the minimum-to-maximum difference of the orthogonal residuals). Qualitatively, my new slope and roughness maps display a wealth of glaciological information, including ice shelf flow patterns and the locations of megadune fields in East Antarctica. These findings enhance understanding of the AIS's geometrical properties and represent, to the best of my knowledge, the first continent-wide analysis of Antarctic surface roughness, that is independent of slope.

During operation, the Sentinel-3 onboard tracker aims to position the range window so that it captures the underlying topographic surface. One way to measure its success in doing so, whilst acknowledging the inherent limitations imposed by the range window dimension, is to assess the proportion of the surface illuminated by the instrument that falls within the range window. Performing such an assessment at the ice sheet scale, through the analysis of over 8 million echoes, I find that in 57.4% of cases, capturing the full topography within the beam-limited footprint is impossible, due to the relatively small ∼60 m range window of Sentinel-3, relative to the high topographic variance found across some regions of the ice sheet. Nonetheless, my analysis shows that with optimal placement of the range window, 80% of the illuminated topographic surface could be captured in ∼80% of cases. Building on this, I investigated the performance of the current window placement, by measuring the ratio of actual topographic capture to the maximum possible capture that could be achieved with an adjusted placement of the range window. On average, the current placement achieved a median of 89.2 ± 2.46% of total possible topographic capture, with this value decreasing with increasing slope and roughness. Broader

analysis of the relationship between the range window size and the maximum achievable topographic capture showed that substantial gains can be achieved by increasing the range window beyond 60 m. Doubling the window size to 120 m, for example, could increase the mean capture from 90.6% to 96.9%. Together, this analysis serves to provide a better understanding of the performance of the Sentinel-3 onboard tracker, on-ground processing, and also to support the design of future missions, by providing empirical information with which to optimise the size and positioning of the range window.

Within current Level-2 processing schemes, there is an implicit assumption that the prescribed point of closest approach (POCA) lies within both the instrument's beam-limited footprint and its range window, and hence that the range to the waveform leading edge corresponds to the distance between the satellite and the identified POCA. To evaluate this assumption, I compared the locations of the L2-designated POCA in two product versions (BC-004 and BC-005) to the position of range window for ∼8 million acquisitions collected over the Antarctic Ice Sheet, in addition to that of an independent POCA estimate that I derived from a high resolution DEM. This analysis showed that the current window placement successfully captures the point of closest approach for 90.6% of acquisitions and allowed me to investigate, quantitatively, the improvements offered by the BC-005 extended window processing, in areas of complex topographic terrain.

Finally, I explored the correlation between sequences of altimetry echoes along the satellite orbit track, to provide a first assessment of how surface topography affects waveform decorrelation. I found that inter-record correlation was generally high, but decreased in areas with complex topography. Similarly, taking the larger-scale correlation change across 50-records (spanning ∼15 km) indicated that the rate of decorrelation was more rapid in regions with high surface slope and roughness. This analysis therefore provides direct evidence of the relationship between surface topography and along-track altimetry echo correlation at the ice sheet scale, offering new insight into how surface morphology influences waveform structure.

More broadly, this study has shown how a new generation of high-resolution DEMs can be used to assess and understand the performance of current altimetry in-flight systems and on-the-ground processing. This information provides new insight to help understand the current performance of the Sentinel-3 fleet of satellite altimeters, and also serves as a template for future analysis that can guide the design and optimisation of planned missions such as the Copernicus Polar Ice and Snow Topography Altimeter (CRISTAL) and the Sentinel-3 Next Generation Topography mission.

Chapter 6

Quantifying Full Antarctic Ice Sheet Topographic Ambiguity in Non-Interferometric Satellite SAR Altimetry using Deep Learning

Joe Phillips[1], Malcolm McMillan[1]

[1]UK Centre for Polar Observation and Modelling, Centre of Excellence in Environmental Data Science, Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence: Joe Phillips (j.phillips5@lancaster.ac.uk)

The following work is being prepared as a manuscript to be submitted for publication.

JP performed all code development, analysis and manuscript preparation, with supervision from MM.

Abstract

Conventional, non-interferometric radar altimetry faces a fundamental ambiguity: each waveform corresponds to many different possible surface configurations within the illuminated footprint. Standard processing endeavours to resolve this by reducing the signal to a single elevation estimate, typically tied to the leading edge of the waveform and attributed to the point of closest approach. As such, these conventional approaches, which have been employed for decades, discard much of the information encoded in the waveform return, and are susceptible to data loss and degradation in areas of complex terrain, where these simplifying assumptions break down. This study takes a fundamentally different approach: rather than attempting to remove this ambiguity through dimensionality reduction, it explicitly models it, treating waveform multiplicity as a feature to be learned and quantified, with a view to the future redesign and improvement of Level-2 altimetry processing chains. Specifically, a novel probabilistic deep learning framework is developed and assessed that predicts full distributions of plausible surface profiles from CryoSat-2 SARIn waveforms, using no interferometric phase information. By directly modelling both aleatoric uncertainty (arising from the multiplicity of possible scattering surfaces) and epistemic uncertainty (reflecting uncertainty in the model predictions themselves), the framework produces spatially distributed, probabilistic elevation predictions that capture the inherent ambiguity in the waveform. In more detail, an ensemble of 16 ResNet-RS models is trained with Pinball loss to estimate multiple elevation quantiles (5^{th}, 50^{th}, 95^{th}) across a 15 km across-track swath. Evaluation against an independent Antarctic test set shows that the ensemble captures the across-track surface geometry well visually, with predicted quantiles varying systematically in ways consistent with known radar scattering behaviour. Quantile coverage analysis reveals a mean underestimation of 10.4% in the percentage of unseen data occupying the central 5-95% interval, driven primarily by the 5^{th} and 50^{th} percentiles being systematically overestimated in elevation relative to the true surface, a tendency that increases as terrain grows more complex.

In contrast, the 95$^{\text{th}}$ percentile estimate of surface elevation remains robust, showing little sensitivity to slope or roughness, which is consistent with its physical anchoring in the leading-edge return. As such, by reframing waveform ambiguity as a source of information rather than a limitation, this study demonstrates that an ensemble of deep learning models can be trained to extract meaningful across-track distributional elevation information from SAR power waveforms alone, without the introduction of interferometric information. This represents a first step toward unlocking the full potential of non-interferometric altimetry, paving the way for greater measurement density, improved robustness and uncertainty characterisation, and new elevation retrieval methods from current and future SAR missions such as Sentinel-3 and CRISTAL, with potential for adaptation to historical LRM records from missions such as ERS-1 and ERS-2.

6.1 Introduction

6.1.1 Context

The Antarctic Ice Sheet (AIS) is undergoing rapid and accelerating change, making it one of the largest sources of uncertainty in projections of future sea-level rise. Since 1992, the AIS has lost over 2,600 billion tonnes of ice (Otosaka et al., 2023), with the greatest losses observed in West Antarctica and the Antarctic Peninsula. This ongoing imbalance has already contributed more than 7 mm to global mean sea level over the past 4 decades (Otosaka et al., 2023), and projections suggest that Antarctic contributions could exceed one metre by 2100 under high-end scenarios (Frederikse et al., 2020). However, such estimates remain highly uncertain - not only due to variation in future climate forcing, but also because of limited understanding of the AIS's dynamics and often nonlinear response to external drivers. In this regard, long-term, high-resolution observations are critical to reduce this uncertainty, offering the means to monitor evolving changes and to improve the representation of ice-sheet

processes in predictive models (Davis et al., 2005; Price et al., 2011; Shepherd et al., 2004).

6.1.2 Satellite Radar Altimetry

Further background on the theoretical principles of radar altimetry is given in Chapter 2; here I briefly re-introduce these concepts and summarise key technical aspects.

Satellite observations are critical for understanding ice sheet change, with radar altimetry providing the longest continuous record since the early 1990s. Satellite radar altimeters work by transmitting radio-wave pulses towards Earth's surface and recording the returned echo as a discrete waveform (Brown, 1977). Each waveform, comprising the aggregation of surface reflections within the altimeter footprint, encodes information relating to surface topography, electromagnetic scattering characteristics, and over oceans, wind speed and wave height (Quartly et al., 2020). Over flat surfaces, waveforms exhibit a distinctive shape, with a clear peak in power corresponding to the energy backscattered from the point of closest approach (POCA) on the surface, followed by a gradual decay in the received power as reflections further from the antenna are captured. As surface topography becomes more complex, however, the returned waveform becomes increasingly variable, reflecting the wider range in travel-times of surface reflections due to variations in surface elevation within the illuminated footprint.

The era of systematic polar monitoring by radar altimeters began with the launch of the ERS-1 and ERS-2 missions in the early 1990s. These missions were designed primarily to measure the ocean geoid, and as a result offered relatively poor performance over complex ice sheet topography, due in part to their relatively large pulse limited footprint associated with their operation in Low Resolution Model (LRM) (Joughin et al., 1996; Rémy et al., 2009). As such, following technological advancements in subsequent missions such as Jason-1 and Envisat, CryoSat-2 (CS2) was eventually launched in 2010, which introduced two key innovations in

instrumental design for improving ice surface measurements: (1) the use of synthetic aperture radar (SAR), and (2) interferometry (SARIn when used alongside SAR processing) (Wingham et al., 2006b).

In conventional LRM operation, the altimeter must wait for each radar pulse to complete its round trip before transmitting the next. This constraint enforces a low PRF and limits along-track resolution to the kilometre scale, preventing the detection of fine-scale topographic variation. Synthetic aperture radar (SAR) altimetry overcomes this limitation by transmitting bursts of pulses at a much higher PRF, enabling overlapping emissions and finer along-track sampling. The resulting echoes are then separated using Doppler processing, which exploits the frequency shifts induced by satellite motion to achieve improved along-track resolution.

With a single antenna, altimetry is one-dimensional: each waveform records only the return time of reflections, not their origin within the beam footprint. Each reflection could originate anywhere along an arc defined by the antenna gain pattern and echo return time. Thus, a waveform represents a distribution of possible surfaces within the footprint, unresolved without further information. Interferometric altimetry resolves this ambiguity by employing two antennas to measure the angle of arrival, allowing for detailed elevation profiles consisting of numerous geolocated points across-track.

Following CS2, Sentinel-3 A and B were launched in 2016 and 2018, further advancing operational monitoring and global SAR altimetry coverage, but lacking interferometric capability. The upcoming Sentinel-3 Next Generation (S3NG) satellites (Egido, 2023), along with the planned CRISTAL mission (Kern et al., 2020) (which will employ SAR interferometry at Ku-band but SAR only at Ka-band), will ensure the widespread availability of non-interferometric SAR altimetry over the ice sheets well into the next decade. With these measurements set to increase in both volume and temporal coverage, there is a strong case for investing in more fundamental approaches to SAR waveform data processing and information extraction, laying the groundwork for long-term scientific and operational gains.

6.1.3 Key Steps in Current SAR Altimetry Level-2 Processing Logic

Further theoretical detail and background on current SAR altimetry Level-2 processing logic is provided in Chapter 2; in this section, I briefly re-introduce these concepts.

Due to the range ambiguity within conventional, non-interferometric waveforms (LRM and SAR), each discrete waveform theoretically corresponds to a vast number of possible topographic surfaces within the satellite footprint. As such, it is impossible to fully resolve this ambiguity in the origin of reflections without additional data and simplifying assumptions. Current ice sheet SAR altimetry processing chains, which have maintained broadly the same logic for the past three decades, address this ambiguity via two steps: (1) retracking, and (2) slope correction. Together, these aim to reduce the dimensionality of the problem such that a singular measurement corresponding to the reflection from the point of closest approach on the surface is obtained. Specifically, the assumption is made that the leading edge of a captured echo - the initial steep rise of the pulse from the noise floor to the inflection point at peak power - corresponds to the first surface return at the point of closest approach (POCA) to the satellite. Using this assumption, retracking works to first obtain a single range measurement from the leading edge, corresponding to the distance between the satellite and the surface, with the slope correction step then used to identify the POCA location on the surface to attribute this to.

Retracking algorithms can be broadly divided into physical approaches, which fit analytical models of radar-surface interaction but are sensitive to irregular echoes and surface scattering properties (Martin et al., 1983; Ridley et al., 1988; Passaro et al., 2022; Villadsen et al., 2016), and empirical approaches, which instead rely on waveform geometry and are more robust to deviations from the theoretical form (Wingham et al., 1986; Bamber, 1994; Davis, 1997; Helm et al., 2014; Huang et al., 2024). The subsequent slope correction step was initially implemented

through slope-based methods that assume a constant surface slope within the footprint (Bamber, 1994; Brenner et al., 1983; Cooper, 1989; Remy et al., 1989), but more recent point-based approaches incorporate auxiliary DEMs or waveform-derived information to better locate the POCA (Levinsen et al., 2016; Roemer et al., 2007; Li et al., 2022; Huang et al., 2024; Aublanc et al., 2025a). Despite these developments, slope correction remains the dominant source of error in SAR altimetry, with uncertainties commonly on the order of tens of metres in elevation and kilometre-scale horizontal offsets (Brenner et al., 2007; Li et al., 2022).

In contrast to these conventional pipelines, when interferometric information is available, swath processing can be used to retrieve elevation estimates across a greater proportion of the full radar footprint (Gourmelen et al., 2018). Rather than retracking just the strongest return, swath methods interpret the entire waveform - combining range and interferometric phase - to resolve multiple elevation points across-track. This reduces the need for both retracking and slope correction, enabling dense, high-resolution mapping of surface topography, even in steep or heterogeneous terrain. However, although SARIn allows for these improved capabilities, it remains subject to the same range window constraints discussed in Chapter 5. As such, returns from surfaces beyond the window are still truncated, limiting coverage. Furthermore, swath processing is currently restricted to cases where interferometric information is available (i.e. CryoSat-2 SARIn mode) and cannot be applied to SAR altimeter missions that provide only power waveforms. Consequently, in these cases, the simplifying assumptions involved in reducing the full waveform to a single elevation measurement remain necessary. Thus, in the latter scenario, the simplifying assumptions and reduction associated with converting the full waveform to a single elevation measurement is still required.

6.1.4 Towards New Data-driven Approaches

Although measurements retrieved using conventional SAR altimeter processing chains have proven invaluable for investigating ice sheet dynamics over Antarctica

(Schröder et al., 2019; Fricker et al., 2012; Nilsson et al., 2022), both the quantity and quality of processed data is reduced over complex topographic regions like the Antarctic Peninsula and many coastal regions of Antarctica (see Chapter 5). Furthermore, despite waveforms encoding data pertaining to the entire illuminated surface, ambiguity in the origin of different parts of the echo means these processing approaches extract only singular (or in the case of recent advances in multi-peak retracking approaches, several) measurements, which are focused primarily on retrieval of surface height at POCA. As such, these approaches fail to exploit a vast amount of information acquired by current and future missions, which has the potential to improve the quality and utility of the satellite radar altimeter record.

In this regard, efforts to extract more information from each waveform - for example, through advances in retracking and slope correction - have demonstrated that multiple elevation estimates can, in some cases, be recovered from a single waveform (Huang et al., 2024). These developments represent important progress, enabling partial use of previously discarded data and increasing measurement density in challenging regions. However, importantly, such methods still remain anchored in the logic of conventional pipelines, whereby they aim to distil the waveform down to a small number of elevation values.

Within this chapter, I therefore propose, develop and implement a fundamentally different approach to Level-2 SAR altimetry processing over ice sheets, which instead of focusing on resolving range ambiguity through reducing dimensionality, instead models this ambiguity directly. This is motivated by the knowledge that each SAR waveform encodes a distribution of plausible surface configurations. Thus, by training a model to learn this underlying distribution, I aim to develop a new method which is capable of predicting this topographic distribution directly from raw waveforms, without requiring the prevailing Level-2 processing steps of retracking or slope correction, or indeed any interferometric phase information. This new approach enables the extraction of significantly more information from non-interferometric measurements (comparable to that achieved by SARIn swath processing), and avoid

the limitations and assumptions associated with the retracking and slope correction processing steps. Given that billions of altimetric observations have been collected over the Antarctic Ice Sheet in recent decades, I suggest that deep learning may offer a viable path to learning these underlying topographic distributions - and in doing so offer a fundamentally different, data-driven Level-2 processing logic.

6.1.5 Deep Learning Frameworks for Topographic Retrieval from Radar Altimetry

Conventional SAR altimetry reduces each waveform to one or a few elevation estimates near POCA, discarding much of the signal. As a single non-interferometric waveform may correspond to many plausible surface configurations, this makes topographic retrieval an intrinsically distributional problem. This ambiguity reflects aleatoric uncertainty in the mapping from waveform to surface geometry, while any learned model also carries epistemic uncertainty in how that mapping is represented. A suitable framework must therefore address both.

For this study I adopt ResNet-RS as the backbone architecture. Bello et al. (2021) demonstrated that ResNet-RS combines strong predictive accuracy with training efficiency and architectural simplicity, while still offering the representational depth needed to capture the nonlinear structure of SAR waveforms. These properties make it a suitable foundation for a first application of a probabilistic framework to radar altimetry. Full architectural details are provided in Chapter 3.

On top of this backbone, I evaluate four complementary approaches to modelling the underlying distribution of plausible surface configurations given power waveforms, each of which are also described in depth in Chapter 3, with the main formulations recalled here for clarity:

1. **Gaussian likelihood regression (Russell et al., 2021)**. The network outputs the parameters of a multivariate Gaussian predictive distribution - a mean vector and covariance matrix - trained by minimising the negative

log-likelihood (Equation 3.1). The means provide central elevation predictions across the footprint, while the covariance matrix captures conditional aleatoric uncertainty and inter-variable correlations, together parameterising the distribution of plausible surface configurations consistent with the observed waveform.

2. **Evidential regression (Meinert et al., 2022)**. Rather than predicting a Gaussian likelihood directly, the network outputs the parameters of a higher-order prior over Gaussian likelihoods (Equation 3.9). The predicted means again provide central elevation estimates across the profile, while the inferred evidence quantifies how strongly the waveform supports those estimates. In principle, this allows aleatoric uncertainty - arising from the multiplicity of plausible surface configurations - and epistemic uncertainty - reflecting model confidence in the learned waveform-to-surface mapping - to be estimated jointly, without repeated sampling or ensembles at inference time.

3. **Quantile regression with Pinball loss (Koenker et al., 1978)**. A non-parametric approach in which the network directly estimates selected conditional quantiles of the across-track elevation distribution by minimising the Pinball loss (Equation 3.3). This yields interval-based descriptions of aleatoric uncertainty - capturing the range of surface elevations consistent with the waveform - without assuming a parametric form for the underlying distribution, though it does not define a full continuous distribution between estimated quantiles.

4. **Geometric quantiles (Garcia-Cardona et al., 2021)**. A related non-parametric approach using a geometric loss formulation to estimate ordered conditional quantiles of the across-track elevation profile while enforcing a non-crossing constraint (Equation 3.4). As with Pinball loss, the spread and asymmetry of the resulting quantile surfaces characterise the aleatoric uncertainty in the plausible surface configurations associated with a given

waveform.

When only aleatoric structure is predicted, I additionally estimate epistemic uncertainty using two widely adopted strategies. The first is Monte Carlo dropout, where dropout is applied during inference to generate an approximate posterior via repeated stochastic forward passes (Gal et al., 2016). The second is deep ensembles, in which multiple independently initialised ResNet-RS models are trained, and the variance across their predictions is taken as epistemic uncertainty (Lakshminarayanan et al., 2017). Alongside the evidential method, these strategies provide complementary measures of model confidence, particularly valuable in regions of sparse training data or out-of-distribution inputs. Full details of the training procedure, including dataset preparation, loss function, optimiser, learning rate schedule, and regularisation choices, are provided in Section 6.3.

Taken together, this dual uncertainty modelling supports a fully probabilistic approach to SAR altimetry - moving beyond deterministic, single-point elevation estimates toward a richer, distributional understanding. By explicitly testing several alternative probabilistic strategies, this study aims to establish a framework capable of learning and predicting the full range of possible surface configurations from non-interferometric radar waveforms. This reframing treats waveform ambiguity (the fact that a single non-interferometric waveform is consistent with many distinct across-track surface configurations, due to the absence of directional phase information) not as a limitation but as an information source, opening a path to enhanced topographic retrievals from both historical and future missions.

6.2 Data

6.2.1 CryoSat-2

CryoSat-2 (CS2) is ESA's dedicated ice altimetry mission, launched in 2010 under the Earth Explorer programme. Designed to provide long-term, high-precision

monitoring of polar ice sheets and sea ice, it has now collected over 15 years of continuous data and remains central to understanding of cryosphere-climate interactions. The mission, its SIRAL instrument, operating modes (LRM, SAR, SARIn), and data products are described in detail in Chapter 2.

CryoSat-2 was selected for this proof-of-concept study primarily due to its extensive SARIn range window of approximately 240 meters (European Space Agency, 2021). This larger range window is critical for mitigating waveform truncation, which occurs when topographic elevation within the satellite footprint varies beyond the available range, thereby enabling more complete capture of the illuminated surface response over complex terrain. This is in comparison to the Sentinel-3 mission, which is only able to fully capture ∼42.6% of the ice sheet due to its smaller SAR range window of approximately 60 m (see Chapter 5) (EUMETSAT, 2017). The longevity of CryoSat-2, with over 15 years of continuous data collection, totalling billions of measurements, is also beneficial in terms of offering a substantial dataset for training deep learning models. This large data volume enables the opportunity for more extensive data cleaning and preprocessing, which in turn supports improved model performance. Furthermore, the availability of interferometric data provides the future opportunity to benchmark predictions against state-of-the-art interferometric swath elevations (Chapter 7). Finally, as the dataset is based on SAR acquisitions, it also offers advantages over conventional LRM data by facilitating dimensionality reduction. Specifically, SAR enables the prediction of across-track swaths with reduced sensitivity to the along-track dimension, thereby simplifying the learning task.

For this study, I utilise the Level-1B Cryo-TEMPO baseline C product (available at `ftp://science-pds.cryosat.esa.int`) as my inputs, restricted to data over the CryoSat-2 SARIn zone of the Antarctic Ice Sheet spanning 2012-2021 and 2023.

6.2.2 REMA

The Reference Elevation Model of Antarctica (REMA) is a high-resolution digital elevation model (DEM) that covers nearly 98% of Antarctica (Howat et al., 2019). Created using stereophotogrammetric techniques applied to submeter resolution commercial optical satellite imagery, including data from Maxar, WorldView-1, WorldView-2, WorldView-3, and GeoEye-1, REMA provides a comprehensive topographic picture of the continent. The REMA mosaic was registered to satellite altimetry measurements from CryoSat-2 and ICESat, leading to an estimated absolute uncertainty of less than 1 m and relative uncertainties on the order of decimeters. In this study, the 100 m resolution REMA V2 mosaic (available at `data.pgc.umn.edu/elev/dem/setsm/REMA`) is used as target data for model training and as validation data for the final analysis.

6.2.3 Antarctic Coastline Mask

To ensure that my study focuses exclusively on terrestrial ice data and does not inadvertently capture sea ice of ocean data, I applied a coastline mask to the REMA dataset. The REMA V2 mosaic, which was compiled over several years, is not inherently masked to ocean-land/ice boundaries, necessitating the use of an external dataset for masking purposes. Thus, I utilised the MEaSUREs Antarctic Boundaries for IPY 2007-2009 from Satellite Radar, Version 2 dataset (available at `nsidc.org/data/data-access-tool/NSIDC-0709/versions/2`) to create this mask.

6.3 Methodology

This study aims to develop a deep learning framework capable of predicting full surface elevation distributions from non-interferometric SAR altimetry waveforms. As outlined in Section 6.1.3, these waveforms contain inherent ambiguity, reflecting multiple plausible surface configurations that traditional processing chains reduce to singular (or several) point estimates. My approach instead models this ambiguity

directly, incorporating both aleatoric and epistemic uncertainty to produce spatially distributed, probabilistic elevation predictions.

To train this framework, I compile a dataset of CryoSat-2 SARIn waveforms collected over the Antarctic Ice Sheet between 1st February 2012 and 31st December 2021. Pine Island Glacier (Figure 6.1) is withheld entirely from training and used as a geographically distinct validation region, allowing me to monitor overfitting and explicitly test how well the model generalises to an unseen area of the ice sheet. In addition, I reserve all data from 2023 as a temporally distinct test set, excluding 2022 to ensure sufficient separation between training and testing. Waveforms were retained only if their nadir coordinates, projected on an Antarctic polar stereographic grid, fell within the MEaSUREs IPY 2007-2009 coastline boundary. Further, additional filtering, such as removing records where the surface topography within the footprint exceeds the range window extent, are described in Section 6.3.1.2.

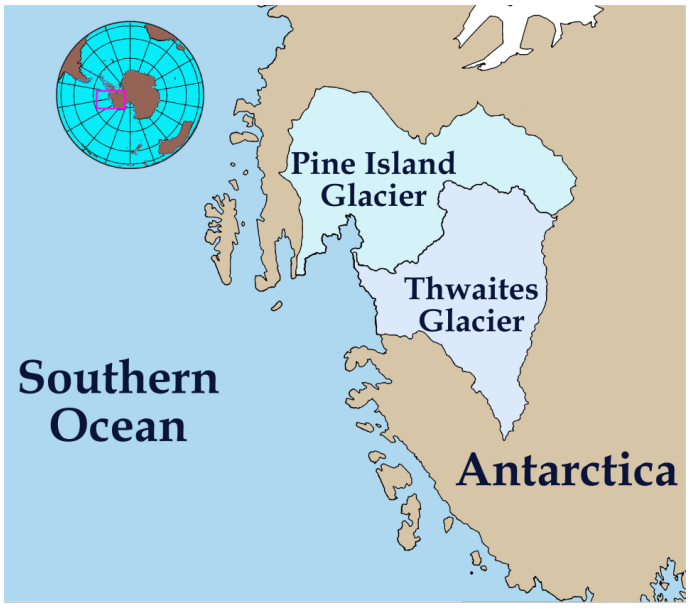

Figure 6.1: Location of Pine Island Glacier in West Antarctica. From Nwbeeson (2024).

The following sections detail the processing of (1) the input SARIn waveforms and REMA elevation profile targets, (2) the simulation of synthetic waveforms to reduce waveform-profile mismatches, and (3) the balancing of the training dataset

across surface slope and roughness. I then describe (4) the training procedure, and finally (5) the strategy for analysing the performance of the final framework. An overview of the complete pipeline is provided in Figure 6.2.

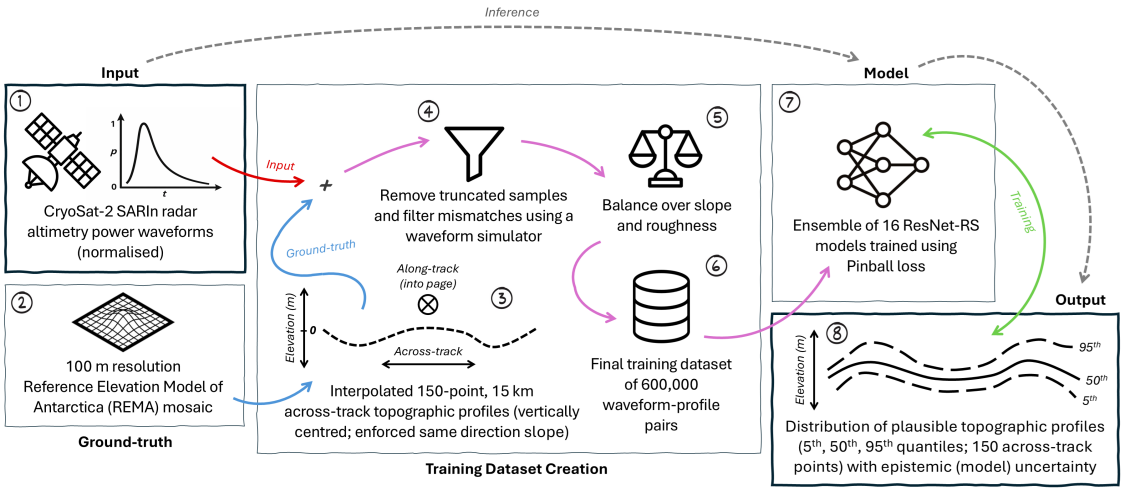

Figure 6.2: Overview of the deep learning framework for predicting topographic distributions from non-interferometric SAR altimetry waveforms. The pipeline shows eight key steps: (1) CryoSat-2 SARIn power waveforms are acquired and normalised (1024 bins); (2) REMA 100 m resolution elevation mosaic data is extracted from within the satellite footprint; (3) 150-point across-track topographic profiles are interpolated from REMA, vertically centered about zero, and enforced to slope in the same direction to eliminate mirror ambiguity; (4) input waveforms and target profiles are paired, with truncated waveforms removed and waveform-profile mismatches filtered using a simple waveform simulator; (5) samples are balanced across surface slope and roughness; (6) producing a final training dataset of 600,000 waveform-profile pairs (2012-2021); (7) an ensemble of 16 independently-initialised ResNet-RS models is trained using Pinball loss to predict the 5^{th}, 50^{th}, and 95^{th} elevation quantiles simultaneously across all 150 across-track points; (8) generating distributional predictions of plausible topographic profiles, with aleatoric uncertainty quantified through predicted quantile ranges and epistemic uncertainty through ensemble variance. During inference, only steps 1, 7, and 8 are required, with all other steps pertaining to model training. Outputs require the corrected range to be added for proper vertical placement and correct across-track orientation using a-priori DEM data.

6.3.1 Data Preparation

To generate paired waveform-topographic profiles suitable for training and evaluation, I combined CryoSat-2 SARIn altimetry with REMA digital elevation data, applying a series of preprocessing steps to ensure internal consistency, spatial alignment, and representativeness of surface conditions.

6.3.1.1 CryoSat-2 SARIn Waveform Acquisition

The SARIn L1B product data is split into numerous NetCDF files, each providing measurements along a single orbital arc corresponding to a continuous period during which CryoSat-2 operated in SARIn mode. Each month comprises several thousand distinct files, each of which consist of up to several thousand individual measurements, cumulatively amounting to hundreds of millions of measurements from July 2010 to the present day. In practice however, the product is self-consistent only as far back as February 2012, due to waveforms being oversampled by a factor of two after this point in order to avoid aliasing over specular surfaces (European Space Agency, 2021).

For each measurement within the AIS, I then extract the processed, multi-looked CS2 SARIn waveform, which consists of 1024 bins, capturing power returns within the ∼240 m range window (or $\sim 8 \times 10^{-7}$ seconds). Each waveform is normalised in the product (0-65535), and can be converted to power in Watts using the provided echo scale factor and echo scale power (Equation 6.1). I therefore additionally extract the echo scale factor and echo scale power variables.

$$W_w = W_n \cdot S_f \cdot 2^{S_p} \tag{6.1}$$

Where W_w is the waveform in Watts, W_n is the 0–65535 normalised waveform, S_f is the echo scaling factor, and S_p is the power scaling exponent.

In addition to the waveform, I also extract the satellite altitude for each measurement in meters and the return time of the pulse to and from the middle of the range window (window delay) in seconds. The return time (seconds) is then

converted to the distance from the satellite to the reference bin index (meters) using Equation 6.2.

$$r = \frac{t_{wd} \cdot c}{2} \tag{6.2}$$

Where r is the range (one-way distance) in meters, t_{wd} is the window delay, and c is the speed of light. The division by 2 accounts for the fact that the calculation is for the one-way distance, not the round-trip distance.

Before using the range measurements, it is necessary to apply geophysical and atmospheric corrections provided in the product. These include path delays due to dry tropospheric, wet tropospheric, and ionospheric effects; tidal effects; and variations induced by atmospheric pressure. As these corrections are provided at a native sampling of 1 Hz, I use nearest neighbour interpolation to oversample to the 20 Hz sampling of the measurement parameters.

Some corrections, specifically ocean tide and atmospheric pressure, are relevant only over floating ice and water, not land. To determine the classification of the surface below the satellite, I use the surface type variable, which categorises the surface into four types: ocean, lake/enclosed sea, ice, and land. For all non-land measurements, these additional corrections are applied. I also determine the satellite heading for each measurement in degrees, based on the difference in nadir coordinates from the current measurement to the next, additionally extracting the corresponding year, month, and time of each observation. In total, I process 262,076,191 observations over the AIS spanning 2012-2021 and 2023

6.3.1.2 REMA Acquisition

Although no ground truth data is available at sufficiently high resolution and over the entire continent, REMA serves as a suitable proxy in terms of providing a comprehensive and highly detailed description of topographic variability. Nonetheless, as a consequence of using REMA as the target dataset, it should be noted that the model is essentially trained to be a predictor of REMA rather than a direct regressor

of real-world topography; i.e. I am training the model to predict REMA profiles from the detected waveforms. For each record taken from the SARIn L1B product, I therefore extracted the REMA V2 100 m resolution DEM points contained within the satellite footprint, defined as a 15 km by 380 m rectangle in the across-track and along-track directions, respectively. This rectangle was centred on nadir, and orientated according to the satellite heading.

Subsequently, for each record, I interpolated the DEM points contained within the footprint onto a 1-D across-track profile at nadir, at 100 m resolution, consistent with the native resolution of the REMA mosaic, yielding 150 elevation points spanning 15 km across-track. This follows the methodology outlined in Chapter 5 for generating similar across-track profiles for exploring the performance of Sentinel-3 SRAL. Interpolation was performed using a piecewise cubic, C1 smooth, curvature-minimising interpolator (Clough-Tocher), with endpoints interpolated using nearest-neighbour. If interpolation failed, or if any point on the line had no DEM points within 250 m, the entire record was deemed unusable and omitted from the resulting dataset.

To identify instances of waveform truncation, the range to the satellite was measured for each of the 150 across-track points in every interpolated topographic profile. If any of these range measurements fell outside the range window, the entire record was omitted from any further processing. Considering the dataset as a whole, 16.6% of acquired records exhibited greater topographic variance than could be captured within the CryoSat-2 SARIn range window of approximately $\sim$240 m. In such cases, the waveform was truncated, resulting in the loss of information pertaining to surfaces within the beam footprint but outside of the range window, making prediction of the full topographic profile impossible.

REMA provides elevation measurements relative to the geoid, which is information not directly captured within the raw waveform. Instead, the waveform encodes elevation relative to the range window, of which the vertical placement varies between measurements. Without the corrected range, a model using the raw

waveform as input would lack the necessary information to predict elevation, only deviation from the reference index. Therefore, I vertically centred each topographic profile about zero by subtracting the reference range for each record. Converting back to elevation then involved simply adding the relevant corrected range to any topographic profile predictions post-inference.

Topographic profiles inherently exhibit a mirror ambiguity about nadir in the across-track direction: a profile mirrored about nadir in theory produces the same waveform. While it may be possible to characterise the general uncertainty in surface configurations from a given waveform, this ambiguity is effectively insurmountable for individual records, as it doubles the number of plausible configurations, with no clear information held within the waveform to determine this orientation. To reduce redundancy and ensure the model focuses on the true geometric relationship between waveform shape and underlying topography, I therefore enforced all surfaces to slope consistently in the same across-track direction. This can then be reverted post-inference using a priori slope information. As a whole, this process produced 220,935,604 and 21,778,482 waveform-surface topography (input-target) pairs over the AIS for 2012-2021 and 2023, respectively, where the surface topography component corresponds to a centred, 150-point REMA topographic profile representing across-track surface topography within the satellite beam footprint.

6.3.2 Dataset Cleaning and Preparation

Following data acquisition, a number of preprocessing steps were performed to ensure both internal consistency of the waveform product and spatial alignment with REMA. The following sections describe the preparation and selection of paired waveform-topographic profiles used for model training and evaluation.

6.3.2.1 Simulating Waveforms to Remove Mismatches

The REMA mosaic is a temporally aggregated model with a mean year of 2015 (Howat et al., 2019). Given that surfaces can change over time and the REMA formation process is not always perfectly accurate, it is essential to ensure that the target interpolated topographic profile corresponds accurately to the CS2 waveform. To maximise the quality of the training dataset - a critical determinant of model performance - I therefore developed a simple but efficient waveform simulator. This was used to assess the basic consistency between each REMA-derived 150-point topographic profile and its associated waveform, allowing me to retain only records where the simulated and observed waveforms matched within a defined threshold. Since the simulator was used solely as a filtering step, a lightweight and efficient implementation was favoured.

To evaluate whether a given REMA profile is a plausible surface configuration for an observed waveform, I simulated the expected return from that topography using a simple geometric ray-casting approach. I initialise a 128-bin empty waveform and cast 512 rays, evenly spaced in angle, from a virtual sensor located at the reference range provided in the CryoSat-2 product. Each ray intersects the 1-D REMA-derived elevation profile, and the corresponding waveform bin is determined by the ray's path length. Bin values are then incremented based on these intersections, weighted by the antenna gain as a function of angle from the antenna boresight (Equation 6.3). The resulting simulated waveform is normalised to the range 0-1 and upsampled to 1024 bins to match the format of the CS2 SARIn waveform.

$$G(\theta, \vartheta) = G_0 \exp\left[-\theta^2 \left(\frac{\cos^2 \vartheta}{\gamma_1^2} + \frac{\sin^2 \vartheta}{\gamma_2^2}\right)\right] \tag{6.3}$$

Where $G(\theta, \vartheta)$ is the antenna gain at polar angle θ and azimuthal angle ϑ, measured with respect to the antenna boresight. The constants $\gamma_1 = 0.0133$ and $\gamma_2 = 0.0148$ (radians) set the angular response in the across-track and along-track directions, and correspond to 3 dB beamwidths of approximately 0.66° and 0.70° respectively (Wingham et al., 2006a). Since I am modelling SAR waveforms using an across-

track elevation profile, I set ϑ to zero, using solely variation with polar angle. This is appropriate because the waveform simulator operates on a 1-D across-track elevation profile, which contains no along-track variation by construction. The azimuthal angle ϑ therefore carries no information in this context, and setting it to zero reduces the gain function to its across-track form. In the implementation, the constant G_0 (42 dB) is omitted as the simulated waveform is normalised, preserving only the relative directional characteristics of the antenna pattern.

To determine if a topographic profile is in sufficiently close agreement with its associated waveform, I measured the similarity between the simulated and actual waveforms using the Wasserstein distance. The Wasserstein distance, also known as the Earth Mover's distance, quantifies the minimum amount of work required to transform one distribution into another (Panaretos et al., 2019). This measure is particularly suitable for comparing waveforms, as it accounts for the distribution of values rather than just pointwise differences.

The simulator's parameters, including the number of bins and rays, were tuned using a grid search. This process involved minimising the Wasserstein distance on a small, balanced sample from 2015; a process which indicated an optimal bin size of 128 and 512 rays. Although I considered incorporating the surface normal angle from the satellite - since, in principle, the angle of incidence between the radar beam and the local surface normal modulates the return power, with surfaces tilted toward or away from the sensor reflecting differently than a horizontal surface - testing showed that this did not yield improvements and slightly increased run-time. Consequently, I excluded it from the final simulator.

To identify which waveform-REMA pairs to pass through to training, a relatively strict threshold Wasserstein distance of 0.0125 was selected, with the aim of balancing data retention and simulator accuracy. This intentionally strict threshold resulted in the discarding of 88.4% of records from the dataset formed in Section 6.3.1.2. Although this represents a large proportion of the data, the overwhelming priority is to ensure that the model is trained only on reliable and geometrically

consistent samples. Conversely, retaining incongruous topography and waveform pairings would be likely to degrade model performance by introducing noise or bias during training. The filtering step thus serves as a conservative but necessary safeguard, which is designed to still provide a sufficiently large and diverse dataset for effective model development.

6.3.2.2 Balancing for Slope and Roughness

Balancing the dataset across topographic slope and roughness is essential to ensure that the model is exposed to a representative range of surface conditions during training. This diversity enables the model to learn the nuanced relationships between waveform shape and underlying topography. While the vast majority of the Antarctic Ice Sheet is characterised by smooth, low-slope terrain (61.3% of records below 0.1° slope and 69.1% with less than 5 m roughness; Chapter 5), the most dynamic and scientifically important regions - such as ice sheet margins and coastal zones - tend to be far more topographically complex. Without accounting for this imbalance during the training process, the model would be likely to overfit to the most common, flatter conditions, and underperform in rougher, more variable regions. By enforcing balance, I therefore aim to improve the model's ability to generalise and make accurate predictions across the full range of surface conditions encountered across the ice sheet.

To balance the training dataset according to surface topography, I derived estimates of across-track slope and roughness for each sample using the 1-D adaptation of the SVD-based method described in Chapter 5. Specifically, I applied this approach to the REMA elevation profiles extracted in Section 6.3.2.1, fitting a line of best fit to each across-track profile. Slope was then defined as the angle of this line relative to the horizontal, while roughness was taken as the range of orthogonal residuals. This yielded per-sample estimates of slope and roughness aligned with the geometry of each across-track REMA segment. To balance the dataset using these two values, I used 50 bins each for slope and roughness, spanning 0-0.75° and

0-60 m, respectively, keeping the last bin open-ended. Although, using trigonometry and assuming a non-curving range window, I estimate that $\sim 1^\circ$ is the theoretical maximum for slope (above which a linear topographic profile would not be entirely encapsulated within the 240 m range window), the number of available records with a slope greater than 0.75° post-filtering is significantly reduced, comprising only 0.0519% of the filtered dataset. This represents just 15,991 records due to the conservative filtering threshold Wasserstein distance of 0.0125 chosen in Section 6.3.2.1. Similarly, for roughness, I found that 5.50% of the data had roughness greater than 60 m, comprising 59,845 records. Therefore, I chose these bounds to balance topographic representation and data availability.

For each year from 2012 to 2021, I then sampled, without repetition, from each combination of slope-roughness bins (2500 in total), ensuring I adequately captured the topographic variation seen across the Antarctic Ice Sheet. If there were no available samples, I increased the roughness bin search by ± 1 bin. If the full roughness bin range had been searched, I then increased the slope bin search by ± 1 bin and repeated. This process was repeated for each slope bin until all samples had been obtained. This bin-dilation strategy avoided repeated sampling - which in the inverse-mapping framework would otherwise impact the predicted distribution by artificially inflating evidence for very specific waveform-surface pairs - while still maintaining a balanced representation of slope and roughness conditions. Consequently, the maximum number of samples I could obtain annually was 60,000 (24 samples per bin-combination), above which the proportion of slope and roughness values became unbalanced. I therefore achieve a final fully balanced and filtered training dataset of 600,000 samples spanning the years 2012-2022.

6.3.3 Model Training

Having prepared a paired waveform-topography training dataset, I next turn to the design and training of deep learning models capable of learning the relationship between CryoSat-2 SARIn waveforms and the underlying surface topography. I

adopt ResNet-RS as the architecture, as Bello et al. (2021) demonstrated it to be competitive and efficient, while remaining relatively straightforward to implement. This made it a strong candidate for my purposes, striking a balance between performance and simplicity, which was particularly important for this first application of a probabilistic framework to radar altimetry.

6.3.4 Tuning

For tuning the ResNet-RS model, I employed an incremental strategy, starting with a simple configuration and gradually adding complexity, a commonly used approach in deep learning to build understanding of the problem and guide hyperparameter choices (Google, 2023). For each test I trained for 50 epochs. Although I observed that the model was occasionally still training at this epoch, performance up to this point provided a good basis for comparing hyperparameter choices while keeping computational times low. In order to have the model work in 1-D, I replaced any 2-D layers in the model with their 1-D equivalent in PyTorch. I used four years of the training dataset (2014-2018) to reduce training times, totalling ∼240,000 samples. The dataset was divided spatially into four quadrants across the Antarctic Ice Sheet (AIS) using a 4-fold cross-validation approach. For each fold, the model was trained on data from three quadrants. The remaining quadrant was split in half: one half served as a validation set, and the other as a test set. The validation set was used to monitor the model's performance on unseen data during training. Specifically, I evaluated the model on the validation set at the end of each training epoch and selected the model checkpoint corresponding to the epoch with the lowest validation loss once training ended. This approach helped prevent overfitting by ensuring that the final model used for testing was selected based on its performance on independent data, rather than on the training set. The final test set was then evaluated using this selected model. This process resulted in four models per test, each withholding a different spatial quadrant, providing a final averaged loss across all runs. Spatial quadrants were chosen instead of a random 25% split to avoid high correlations of

subsequent records along-track, ensuring that validation and testing data were well-separated in space and thus entirely unseen by the model. The 4-fold approach also mitigated the impact of poor weight initialisation by averaging performance across multiple folds, leading to a more stable and reliable assessment.

In terms of introducing a probabilistic component into my models, I tested (1) the Gaussian likelihood regression approach outlined by Russell et al. (2021), (2) the evidential extension proposed by Meinert et al. (2022), (3) the geometric quantile framework outlined by Garcia-Cardona et al. (2021), and (4) the Pinball loss-oriented quantile approach by Koenker et al. (1978). Based on preliminary tests, I encountered difficulties in converging to optimal solutions using both the Gaussian likelihood regression approach and the evidential extension. In both cases, the primary challenge was maintaining positive semi-definite covariance matrices, which are required for computing the log-determinant term in the loss function. This issue primarily arose from the models predicting increasingly small values in the covariance matrix - particularly in the diagonal variance components - which led to numerical instability and frequent failures during training. To address this, I implemented several stabilisation strategies that are commonly used in probabilistic deep learning to ensure well-behaved uncertainty estimates. First, I added a small constant (1×10^{-6}) to the diagonal variance elements to prevent them from collapsing toward zero - a standard regularisation technique that preserves numerical stability while minimally impacting model expressivity (Hastie, 2020). Second, I replaced the exponential activation with softplus for the variance terms, and tanh with softsign for the Pearson correlation coefficients. These alternatives are smoother and saturate more gradually, which improves gradient behaviour and helps prevent extreme values that would otherwise destabilise training (Shridhar et al., 2019; Szandała, 2021).

In addition, I implemented a custom log-determinant function to address the instability of PyTorch's default operation at very low values, where numerical precision issues arise. For weight initialisation, I adopted Xavier (Glorot) initialisation, a widely used method in convolutional neural networks (Glorot et al.,

2010). This technique maintains gradient scales across layers by drawing weights from a distribution with zero mean and a variance determined by the number of input and output units. As a result, it preserves signal variance throughout the network and supports stable, effective training, particularly in deep architectures. Secondly, ReLU activations were replaced with Leaky ReLU, which allows a small, non-zero gradient for negative inputs (Xu et al., 2015). This adjustment helps maintain gradient flow and can improve both convergence and overall performance in certain scenarios. Finally, I observed that batch normalisation layers were contributing to instability, likely due to batch-to-batch variation in predicted uncertainty components. I replaced these with group normalisation layers, which is an alternative introduced by (Wu et al., 2018) that addresses some limitations of batch normalisation, especially when dealing with small batch sizes. Unlike batch normalisation, which normalises across the batch dimension, group normalisation divides channels into groups and normalises within each group by computing the mean and variance. This method is independent of batch size and helps mitigate performance degradation in small batch regimes, ultimately enhancing generalisation. While these changes did not yield stable convergence for the evidential learning approach, they did enable stable training under the Gaussian likelihood regression framework. However, despite improved optimisation stability, the resulting models still produced suboptimal validation loss values, suggesting limited performance.

Fortunately, no such issues arose when training using the two multivariate quantile loss functions (Garcia-Cardona et al., 2021; Koenker et al., 1978). Nonetheless, despite my efforts, the geometric approach outlined by Garcia-Cardona et al. (2021) also resulted in poor final validation loss values. In contrast, I observed promising preliminary performance with the Pinball loss function leading me to select it as the final loss function for my model. Doing so, I therefore target quantile predictions, making no parametric assumptions.

To train the ResNet-RS model using the Pinball loss function, I targeted the 5^{th},

50^{th}, and 95^{th} quantiles simultaneously across the entire topographic profile. This was achieved by summing the loss for all three quantiles and then averaging over all 150 points in the profile to attain a single loss value per record. Although I attempted to learn the quantiles separately, this approach resulted in negligible performance improvements and significantly increased computational overhead during both training and inference.

I largely followed the training regime used by Bello et al. (2021) although modified it slightly by employing the more sophisticated RMSprop optimiser, as used by Tan et al. (2020) and EfficientNet (Chapter 3). Additionally, I opted to keep several of the adjustments made when stabilising training for the probabilistic regression and evidential learning approaches; namely, using Leaky ReLU activation functions and Xavier initialisation, the latter of which is in contrast to the pre-trained weights used by Bello et al. (2021). As my 1-D ResNet-RS architecture had to be trained entirely from scratch, Xavier initialisation was adopted as a standard and robust alternative to partial pre-training. Leaky ReLU was retained as it had been shown to improve optimisation stability in the prior probabilistic frameworks. As with ResNet-RS, I implemented a cosine learning rate scheduler with a linear warmup, increasing the learning rate linearly from 0 to a maximum value over 5 epochs, followed by a cosine decay to zero over the remaining 45 epochs. This approach allows the model to explore the loss landscape broadly in the initial phases and then make finer adjustments towards convergence, enhancing stability and performance (Loshchilov et al., 2017; Goyal et al., 2018). Early stopping with a patience of 25 epochs was also included to avoid unnecessary computation; essentially stopping training if validation performance did not improve for 25 epochs. As with Bello et al. (2021), I also employed dropout after the final global average pooling layer, stochastic depth, and weight decay applied to the weights in convolution and fully connected layers.

I began model tuning by conducting preliminary tests on various data preprocessing strategies and input configurations using an untuned mixture of the original

ResNet-RS and ResNet-D hyperparameters. These initial parameters were broadly selected through brief, exploratory tests aimed at identifying the rough boundaries of the hyperparameter search space. My goal was to first determine which data configurations (such as waveform scaling, standardisation, and covariate inclusion) led to the most stable or promising performance. This helped minimise variability and complexity during subsequent tuning phases, allowing for a clearer attribution of performance changes to hyperparameter adjustments. To reduce training time, I used a lightweight model configuration, with a depth of 101 layers and a width factor of 0.25, both identified through quick testing. Although I tested a broad set of approaches, I focus here on the most impactful for brevity.

I first compared normalising the input waveform (0-1) against converting it to Watts. Normalisation is a common input transformation in training neural networks as it helps mitigate varying input magnitudes leading to gradient problems during optimisation (Goodfellow et al., 2016). Given the very low resulting values when converting to Watts, I also tested both scaling (to a mean of approximately zero) and standardising (subtracting the mean and dividing by the standard deviation).

In theory, the waveform amplitude (as given in Watts) is important because the return power (Sigma-0) contains useful information about surface conditions, which could be lost if waveform normalisation is applied. To address this, I also explored adding waveform power scaling variables as additional covariates alongside the normalised waveform. Specifically, I tested two methods: (1) concatenating this variable to the end of the flattened layer before the final fully-connected layer and (2) inputting it into its own fully connected layer, and then multiplying this with the final fully-connected layer. Additionally, I examined incorporating the total atmospheric corrections as extra covariate information, following the same methods used for the power scaling variables. I hypothesised that these corrections, which can shift returns by meters, might provide valuable context since the model otherwise lacks knowledge of these external factors.

I also investigated the impact of smoothing the waveform using a Savitzky-Golay

filter with a window length of 9 and a polynomial order of 3, a technique used for noise reduction in CS2 Cryo-TEMPO Land Ice processing baselines (Muir, 2024). In addition, I explored standardising the input waveforms across the training set - that is, subtracting the mean and dividing by the standard deviation at each waveform bin (computed over the full training set). This form of per-bin standardisation is commonly used in deep learning to improve convergence and reduce sensitivity to scale differences across features (LeCun et al., 2012).

Further, I tested different resolutions for the output topographic profiles, specifically 300 m (50 points) and 500 m (30 points), and varied the across-track distance considered, using 10 km and 5 km. Lastly, I assessed the necessity of enforcing all topographic profiles to slope in the same direction across-track, as outlined in Section 6.3.1.2, by comparing training results with and without this enforcement.

As a result of these preliminary tests, I found that the full 15 km topographic profiles, with the original resolution of 100 m (150 points), and enforced to slope in the same direction, produced the best results, attaining lower validation loss values. Additionally, the raw, non-smoothed, normalised (0-1) waveform outperformed other variations. This was likely because (1) the waveform is highly responsive to surface features, so smoothing essentially removes valuable information, and (2) the network struggles to learn meaningful geometric features when the waveform amplitude varies significantly between records when converted to Watts. I also discovered that adding covariate information (atmospheric corrections and/or scaling variables) hindered performance (higher validation loss), likely due to providing easy avenues for overfitting during training.

Following this, I then conducted a larger-scale training regime. While a grid search would have provided a more exhaustive exploration, I opted for a sequential greedy search due to the impractical number of possible permutations to cover comprehensively (see Table 6.1 for parameters tuned). In this approach, I began with the initial choices from the preliminary tests and then systematically varied

the parameters in turn. After each step, I advanced the best-performing choice to the next round of tuning.

It is generally accepted that, as long as key hyperparameters such as the learning rate and regularisation terms are well-tuned and the number of training steps is sufficient, the final performance should be attainable regardless of the batch size (Shallue et al., 2019). Despite this, my experiments revealed that smaller batch sizes significantly improved validation performance, with alternative parameter configurations failing to match this improvement. As such, in spite of increased training times, I opted for a smaller batch size of 16. This observation aligns with findings in the literature, where lower batch sizes are known to enhance generalisation by creating flatter loss landscapes due to the noise introduced in gradient estimation (Keskar et al., 2017). As a consequence of this smaller batch size, I replaced my batch normalisation layers with group normalisation layers, which further improved performance. Following this tuning process, I obtained the final parameter choices detailed in Table 6.1.

To train the final model, I then took the full dataset of 600,000 samples spanning 2012-2021. For validation data, instead of the 4-folds approach taken in tuning, I withheld data acquired over the entire Pine Island Glacier basin, chosen as a representative subset of the full ice sheet which covers a range of topographic complexities (Otosaka et al., 2023; Park et al., 2013; Wingham et al., 2009). This allowed the final model to observe a far broader distribution of training samples. Over Pine Island, I selected 6,250 samples - the maximum number possible while retaining a balanced subset (as per Section 6.3.2.2). Here, balancing the validation data is equally important as in training, to avoid biasing model evaluation toward any particular surface type, topographic regime, or temporal window. To reduce sensitivity to poor initialisation, I trained four final models for 200 epochs with different random seeds and selected the one with the best validation performance, similar to my tuning approach.

Hyperparameter	Final Tuned Value
Model depth	3, 4, 23, 3 (ResNet-101)
Bottleneck channels	16, 32, 64, 128 (0.25× ResNet default)
Expansion factor	4 (ResNet default)
Batch size	16
Weight decay	0.0001
Optimiser momentum	0.1
Squeeze-and-excitation reduction ratio	0.125
Stochastic depth ratio	0.2
Number of GroupNorm groups	8
Dropout ratio	0.0
Stem width	64
Maximum learning rate	0.001

Table 6.1: Model hyperparameters and their final tuned values.

6.3.5 Epistemic Uncertainty Quantification

To estimate epistemic uncertainty, I first explored using a Monte Carlo dropout approach. I started by placing dropout layers after every convolutional layer. However, I found that the best results were achieved by applying dropout solely to the layer before the final fully connected layer, aligning with the placement in the original ResNet-RS model.

To tune the dropout rate, I generated quantile predictions for the Pine Island validation set using my final model. I tested 75 equidistant dropout rates, ranging from 0 to 0.75, where for each, I ran the model for 30 Monte Carlo samples. For each tested dropout rate, I calculated the mean and standard deviation of the quantile predictions across the Monte Carlo samples.

Next, I separated the predictions into groups of 50 each. For each set of 50 predicted quantiles in each group, I determined the number of individual points (150 across-track) in the target topographic REMA profile that fell within a given

quantile prediction (5^{th}, 50^{th}, and 95^{th}) $\pm$ the associated epistemic uncertainty as given by Monte Carlo dropout. This provided me with three values per model prediction (50×3 per group). For each group of 50, I then evaluated the overall proportion of these counts that correctly encompassed the expected proportion of data for each quantile (i.e., 5% below the 5^{th} quantile, 50% below the 50^{th} quantile, and 95% below the 95^{th} quantile). Assuming that epistemic uncertainty followed a Gaussian distribution, I tuned the dropout rate by selecting the value that minimised the difference between the observed proportion and the expected 68% (captured by one standard deviation).

Unfortunately, this approach did not yield a clear optimal dropout rate, and visually, the quality of the epistemic uncertainty was poor in all cases. This outcome aligns with findings in the literature, where Monte Carlo dropout often produces low-quality uncertainty estimates in deep learning networks (Huang et al., 2025a; Liu, 2021; Ovadia et al., 2019; Detlefsen et al., 2019; Folgoc et al., 2021). Therefore, I instead opted for an ensemble approach despite the associated increased computational demands. While 5-10 ensemble members are typically sufficient (Lakshminarayanan et al., 2017), I chose 16 members to ensure robust epistemic uncertainty estimation. Since my approach aims to predict aleatoric uncertainty in surface configurations, cleanly separating epistemic uncertainty (model confidence) from aleatoric uncertainty (true variability in plausible surface configurations) is critical for reliable uncertainty quantification. The ensemble of 16 models substantially outperformed Monte Carlo dropout, providing significantly more stable epistemic uncertainty estimates. Training and validation loss curves over the ensembles can be found in Figure 6.3, and details on the final architecture can be found in Figure 6.4.

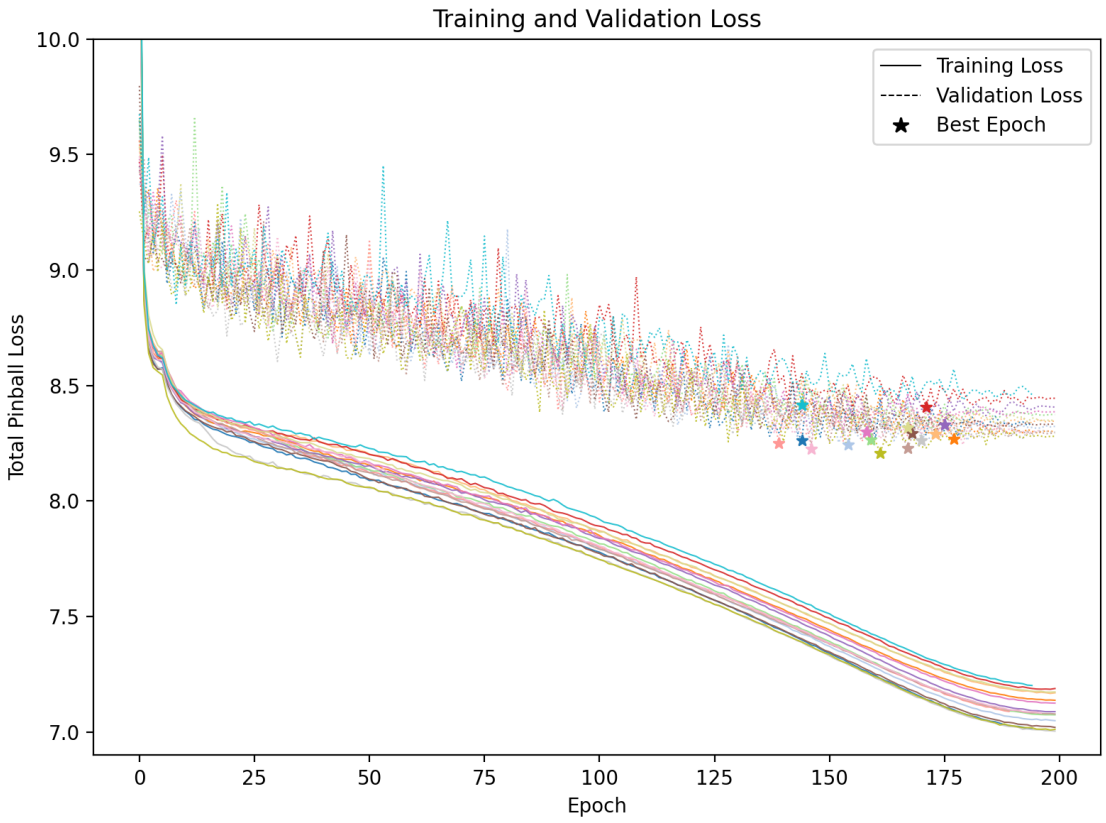

Figure 6.3: Training and validation loss curves for each of the 16 models in the final ensemble. Each model is shown in a different colour, with solid lines representing training loss and dashed lines representing validation loss. Stars mark the epochs at which the final models were selected, based on early stopping after 25 epochs without improvement in validation loss.

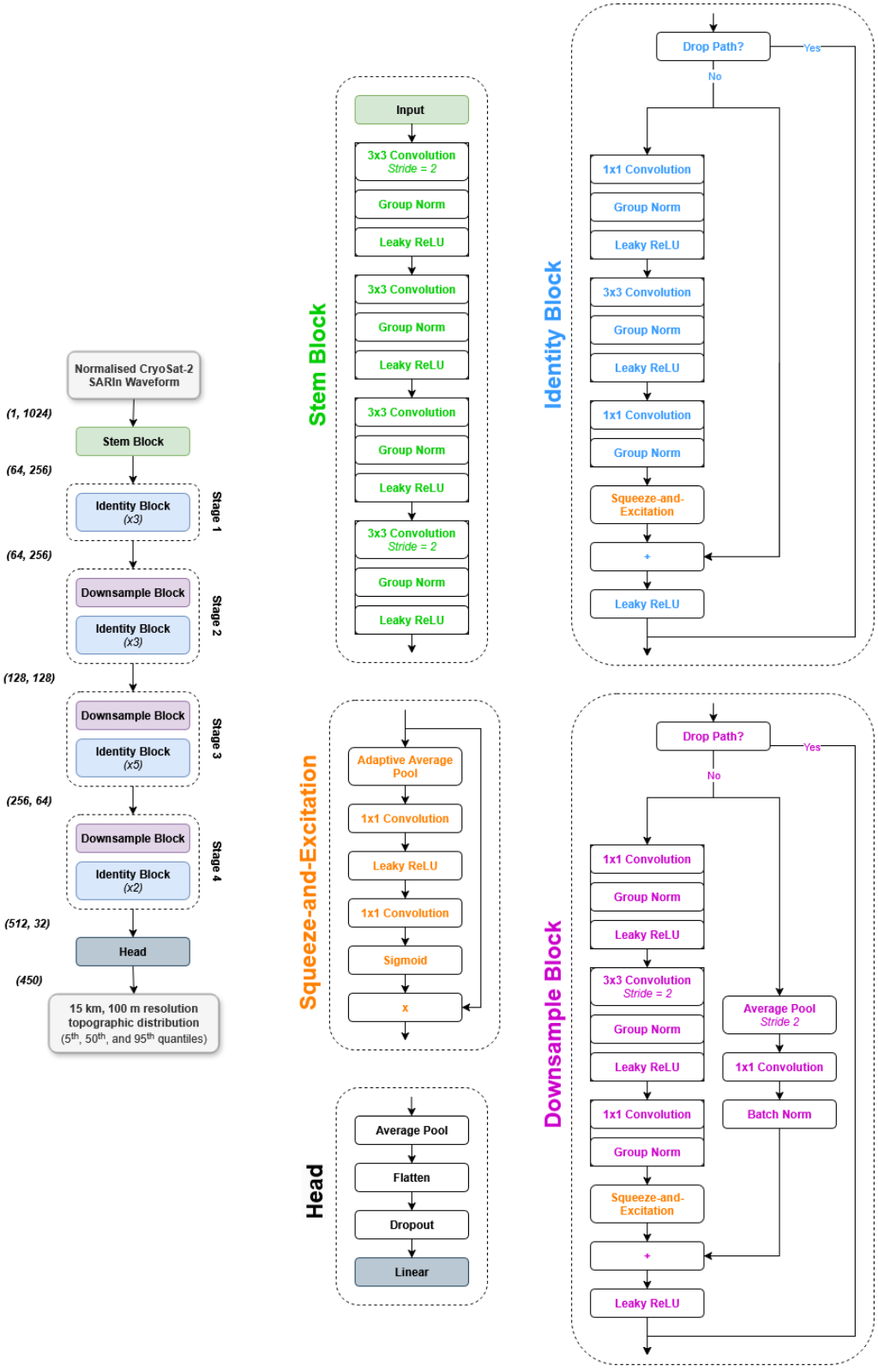

Figure 6.4: ResNet-RS architecture used for each model in the final ensemble, showing inputs, outputs, and dimensionality at each stage.

6.3.6 Inference and Testing

To analyse the performance of the trained ensemble, I take in unseen, normalised CS2 SARIn waveforms, predicting the underlying topographic distribution (aleatoric uncertainty) for each record over my 16 ensemble members. As my final predictions, I then take the mean across the individual ensemble predictions, and the standard deviation as the measure of epistemic uncertainty. While this study uses CryoSat-2 SARIn waveforms, the methodology is designed for conventional SAR waveforms, where no interferometric phase information is assumed to be available to resolve ambiguities in surface configuration. Accordingly, the interferometric phase contained within the SARIn data is disregarded for the purposes of the method, with analysis focusing solely on the power waveform; the phase information is instead retained for comparison in Chapter 7.

For evaluation, I use the complete 2023 AIS dataset prior to balancing and simulated waveform mismatch filtering (Section 6.3.1.2), comprising 21,778,482 records. This enables me to evaluate predictions over the completely unseen topography of the Pine Island Glacier catchment, in addition to acquisitions over the remainder of the ice sheet from the temporally separate year of 2023. The latter element is intentionally designed to mimic an operational scenario in which the model is initially trained on a balanced subset of early-mission data and then applied subsequently to all future acquisitions within an operational processing scheme.

Notably, the training dataset was curated across slope (0-0.75°) and roughness (0-60 m) ranges, with open-ended final bins. Although some higher slope ($> 0.75°$) and roughness values (>60 m) were included, they were significantly underrepresented in the training dataset, particularly near the upper bounds. As such, performance metrics in these higher slope and roughness regimes - which are more prominent in the testing dataset - offer valuable insights into the model's ability to generalise beyond its primary training distribution. Consistent performance trends in these out-of-distribution areas would suggest successful generalised learning of the underlying task - namely, inverse mapping of waveforms to surface distributions -

without overfitting to data found in these specific slope and roughness ranges.

As outlined in Section 6.3.2.1, all model outputs initially have slopes of the same positive sign (> 0), reflecting a training constraint in which profiles were restricted to a single across-track direction to reduce the space of possible surface configurations. This constraint is necessary because slope direction cannot be inferred uniquely from the waveform alone: profiles mirrored about nadir produce indistinguishable returns, making the ambiguity effectively unlearnable (Section 6.3.4). This left-right mirror ambiguity is particularly severe - without constraining it, the problem becomes effectively intractable - so it is addressed explicitly in this manner. Accordingly, slope direction is resolved only after inference using a priori estimates derived as described in Section 6.3.1.2, under the assumption that slope orientation remains stable between the REMA acquisition and the altimetry observation - a reasonable approximation over the timescales considered. More broadly, this ambiguity is one example of the inherent non-uniqueness of the inverse problem; the probabilistic framework addresses this by predicting a distribution of plausible surfaces rather than a single deterministic solution.

To assess the performance of my ensemble, I use the Prediction Interval Coverage Probability (PICP), which tests whether the nominal confidence level (e.g. 90%) matches the fraction of observations falling within the predicted intervals (Sluijterman et al., 2024). Formally, for a set of N predictions, PICP at nominal level α is defined as:

$$\mathrm{PICP}_{[\alpha_{\mathrm{lo}},\alpha_{\mathrm{hi}}]} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left[\hat{q}_{\alpha_{\mathrm{lo}},i} \leq y_i \leq \hat{q}_{\alpha_{\mathrm{hi}},i}\right] \tag{6.4}$$

Where y_i is the i-th observed elevation, and $\hat{q}_{\alpha_{\mathrm{lo}},i}$ and $\hat{q}_{\alpha_{\mathrm{hi}},i}$ are the predicted lower and upper quantiles of the interval. For one-sided cases, this reduces to the fraction of observations falling below a given quantile $\hat{q}_{\alpha,i}$, with a well-calibrated model achieving a fraction of α. This choice reflects the fact that my framework produces full predictive distributions rather than single elevation estimates, so conventional error metrics based on pointwise elevation differences to REMA are not sufficient.

Instead, PICP directly evaluates whether the predicted quantiles achieve their intended statistical coverage. Specifically, I evaluate four cases: the fraction of observations falling below the 5^{th} percentile (target: 5%), above the 95^{th} percentile (target: 5%), below the 50^{th} percentile (target: 50%), and within the central 5^{th}–95^{th} interval (target: 90%). These cases together probe calibration across the full predicted distribution, from the tails to the central interval.

In calculating PICP, to mitigate waveform-REMA mismatches, I filtered for simulated waveform mismatches (Section 6.3.2.1). Here I used a slightly less conservative Wasserstein distance of 0.075 - as opposed to 0.015 when generating the training dataset - to balance the removal of gross mismatches whilst maintaining sufficient data for reliable trend extraction. This adjustment resulted in 14,546,554 records, a reduction of 33.5%. Because PICP requires a large number of samples and only measures coverage in aggregate, it can mask systematic miscalibration under specific conditions (Sluijterman et al., 2024). To address this, I additionally stratify PICP by space, slope, and roughness, which tests whether prediction intervals remain valid across varying surface conditions. Such an evaluation is critical for judging the overall reliability and potential limitations of my method when extended to continent-wide monitoring.

6.4 Results and Discussion

Here, I analyse the final ensemble outputs as well as results corresponding to the various evaluation metrics outlined in Section 6.3.6 over the full 2023 Antarctic Ice Sheet (AIS) testing dataset, as detailed in Section 6.3.1.2, and comprising 21,778,482 records. In order, I present and assess (1) a randomly-selected sample of ensemble predictions, (2) 5-95^{th} distributional widths (to assess the degree to which surface characteristics impact distributional certainty), (3) ensemble epistemic uncertainties (to assess the degree to which surface characteristics impact model certainty in the predictions themselves), and (4) the overall model performance as given by PICP. I

then include two auxiliary diagnostics: (5) Pinball loss, to provide a complementary view on distributional accuracy, and (6) differences between the predicted median and REMA, to quantify systematic offsets. Each of these components are reviewed in-turn, and additional analysis is performed to compare them to ice sheet slope and roughness. Finally, I examine spatial patterns by binning records into 5 km grids and averaging within each bin, with 10 km grids taken in the case of PICP to ensure sufficient sample volume.

6.4.1 Random Samples

First, I visually inspected a number of randomly-selected outputs from my ensemble, to assess the topographic profile predictions made by my trained ensemble, based only on unseen waveform data from 2023. A selection of these sample profiles are shown in Figure 6.5, ordered such that average Pinball loss with respect to REMA - which measures the accuracy of predicted quantiles by penalising both over- and under-estimation in proportion to the error magnitude (Chapter 3) - increases down the figure. To provide context regarding where these rows lie with respect to the overall loss distribution, I include the percentage of records over the AIS with average losses below the maximum bounds for each row (see Section 6.4.5 for an in-depth analysis on these loss distributions). Records flagged as REMA-waveform mismatches (>0.075 Wasserstein distance between actual and simulated waveforms) were excluded from random selection.

In the random samples shown in Figure 6.5, I find overall highly encouraging model performance in terms of its ability to be able to predict the unseen REMA topographic profiles from the CryoSat-2 power waveforms alone. Particularly for the lower range of loss values, which prevail over most of the ice sheet (top row of Figure 6.5; Pinball loss representative of 71% of ice sheet records), the model predictions accurately resolve the dominant large-scale footprint topography associated with surface slope, in addition to the more modest undulations in topography. This, in itself, is an achievement, when considering the waveforms which constitute the

source of these predictions, and the lack of interferometric phase information.

As loss increases (e.g. Figure 6.5c-f), the surface topography becomes more rugged and, alongside this, the radar waveforms become far more complex, with multiple peaks, irregular shapes, and steps often apparent in the leading edge. Nonetheless, the model is still able to accurately represent the overall shape of the topography within the illuminated footprint, including where it deviates from a broadly uniform slope (e.g. in cases where there are steps or curvature in the topography, as evident in Figure 6.5d-f); albeit it may mislocate the position or gradient of the step, or miss more subtle small-scale topographic features. Importantly, however, the 5^{th}-95^{th} quantiles of the model (shown by the orange dashed lines in Figure 6.5) successfully capture the REMA profiles in the majority of cases, indicating that the model appropriately quantifies the underlying topographic distribution. For example, in panel 6.5b, the model exhibits low aleatoric uncertainty (i.e. the orange dashed lines converge) in its prediction of the surface elevation in the flat surface region close to nadir (across-track distance of $\sim +1500$ m); whereas in panel 6.5d, the model exhibits high aleatoric uncertainty (i.e. the orange dashed lines diverge) in its ability to predict the steep 100 m step in elevation at $\sim$ -4000 m across-track distance from nadir. In the latter case, it is unsurprising that the model attributes high uncertainty to this feature, given that it occurs over a distance that is smaller ($\sim$ 1500 m) than the across-track resolution of the altimeter. However, what is important is that the model successfully recognises this, and is able to assign appropriate uncertainty to this part of the profile accordingly. Thus it provides a credible confidence estimate to the user. It is only within the very extreme values of Pinball loss (Figure 6.5g-h; comprising only 0.3% of records ice sheet wide) that the model fails to assign appropriate uncertainty bounds to at least part (Figure 6.5g) or all (Figure 6.5h) of the profile. In the latter example, it is clear from the waveform itself that it contains little useable geophysical information, and is likely associated with a case where the altimeter is failing to track the ice surface.

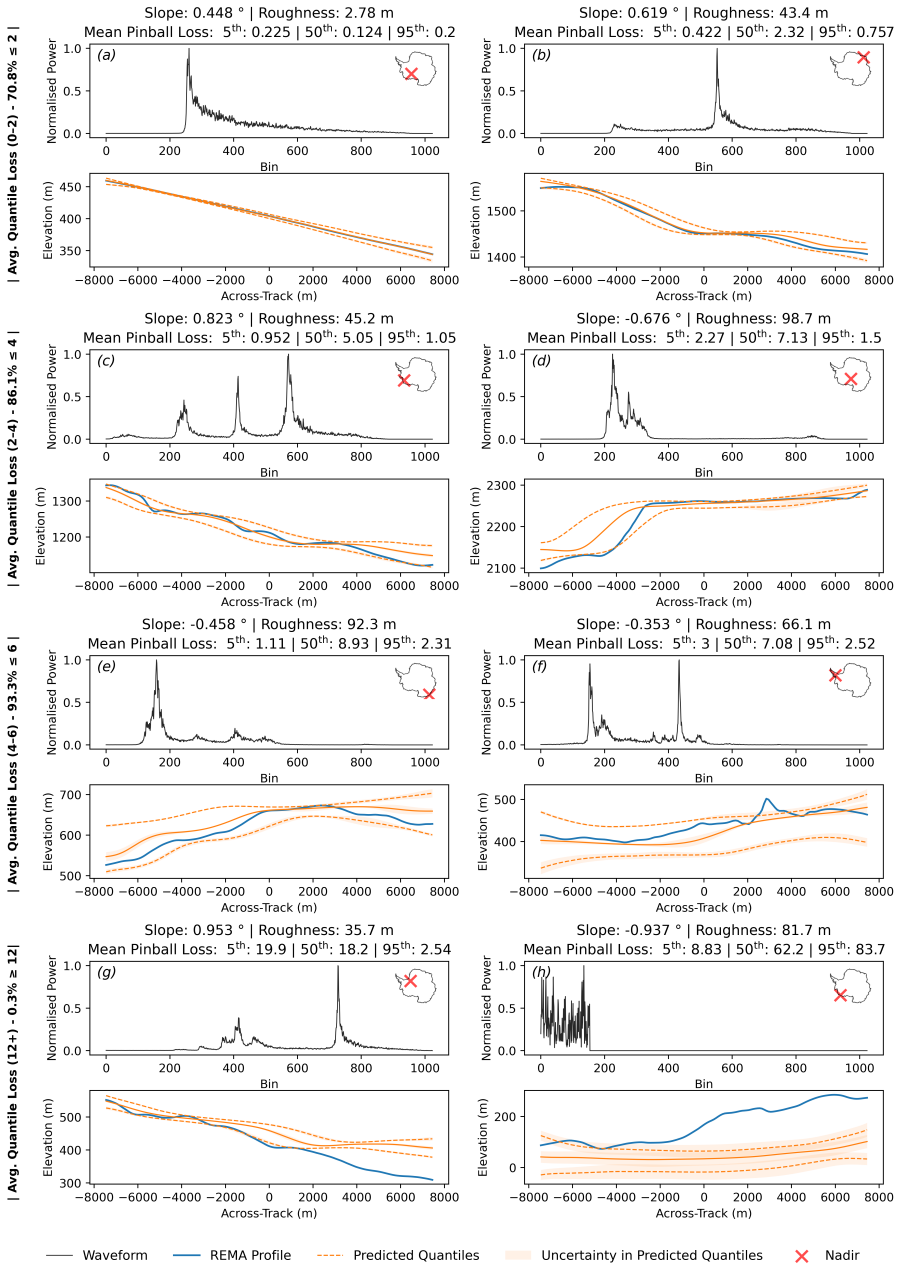

Figure 6.5: Randomly selected ensemble predictions from unseen waveforms over the Antarctic Ice Sheet (2023). Predictions are derived solely from the normalised SARIn power waveform amplitude, making the input effectively analogous to a conventional SAR waveform; interferometric capability is not used. Across-track slope direction is resolved post-inference using a priori slope data (Section 6.3.6). Each panel shows the input waveform (top) and corresponding unseen surface topography (bottom), with the REMA reference profile (blue), predicted quantiles (orange), and epistemic uncertainty (shaded orange). Rows are ordered by increasing average Pinball loss relative to REMA; percentages indicate the fraction of AIS records falling below the upper loss bound for each row. Insets show record location (red cross) on the AIS outline.

Alongside the mean ensemble predictions for each quantile (in orange), I present the epistemic uncertainty as semi-transparent bands around the corresponding quantiles, given as $\pm$ 1σ of the ensemble predictions. As expected, this ensemble spread is generally much smaller than the 5-95% quantile range, highlighting that most of the predictive uncertainty arises from aleatoric ambiguity in the waveform-surface mapping, rather than instability of the models themselves. This uncertainty tends to vary consistently across quantiles, as all are sampled from the same underlying distribution. In some cases the model assigns low epistemic uncertainty to predictions that still lie far from the true value. This does not necessarily indicate poor uncertainty estimates, but reflects the probabilistic nature of the outputs: by design, a fraction of observations are expected to fall outside the predicted intervals. This can be seen in Figure 6.5g, where the model appears confident in its distributional prediction (low epistemic uncertainty) but incurs high Pinball loss because the observed profile lies well outside the predicted bounds. In contrast, Figure 6.5h also shows high loss, but here the larger epistemic uncertainty indicate the model was far less confident in its estimate, caused by the noisy and truncated waveform.

Throughout the samples shown in Figure 6.5, it is also noticeable that the 5-95$^{\text{th}}$ distributional width consistently expands towards the edges of the profiles. This can be attributed to a drop off in the returned signal towards the edges of the footprint, following the gain pattern equation (Equation 6.3), with information pertaining to nadir dominating the final waveform. As such, resolving the surface topography in these peripheral regions is more difficult, translating to greater aleatoric uncertainty. At the same time, the distributional width often narrows around the point of closest approach, where the leading-edge return provides clearer, less ambiguous information that constrains the distribution more tightly than later, mixed returns. This is discussed in more detail in Sections 6.4.2 and 6.4.3. Within the selected profiles, the distributional width also tends to increase with slope and roughness. Flat regions exhibit narrower ranges, indicating greater distributional

certainty, as sharp, well-defined waveforms typical of flat surfaces have fewer possible topographic origins. In contrast, complex terrain produces waveforms that can arise from numerous surface configurations, increasing the width of the distribution. This is explored in greater, more quantitative detail in Section 6.4.3.

In general, I observe that the model tends to produce relatively smooth distributions, with gradual changes across the predicted profiles rather than sharp inflections. This reflects not only the pulse-limited resolution of the altimeter, across-track, but also the probabilistic nature of the framework; because the model estimates the range of plausible surfaces that are consistent with the waveform, localised extremes are dampened across the distribution. As a result, narrow peaks or abrupt valleys in the true surface often protrude beyond the predicted intervals, as illustrated in Figure 6.5f. These discrepancies do not necessarily indicate failure of the approach, but rather highlight a structural tendency of the model to prioritise smoother, broadly consistent solutions over highly localised features. Incorporating prior topographic information (e.g. from a coarse DEM or slope estimate) could help constrain the predicted distribution and sharpen these features, though this was not pursued here to avoid imposing a fixed, time-dependent prior on the solution.

6.4.2 Ensemble Epistemic Uncertainty

Within this section, I move beyond individual prediction examples (Figure 6.5) to develop and analyse a larger-scale characterisation of epistemic uncertainty across the full test set. Represented as the standard deviation of the ensemble predictions, epistemic uncertainty captures the model's internal confidence - namely, quantifying how much the model predictions vary when sampling from the learned distribution. This reflects uncertainty in the model itself, rather than ambiguity in the input waveform (aleatoric uncertainty). By analysing how the ensemble epistemic uncertainty varies across predicted quantiles and between regions (Figure 6.6), I aim to gain insight into how the model expresses confidence under different surface conditions.

Figure 6.6a shows the overall distribution of epistemic uncertainty across the Antarctic Ice Sheet SARIn zone for each of the 5^{th}, 50^{th} and 95^{th} quantiles, and demonstrates clear differences in epistemic uncertainty between the quantiles. The uncertainty in the 95^{th} quantile follows a unimodal, positively skewed distribution, with a median of 1.48 m and median absolute deviation (MAD) of 0.531 m. Similarly, uncertainty in the 50^{th} quantile predictions appears positively skewed towards lower values, but has a much broader shape, attaining a median of 2.17 m and MAD of 0.910 m. Conversely, uncertainty in the 5^{th} quantile predictions appears somewhat bimodal and is much broader than either of the other quantiles, attaining a median of 3.19 m and MAD of 1.34 m. The peaks correspond to higher peak at around 1-2 m and a lower, broader peak at 3-4 m in uncertainty.

Intercomparing the three different quantiles, the distributional differences between quantiles follow a clear trend, with greater consistency (i.e. lower variance) across the ensemble of models at the top end (95^{th} quantile) of the distribution, compared to the middle (50^{th} quantile) and then the bottom (5^{th} quantile). This pattern is likely to reflect the fundamental physics of radar altimeter returns over ice sheets. Namely, that the leading edge of the waveform corresponds to the first returns from the highest elevations within the footprint, providing clear, and less ambiguous information about these surfaces (Bamber, 1994; Scott et al., 1994). Subsequent parts of the waveform contain progressively more mixed and noisy returns as signals from greater ranges from the satellite become intermixed with multiple distinct surface reflections and volume scattering. This physical reality is well-established and forms the basis for current L2 POCA processing pipelines, which primarily utilise the leading edge for elevation measurements due to its reduced ambiguity (Section 6.1.3). The observed pattern in epistemic uncertainty therefore suggests that the ensemble of models has inherently learned this same underlying physical considerations - showing strong agreement between ensemble members when predicting the 95^{th} percentile (where they can leverage the clear information in the leading edge), and progressively more disagreement when predicting lower quantiles

where they must interpret more ambiguous signals from later in the waveform.

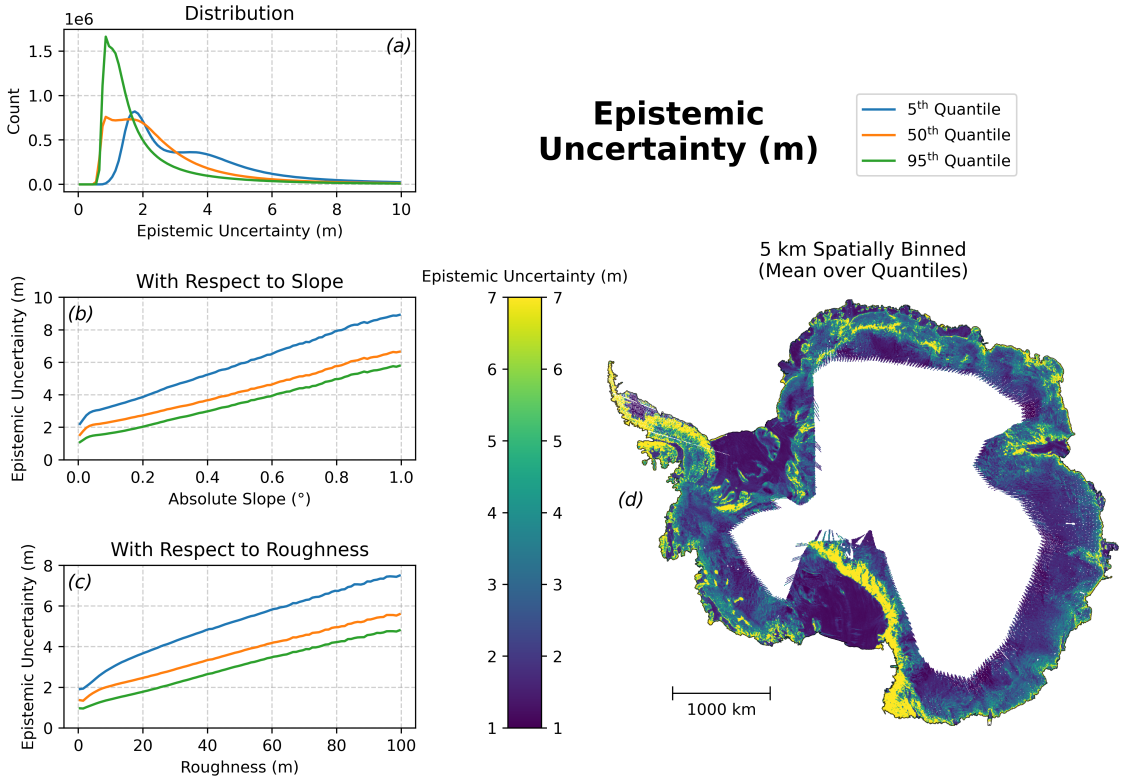

Figure 6.6: Epistemic uncertainty across the Antarctic Ice Sheet (2023), expressed as the standard deviation of ensemble predictions. The 5$^{\text{th}}$, 50$^{\text{th}}$, and 95$^{\text{th}}$ quantiles are shown in blue, orange, and green, respectively. (a) Distribution of uncertainty across quantiles. (b) Uncertainty as a function of ice sheet slope. (c) Uncertainty as a function of ice sheet surface roughness. (d) Gridded mean epistemic uncertainty across the Antarctic Ice Sheet SARIn zone, spatially binned at 5km resolution, where the mean is taken across the three predicted quantiles (5$^{\text{th}}$, 50$^{\text{th}}$, and 95$^{\text{th}}$) within each cell. Spatial patterns are systematic, with elevated uncertainty concentrated in topographically complex regions, consistent with the slope and roughness dependencies shown in panels (b) and (c).

The uncertainty in the quantiles increases linearly with surface slope and roughness (Figure 6.6c-d). With respect to slope, the quantiles increase from 2.2 m, 1.5 m, and 1.1 m at 0°, to 8.9 m, 6.7 m, and 5.8. m at 1° for the 5$^{\text{th}}$, 50$^{\text{th}}$, and 95$^{\text{th}}$ quantiles, respectively (Figure 6.6b). With respect to roughness, the uncertainty increases from 1.9 m, 1.4 m, and 0.9 m at 0 m roughness, to 7.5 m, 5.6 m, and 4.8

m at 100 m roughness (Figure 6.6c). This increase in uncertainty with topographic complexity can be attributed to the fact that well-defined waveforms over flatter and less complex surfaces have fewer possible topographic realisations, whereas complex terrain generates broader distributions as the resulting waveforms can result from many different surface configurations. As the number of possible surfaces grows, so too does the difficulty of the prediction task, which translates into greater epistemic uncertainty overall. In the ensemble framework, this is expressed as increased variance between members, with different network initialisations converging on different plausible realisations of the underlying topography.

Encouragingly, despite slopes greater than 0.75° and roughness values greater than 60 m being out-of-distribution of the training dataset (as discussed in Section 6.3.6), especially towards the upper end of the 0-100 m roughness range shown, there appears to be no clear indication that the uncertainty increases disproportionately in these regions, relative to the in-distribution regions. This suggests that the models have successfully tended towards learning the underlying mapping between waveforms and topography - as opposed to overfitting - and so are able to generalise well out-of-sample, to higher slope and roughness settings. On the contrary, if the models were significantly overfitting, then uncertainties would be expected to rapidly increase in these regions, due to greater variance across the ensemble of model predictions.

Spatially, higher epistemic uncertainty is found in topographically complex regions, consistent with the trends observed in slope and roughness (Figure 6.6d; Figures 5.2 and 5.3, Chapter 5). For instance, high uncertainty is evident across the Transantarctic Mountains, coastal areas, and the Antarctic Peninsula, whereas floating ice shelves and the lower slope interior regions generally exhibit lower epistemic uncertainty.

6.4.3 5-95th Distributional Widths

Following my analysis of epistemic uncertainty, I now turn to aleatoric uncertainty - the ambiguity inherent to non-interferometric retrievals, where a single waveform can arise from many different underlying surface configurations. Specifically, for each record, I analyse the mean width between the predicted 5th and 95th across-track elevation quantiles - hereafter referred to as the distributional width - as a proxy for the overall aleatoric uncertainty in surface topography. In essence, this width captures the range of plausible surface profiles that could have produced a given waveform, offering insight into the inherent ambiguity of non-interferometric SAR altimetry and the spatial and physical factors that shape it. Crucially, this uncertainty is not a measure of model error but a direct output of the model's probabilistic prediction - quantifying how broad the space of plausible surface configurations is, given the limited information in the waveform. As with my assessment of epistemic uncertainty, I consider the overall and spatial distributions of this metric, alongside trends with respect to surface slope and roughness (Figure 6.7).

Over the AIS as a whole, the distribution of the distributional width is positively skewed and unimodal, biased toward lower values, with a median of 28.2 m and a MAD of 8.06 m (Figure 6.7a). Across the ice sheet, distributional widths typically fall in the 20-30 m range. This suggests that even under idealised conditions and with a well-calibrated model, topography across the full satellite swath inferred from a single CS2 SARin waveform remains ambiguous to within tens of metres. Widths below 10 m are rare in the swath-averaged metric, as they are offset by the broader uncertainty at the across-track edges. However, values below this threshold are common in more constrained regions of the footprint, particularly around POCA or near-nadir, where the leading-edge return provides clearer information. Even in these cases, though, the model maintains a baseline level of uncertainty, reflecting the fundamental ambiguity of the inverse problem: multiple distinct surface shapes can yield similar waveforms.

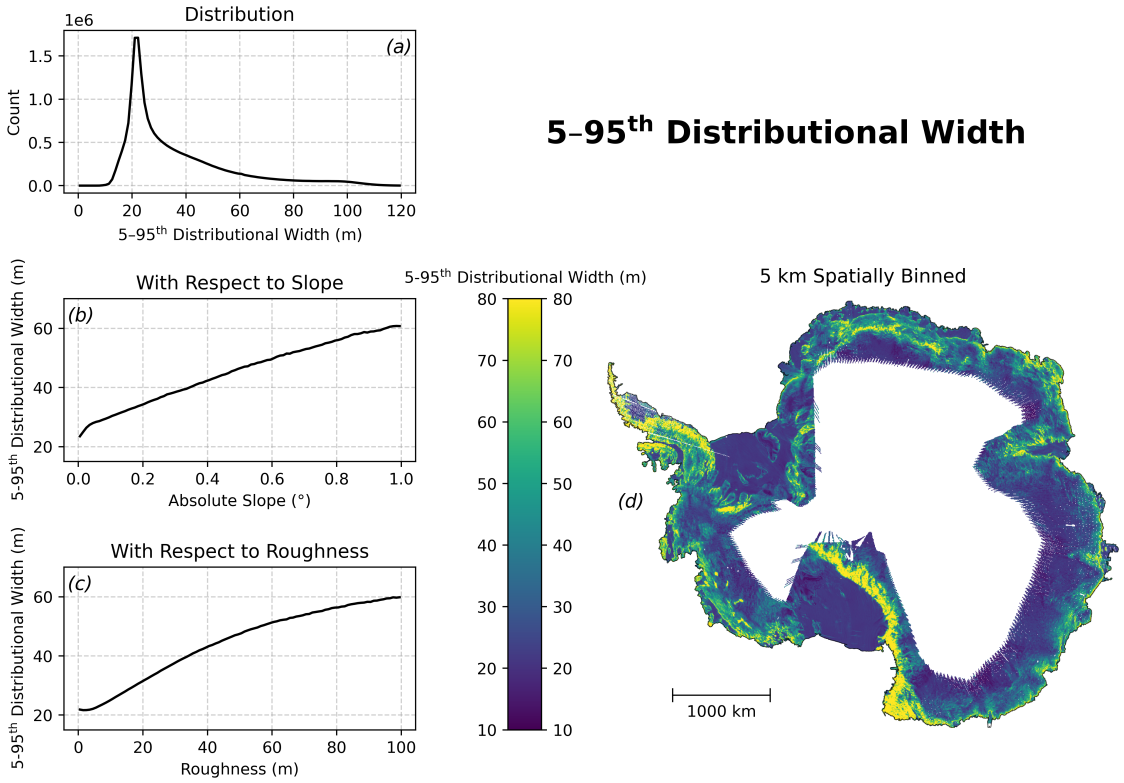

Figure 6.7: Prediction distributional width, defined as the width (difference in elevation) between the 5th and 95th quantiles across the Antarctic Ice Sheet (2023), and used as a measure of the overall aleatoric uncertainty in surface topography. (a) Distribution of 5-95th widths. (b) Widths as a function of surface slope. (c) Widths as a function of surface roughness. (d) Gridded mean 5-95th distributional widths across the Antarctic Ice Sheet SARIn zone, spatially binned at 5 km resolution.

Similar to the analysis of epistemic uncertainty, the distributional width also exhibits a broadly linear trend with respect to slope and (to a slightly lesser extent) roughness. Specifically, the mean distributional width increases from approximately 23.5 m at 0° slope to around 60.8 m at 1° (Figure 6.7b). For roughness, the range grows from 21.9 m at 0 m to 59.8 m at 100 m (Figure 6.7c). As discussed in Sections 6.4.1 and 6.4.2, this quantitively supports the expectations that flat regions exhibit narrower quantile ranges, reflecting greater distributional certainty due to their sharp, well-defined waveforms having fewer possible topographic realisations, whereas complex terrain generates broader distributions as the resulting waveforms

can result from multiple surface configurations.

Spatially, similar trends to that of epistemic uncertainty (Section 6.4.2) are observed, with higher values around more topographically complex areas such as the coastlines and Transantarctic Mountains, and lower values over the smoother and flatter surfaces of the ice shelves and ice sheet interior (Figure 6.7d). Again, this matches the trends seen with respect to slope and roughness.

6.4.4 Prediction Interval Coverage Probability

Next, I evaluate the prediction interval coverage probability (PICP; Equation 6.4), which assesses whether the predicted quantiles achieve their intended statistical coverage with respect to the unseen REMA topographic profiles. Results are presented in Figure 6.8 for four cases: (1) the deviation from the target proportion of data lying within the central interval between the 5^{th}–95^{th} quantiles, (2) the equivalent values below the 50^{th} percentile, (3) the equivalent values above the 95^{th} percentile, and (4) the equivalent values below the 5^{th} percentile. I express these values as deviations from their nominal coverage levels (90%, 50%, 5%, and 5% respectively), which allows for a clearer assessment of calibration across the full distribution. As with my previous assessments of epistemic uncertainty and distributional width, I assess both the spatial trends in PICP and also their relationship with respect to surface slope and roughness.

Notably, interpreting the extreme intervals ($\geq 95^{\text{th}}$ and $\leq 5^{\text{th}}$) requires care. Negative PICP errors in these cases are capped at -5%, which occurs when the predicted quantile lies entirely above ($\geq 95^{\text{th}}$) or below ($\leq 5^{\text{th}}$) the REMA profile. This bound masks the true degree of under-coverage, meaning that a value of -5% for these two quantiles signals unquantifiable misplacement rather than good calibration. Positive errors, by contrast, can rise to +95% and more directly reflect miscalibration. Equivalent bounds apply for the $\leq 50^{\text{th}}$ ($\pm$50%) and 5^{th}-95^{th} (+10%) intervals, so stratification is necessary to interpret these extremes.

On average across the Antarctic Ice Sheet, I find PICP errors of -10.4% for the

5-95% interval, -8.3% for ≤50%, +8.48% for ≥95%, and 1.87% for ≤5%. These values suggest that the model is reasonably well calibrated overall, but with some systematic tendencies. The lower tail (≤5%) is the best calibrated, with a less than 2% PICP error, indicating that the bottom bound of the elevation distribution is well placed on average. In contrast, both the median (50^{th}) and the upper tail (95^{th}) are biased low in elevation, with too few cases below the median and too many above the 95^{th}. This pattern suggests that while the lower bound is accurate, the predicted distribution above it is too tight, and as such does not always capture the full spread of true elevations. This aligns with the previous observation that the predictions tend to better resolve the broad footprint scale topography across the swath, and do not always capture the finer scale topography, such as peaks or localised elevation maxima.

To assess whether this statistical calibration - that is, the agreement between the nominal coverage levels and the observed proportion of elevations lying within those intervals - holds under different surface conditions, I stratified PICP by slope and roughness (Figure 6.8a and 6.8b). Central coverage (5-95%) declines sharply with increasing slope: from values close to the expected ∼90% coverage over flat terrain (within ∼5% error) to more than -40% underestimation on slopes steeper than 0.8°. Decomposition of the intervals shows that this decline is driven primarily by an increase in the fraction of profiles falling below the 5^{th} percentile, which similarly rises upwards of 40% over the steepest terrain. Likewise, the proportion below the median also increases steadily with slope, driven by this systematic tendency to over-placement of the 5^{th} quantile in elevation. In contrast, the proportion above the 95^{th} percentile remains comparatively stable, showing no clear dependencies on slope. A nearly identical pattern is observed for surface roughness, albeit at lower magnitudes, indicating PICP exhibits a stronger dependency on surface slope. Central coverage decreases with increasing roughness, again due to higher fractions of observations lying below both the 5^{th} percentile and the median, while the 95^{th} percentile remains stable. This consistency across both slope and roughness indicates that surface

topographic complexity - whether expressed through large-scale gradients or small-scale variability - reduces calibration primarily through under-predicted estimates in the lower and central parts of the predicted distributions. The robustness of the 95^{th} quantile to slope and roughness demonstrates that the continent-wide mean values are dominated by measurements over flatter, simpler surfaces. These results highlight the importance of stratification, which reveals a degradation in calibration of the lower and central quantiles linked to topographic complexity.

These findings align closely with the uncertainty analyses in Section 6.4.2, where epistemic uncertainty was found to increase disproportionately in the lower and central quantiles with slope and roughness. Here, PICP confirms that these uncertainties manifest as miscalibration: the lower and central bounds of the distribution estimates are systematically biased high in elevation across the swath in complex terrain. Conversely, the relative stability of the 95^{th} percentile across both slope and roughness is consistent with its physical basis: the leading edge of the waveform corresponds to the first returns from the highest elevations within the footprint. This portion of the signal provides clearer, less ambiguous information than later returns, anchoring the upper bound more reliably than the median or lower bounds.

Spatially, for the 5^{th}-95^{th} coverage (Figure 6.8b), I observe trends consistent with those identified in relation to slope and roughness. PICP errors become increasingly negative in topographically complex regions such as the Transantarctic Mountains and the Antarctic Peninsula, meaning that the predicted quantile intervals are too narrow and more than 10% of observations fall outside the 5-95% range - in some places reaching as high as -90%. By contrast, over flatter regions, errors are closer to zero and in some cases slightly positive, indicating that the predicted intervals are wider than required, with more observations falling inside the 5-95% range than expected, such as over parts of the Ross Ice Shelf.

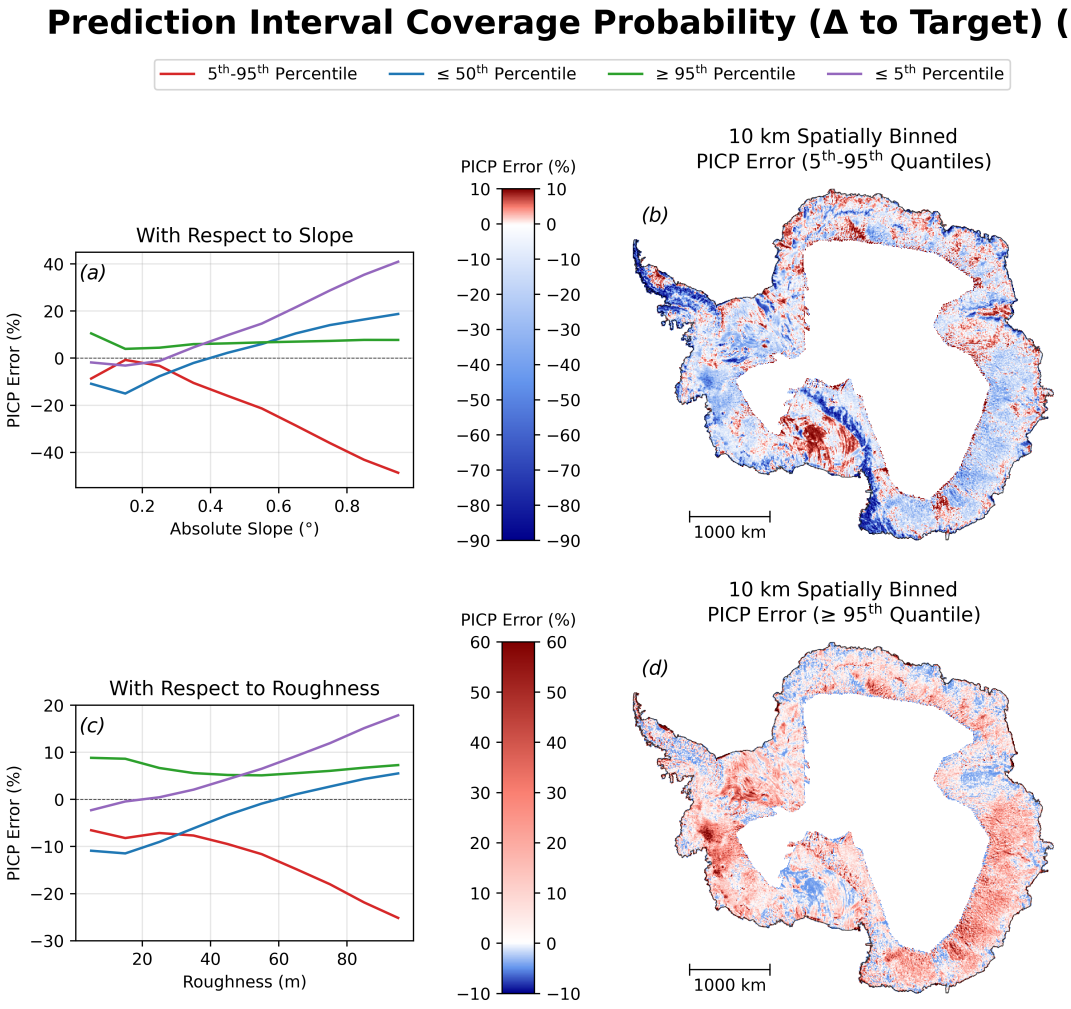

Figure 6.8: Prediction interval coverage probability (PICP) errors across the Antarctic Ice Sheet for predictions made from unseen CryoSat-2 data acquired in 2023 and compared to REMA topographic profiles. PICP errors are expressed as deviations from nominal coverage levels. Values are presented for the (1) central 90% interval (5^{th}-95^{th} percentiles), (2) $\leq 50^{th}$ percentile, (3) $\geq 95^{th}$ percentile, and (4) $\leq 5^{th}$ percentile. (a) PICP error as a function of surface slope over all intervals. (b) Spatial distribution of PICP error for the 5^{th}-95^{th} interval, binned at 10 km resolution. (c) PICP error as a function of surface roughness over all intervals. (d) Spatial distribution of PICP error for the $\geq 95^{th}$ percentile, binned at 10 km resolution.

Because the 95^{th} quantile showed little dependence on slope and roughness, I additionally mapped it in Figure 6.8d to identify whether errors exhibit other spatial structure that was not aligned with slope or roughness. As expected, errors

are generally smaller in magnitude than for the central 90% interval, reflecting the aforementioned improved stability, yet positive biases are evident in several regions, meaning that more observations lie above the predicted 95$^{\text{th}}$ percentile than would be expected under perfect calibration. These are particularly pronounced around areas of dynamic change, including Pine Island and Thwaites Glaciers in West Antarctica (Otosaka et al., 2023; Park et al., 2013; Wingham et al., 2009), and the Larsen C Ice Shelf, where the July 2017 calving of iceberg A-68 occurred (Parmiggiani et al., 2018). These patterns suggest that part of the apparent miscalibration may actually be due to the temporal mismatch, and associated elevation change occurring, between the acquisition of the altimetry and REMA data. This effect is strongest in regions of sustained surface elevation decline, where REMA mosaic elevations are systematically higher than the reality captured by the waveforms, leading prediction distributions to appear biased high. While the filtering in my waveform simulator reduces mismatches, it may not fully remove them - particularly at the less conservative Wasserstein threshold (0.075). Assessing the temporal dimension of the predictions is considered, in detail, in the following chapter.

In summary, PICP highlights both the strengths and limitations of my approach. The 95$^{\text{th}}$ percentile, while showing a small overall bias, proved consistently robust to increasing slope and roughness, reflecting its physical basis in the leading-edge return. By contrast, the central and lower parts of the distribution were systematically miscalibrated, with underestimation becoming progressively more severe in complex terrain. This behaviour points to predictions that have a tendency towards being overly constrained: upper quantiles that are located slightly too low in elevation and, conversely, lower quantiles that are too high. A likely contributing factor is the inherent smoothing effect of probabilistic distributional prediction. As noted in the profile comparisons (Section 6.4.1), the ensemble produces gradual, broadly consistent surfaces rather than reproducing sharp inflections. Localised rises (e.g. peaks) or depressions (e.g. valleys) are therefore dampened across the distribution. This behaviour helps explain why the predicted intervals appear tight

in PICP: by smoothing over extremes, the model yields narrower, more conservative distributions.

At the same time, it is important to place these results in context. Predicting full distributions of possible surfaces without using interferometric phase information is an inherently underdetermined and extremely challenging task. That the model captures the distributional structure at all - and does so with robustness in the 95^{th} percentile even across steep and rough terrain - is a strong indication of its potential for future applications, in terms of efforts to develop capability in probabilistic swath mapping of elevation and elevation change from non-interferometric SAR altimetry systems. Although calibration in the lower and central quantiles remains imperfect currently, this does not preclude its utility for downstream applications, and the clear demonstration of meaningful distributional prediction represents a significant step forward for extracting swath information from non-interferometric measurements, which would otherwise be discarded.

6.4.5 Pinball Loss

As an auxiliary analysis, I also consider the Pinball loss with respect to the REMA profiles over the AIS (2023), averaged across both quantiles and across-track profiles to yield a single value per prediction. While PICP remains the more robust and interpretable diagnostic - directly testing whether predicted quantiles achieve their intended statistical coverage and allowing calibration to be assessed at specific parts of the distribution - Pinball loss provides a complementary perspective. Because it is sensitive not only to whether observations fall within the correct quantiles but also to their distance from them, Pinball loss can penalise broad but well-calibrated distributions more heavily than narrower ones. This makes it less diagnostic of calibration in isolation, but valuable here as it was the ensemble's training objective and therefore offers additional insight into model behaviour.

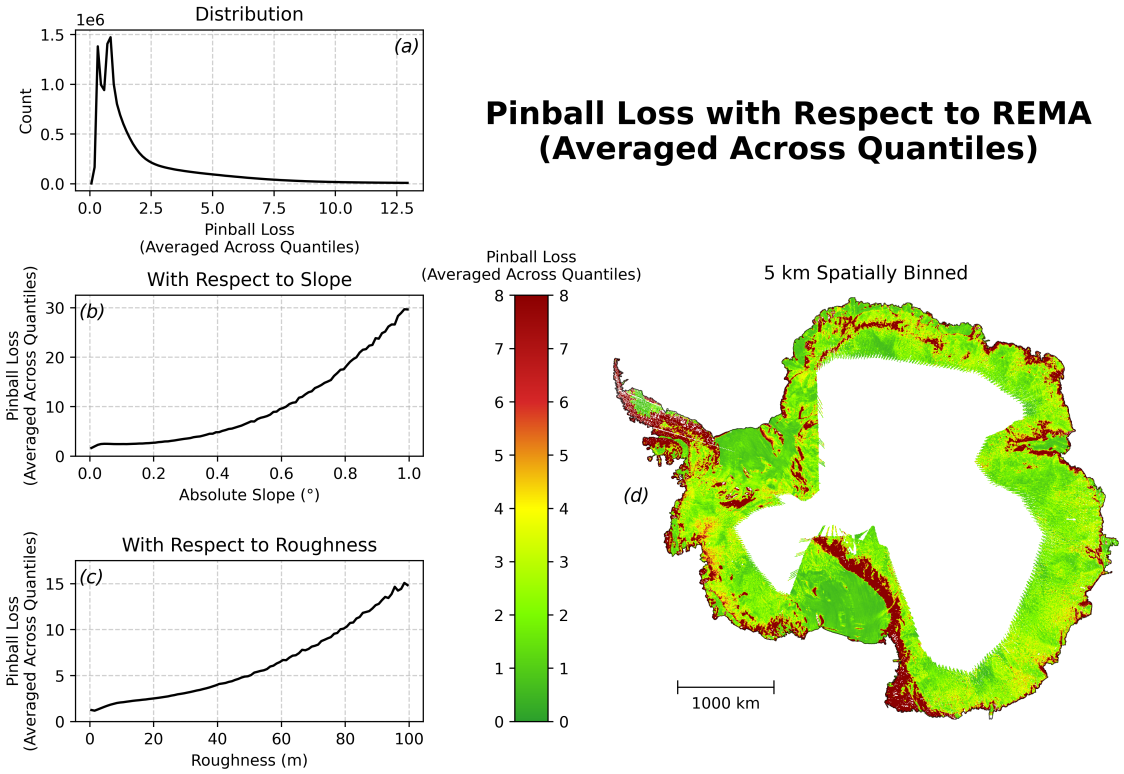

Figure 6.9: Pinball Loss with respect to REMA across the Antarctic Ice Sheet (2023), averaged across both quantiles and across-track profiles. (a) Distribution of Pinball loss. (b) Pinball loss as a function of surface slope. (c) Pinball loss as a function of surface roughness. (d) Gridded mean Pinball loss across the Antarctic Ice Sheet SARIn zone, spatially binned at 5 km resolution.

Pinball loss was computed after filtering waveform-REMA mismatches using the same Wasserstein distance threshold of 0.075 as with Section 6.4.4. The distribution of averaged Pinball losses is unimodal with a long right tail (Figure 6.9a), indicating that most predictions are associated with low errors, though a subset of cases in complex terrain produce much higher values. Median and MAD values are 1.23 and 0.650 respectively, reflecting the overall skew towards low loss. Pinball loss increases systematically with slope (Figure 6.9b), rising from 1.6 at flat terrain to 29.6 at 1°, with the rate of increase itself becoming progressively steeper. Roughness exerts a similar but weaker influence (Figure 6.9c), increasing from 1.3 at 0 m to 14.8 at 100 m. Spatially, these dependencies manifest in the same regions highlighted previously: losses remain low across the ice sheet interior but increase sharply in topographically

complex settings such as the Antarctic Peninsula, Transantarctic Mountains, and many coastal margins (Figure 6.9d). Notably, these trends also reflect the fact that predicted distributions broaden with increasing slope and roughness (Section 6.4.3), meaning that higher Pinball losses in such regions are partly due to the inherent property of Pinball loss penalising distances between points and quantile predictions.

These patterns are consistent with those identified in both the PICP and epistemic uncertainty analyses, with slope exerting the stronger influence relative to roughness, and with performance degrading in complex terrain. Taken together, this consistency shows that the ensemble's optimisation objective during training and independent diagnostics point to the same underlying behaviour, providing additional confidence that the model is capturing meaningful distributional structure.

6.4.6 Difference in the 50$^{\text{th}}$ Quantile

As a final auxiliary diagnostic, I examine the difference between the predicted 50$^{\text{th}}$ quantile (median) and the REMA reference profile, which provides a measure of systematic bias in the central tendency of the predicted distribution. Again, while not as comprehensive as PICP, this metric is straightforward to interpret and provides a simple complementary analysis of where the ensemble median lies relative to REMA.

The analysis was again restricted to waveform-REMA pairs passing the mismatch filtering threshold of 0.075. The distribution of median differences is unimodal, with a median of -2.97 and MAD of 2.96 (Figure 6.10a), confirming that most predictions cluster close to REMA with only moderate spread. As with Pinball loss, deviations increase with both slope and roughness at an accelerating rate (Figures 6.10b, 6.10c). Differences rise from -1.8 m to 85.0 m at 1° slope, and from -1.3 m to 42.4 m at 100 m roughness. Interestingly, both relationships dip slightly below zero at low values (<0.2° slope and <20 m roughness), suggesting a small initial negative bias before positive errors dominate at higher terrain complexity. Importantly, because slope and roughness also drive broader prediction intervals (Section 6.4.3), the median

difference is likely to increase simply as a function of distributional widening: even a well-captured REMA profile may lie far from the ensemble median when the predicted spread is large. Overall, slope exerts the larger influence on prediction differences relative to roughness.

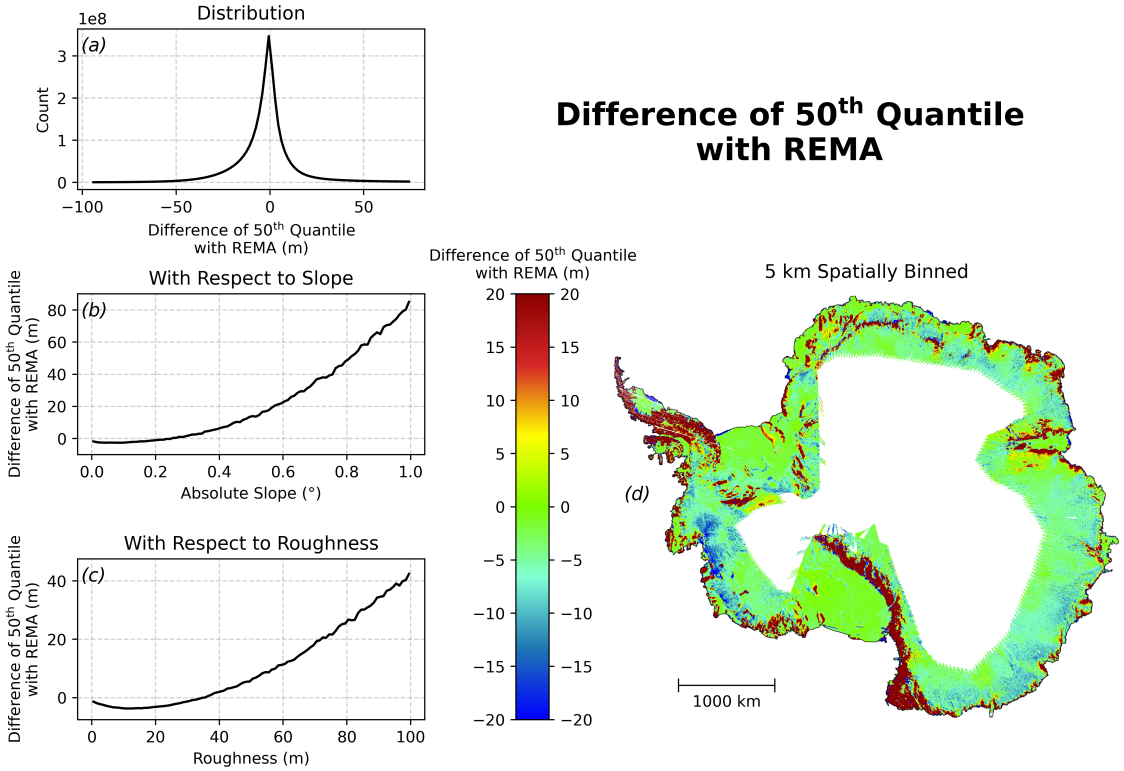

Figure 6.10: Difference of the predicted 50$^{\text{th}}$ quantile (median) with respect to across-track REMA profiles across the Antarctic Ice Sheet (2023). (a) Distribution of median-REMA differences. (b) Median-REMA differences as a function of surface slope. (c) Median-REMA differences as a function of surface roughness. (d) Gridded mean median-REMA differences across the Antarctic Ice Sheet SARIn zone, spatially binned at 5 km resolution.

Spatially, patterns mirror these dependencies: differences are near zero across the flat interior, while strongly positive values occur in high-slope and high-roughness regions such as the Antarctic Peninsula, Transantarctic Mountains, and many coastal sectors (Figure 6.10d). These positive values indicate that REMA elevations are more often located below the predicted median in these areas, consistent with the PICP results. Strong negative differences are also evident, particularly at Pine

Island Glacier, Thwaites Glacier, and Larsen C Ice Shelf - locations that also showed negative PICP errors in the ≥95% quantile PICP spatial analysis (Figure 6.8d). As discussed previously, these are likely driven by temporal mismatches between REMA mosaics and radar altimetry acquisitions, particularly in regions of sustained surface lowering.

Together, the Pinball loss analysis and the median differences relative to REMA show that the ensemble's behaviour is consistent across complementary perspectives, reinforcing confidence that its predictions reflect stable and internally coherent distributional structure.

6.5 Conclusion

Single-antenna radar altimetry presents a fundamental challenge due to the lack of interferometric phase information, making it difficult to determine the exact origin of the radar return within the footprint. Current L2 processing chains, which aim to derive an estimate of the elevation at POCA, address this by assuming a single dominant scattering point at the point of closest approach, attributed to the leading edge of the waveform (Li et al., 2022; Suryawanshi et al., 2025; Huang et al., 2025b). While this approach provides accurate elevation measurements across large parts of the ice sheets, data quality degrades in complex regions such as the Antarctic Peninsula, coastal zones, and the Transantarctic Mountains, and large quantities of potentially useful information are discarded (Brenner et al., 1983; Remy et al., 1989; Bamber, 1994; Brenner et al., 2007; Nilsson et al., 2016) (Chapter 5). In reality, waveforms encode contributions from the entire illuminated surface. More precisely, each waveform can be understood as arising from a range of possible surface configurations consistent with the signal and geophysical constraints. This represents an irreducible (aleatoric) uncertainty in surface geometry for each recorded waveform.

Although this ambiguity cannot be resolved without additional directional

information (e.g. angle of arrival from a second antenna in interferometry), learning to predict the underlying distribution of possible surfaces offers a new, probabilistic, route forward. Such methods have the potential to extract richer measurements from past, current, and future non-interferometric altimetric records, both by producing many elevation estimates where traditional methods return only one, and by recovering useful information from data that current processing chains discard.

This chapter has introduced a novel approach to processing non-interferometric satellite SAR altimetry data, leveraging deep learning to predict the full distribution of possible across-track topographic profiles within a satellite footprint. This approach has been developed and demonstrated using normalised CryoSat-2 SARIn power waveforms and without the use of any interferometric phase information. I train an ensemble of 16 ResNet-RS models with a Pinball loss function to predict multiple quantiles (5^{th}, 50^{th}, and 95^{th}) simultaneously, covering 150 points across the 15 km across-track footprint. Doing so, I demonstrate a robust technique for capturing both the full distribution of possible surfaces for a given waveform (aleatoric uncertainty) and uncertainty in the predictions themselves (epistemic uncertainty). This approach not only provides the means to recover information previously discarded due to the ambiguity of non-interferometric measurements, but also offers a first demonstration of an entirely new data-driven methodology for deriving enhanced topographic understanding from existing and future non-interferometric SAR altimetry missions, which circumvents a large part of existing Level-2 processing schema.

I evaluated the performance of the ensemble on an independent 2023 Antarctic Ice Sheet test set, which included the Pine Island Glacier region that had been withheld during the training process (21,778,482 records, in total). Visual inspection showed robust recovery of topographic profiles, with predicted quantiles capturing the across-track geometry of REMA profiles well. The ensemble generally produced smooth distributions, dampening sharp local extremes and prioritising broadly consistent solutions. Epistemic uncertainty, measured as ensemble standard

deviation, was consistent across quantiles, reflecting their shared origin from the same underlying distribution. Epistemic uncertainty was lowest at the 95th percentile, in line with the established understanding that leading-edge returns provide the clearest, least ambiguous, and most reliable elevation signal (Bamber, 1994; Scott et al., 1994). The predicted 5-95th distribution widths tended to increase at the profile edges, which was attributed to weaker returns due to the antenna gain pattern, indicating greater aleatoric uncertainty. Furthermore, flat regions exhibited narrower, more certain distributions, whereas more topographically complex regions produced broader distributions, reflecting the greater ambiguity in possible surface configurations of the latter. Across the ice sheet as a whole, central coverage (5-95%) averaged about 10.4% lower than expected, with stronger under-coverage occurring in the 50th and 5th percentiles. Combined with slight under-placement of the 95th percentile, this implies a current tendency of the model towards overly constrained distribution predictions, which be attributed in part to the generally smoothed nature of the predictions. Both the 5th and 50th quantiles showed strong dependence on slope and roughness, tending towards increased over-placement in regions of more complex terrain. By contrast, the 95th quantile, while exhibiting a small overall bias, proved strikingly robust to increasing slope and roughness. This resilience reflects its anchoring in the leading-edge return. Auxiliary analyses using Pinball loss and median differences relative to REMA (-2.97 $\pm$ 2.96 m) confirmed the same terrain-dependent patterns, reinforcing the coherence of model behaviour across independent diagnostics. Although calibration remains imperfect - especially in the lower quantiles - the ensemble demonstrates that meaningful distributional prediction is achievable from normalised SAR power waveforms alone, and that useful, physically consistent information can be recovered even in one of the most challenging regimes for radar altimetry.

Despite its successes, and in conjunction with known points of reduced performance, the approach has some inherent limitations. The models were trained to predict REMA-derived topographic profiles, meaning the ensemble acts as a

predictor of REMA rather than a direct regressor of real-world topography. While REMA is highly accurate and based on extensive remote sensing data, the model's performance is inherently constrained by the quality of the mosaic. Additionally, the use of a priori slope data to resolve the ambiguity in profile orientation post-inference, while unavoidable, remains an inherent constraint of non-interferometric approaches. Lastly, the waveform simulator, while efficient and robust, remains a simplified implementation, and more sophisticated simulations may yield improved accuracy in determining waveform-topography relationships, albeit at the expense of increased computational processing load. More broadly, hybrid approaches combining physics-based forward modelling with data-driven inversion - for example, differentiable simulators embedded within the training loop - could enable tighter coupling between learned representations and known electromagnetic scattering physics, and represent a compelling direction for future work beyond the proof-of-concept demonstrated here.

As such, this initial proof-of-concept study paves the way for several future research directions. A clear priority is the development of more extensive datasets (in space and time) using this model, and the associated analysis of the downstream utility of the data, in terms of developing estimates of ice sheet elevation change. This is addressed in the following chapter of this thesis (Chapter 7). Additionally, in terms of methodological development, a number of priority areas for further refinement of the model exist. First, is the development and deployment of improved waveform simulators capable of generating precisely matched surface-waveform pairs, which would provide cleaner training data and allow for prediction across an even broader range of topographic conditions. In this scenario, REMA could serve primarily as a validation resource, while high-fidelity simulations provide the training basis. Parallel avenues for future research include incorporating complementary datasets or alternative geophysical constraints into predictions post-inference to help narrow the range of plausible topographic reconstructions. A related direction is the incorporation of additional prior topographic information to further constrain

predictions, however implementation of high-resolution DEMs in this way were deliberately avoided here due to their temporal rigidity - surface elevations in dynamically active regions can change by metres to tens of metres per year - whereas the slope direction constraint used is stable over much longer timescales. Coarse-resolution or temporally adaptive DEMs could offer a useful middle ground for future implementations. Additional considerations for future work include the potential influence of seasonal variability in surface scattering properties - for example, due to melt-refreeze cycles or fresh snow accumulation - and the role of sub-surface volume scattering in firn, which introduces a 1-D vertical dimension to the scattering problem not captured by the current surface-only framework. Finally, extending the methodology to other satellites, particularly historic LRM missions, offers the prospect of expanding the volume of usable altimetric data and thereby improving long-term reconstructions of Antarctic ice sheet evolution, though adaptation to missions such as ERS-1, ERS-2, and Envisat would require modification to account for differences in measurement characteristics (e.g. footprint geometry).

In conclusion, this paper presents a promising first step in extracting enhanced topographic information from non-interferometric SAR satellite radar altimetry waveforms using deep learning. By quantifying both aleatoric and epistemic uncertainty and providing distributional predictions across the full altimeter swath, rather than single-point estimates, this methodology offers the potential to improve monitoring of the Antarctic Ice Sheet. While calibration of the lower and central quantiles remains imperfect, the robustness of the 95^{th} percentile even over complex terrain demonstrates that meaningful and physically consistent information can be extracted from non-interferometric measurements. By reframing waveform ambiguity as a source of information rather than a limitation, this work lays the groundwork for methodological advances and paves the way toward richer use of non-interferometric altimetry, enabling greater measurement density, improved uncertainty quantification in complex terrain and, ultimately, enhanced reconstructions of past and future ice sheet evolution.

Chapter 7

Computing Ice Sheet Elevation Change from SAR Altimetry and Probabilistic Deep Learning

Joe Phillips[1], Malcolm McMillan[1], Jennifer Maddalena[1]

[1]UK Centre for Polar Observation and Modelling, Centre of Excellence in Environmental Data Science, Lancaster Environment Centre, Lancaster University, Lancaster, UK

Correspondence: Joe Phillips (j.phillips5@lancaster.ac.uk)

The following work is being prepared as a manuscript to be submitted for publication.

JP performed all code development, analysis and manuscript preparation, with supervision from MM and MJ.

Abstract

Accurate monitoring of polar ice sheet elevation change is critical for quantifying sea level rise and improving the process-based understanding that underpins reliable future sea level projections. Synthetic Aperture Radar altimetry provides sufficiently high spatial and temporal coverage to monitor such processes, yet conventional processing reduces the recorded signal to a single-point estimate, discarding potentially useful information. This chapter builds upon the proof-of-concept probabilistic deep learning framework introduced in Chapter 6 to assess the practical application of this approach for estimating ice sheet elevation change. This work demonstrates for the first time how credible ice sheet swath elevation change estimates can be derived from SAR altimetry via probabilistic deep learning.

More specifically, the framework generates ice sheet elevation predictions for two datasets withheld during model training: (1) Pine Island Glacier in West Antarctica, and (2) the entire Greenland Ice Sheet periphery. These predictions are compared against a reference dataset from ICESat-2 laser altimetry. Prediction Interval Coverage Probability (PICP) analysis shows that results derived from the 95^{th} quantile remained stable, though the lower quantiles exhibited increasingly systematic under-coverage in complex terrain. These patterns mirror those observed in Antarctica (in-distribution), indicating that the model successfully captured transferable physical relationships rather than region-specific characteristics.

These datasets are subsequently used to derive ice sheet elevation change rates between 2019-2024, resolving established glaciological signals that dominate current ice sheet mass loss, including 2-3 m yr^{-1} thinning at Pine Island Glacier and major outlet glacier losses in Greenland. These results are compared to elevation change estimates from three independent datasets - ICESat-2, CryoSat-2 point-of-closest-approach (POCA), and CryoSat-2 interferometric swath altimetry (Cryo-TEMPO EOLIS) - finding relatively good agreement in all cases. Notably, both elevation change rate (dh/dt) maps and spatially-integrated relative elevation time series (dh) closely match interferometric performance, despite the deep

learning framework using only power waveforms without interferometric phase information. Additionally, the ensemble predictions detected sub-meter annual changes, demonstrating their ability to resolve subtle waveform variations into realistic elevation signals and confirming that waveform ambiguity represents a valuable information source rather than a processing obstacle. These results provide the first practical demonstration of this data-driven processing approach for monitoring ice sheet evolution, with potential applications for historical missions (ERS-1/2, Envisat), current operational systems (Sentinel-3), and future SAR altimeters (Sentinel-3 Next Generation and CRISTAL).

7.1 Introduction

The polar ice sheets are undergoing rapid change due to contemporary global warming. Between 1992 and 2020, Antarctica lost 2,671 ± 530 billion tonnes of ice, contributing 7.4 ± 1.5 mm to global sea level rise (Otosaka et al., 2023). Over a similar period, 1992-2018, Greenland lost 3,902 ± 342 billion tonnes, contributing 10.8 ± 0.9 mm to sea level rise (IMBIE Team, 2020). Current rates of mass loss from both ice sheets are now six times greater than those observed in the 1990s (IMBIE Team, 2020), with combined contributions accounting for approximately one-third of contemporary sea level rise (Shepherd et al., 2012; IMBIE Team, 2018; IMBIE Team, 2020; Otosaka et al., 2023). Projections suggest that polar ice sheet contributions could reach 70 mm by 2100 (Fox-Kemper et al., 2021).

Satellite radar altimetry has emerged as the primary tool for monitoring ice sheet elevation change across these vast and remote regions, providing the spatial coverage and temporal consistency necessary to track topographic evolution over multi-decadal timescales (Zwally, 1989; Zwally et al., 2005; Zwally et al., 2011; Hurkmans et al., 2014; McMillan et al., 2016; Shepherd et al., 2019; Sandberg Sørensen et al., 2018; Smith et al., 2023). The ability to detect and attribute elevation

changes at the scale of individual drainage basins and glacier systems is critical for understanding the physical processes driving mass loss, constraining ice sheet models, and reducing uncertainty in future sea level projections (Shepherd et al., 2019). However, significant challenges remain in extracting reliable geophysical information from radar altimetry, particularly using non-interferometric instruments over regions of complex topography where conventional processing approaches degrade substantially (Brenner et al., 1983; Remy et al., 1989; Bamber, 1994; Brenner et al., 2007; Nilsson et al., 2016).

In this regard, conventional, non-interferometric radar altimetry faces a fundamental ambiguity due to its capacity to record backscattered power as a function of range only; namely, each waveform corresponds to many different possible surface configurations within the illuminated footprint. Standard Level-2 processing schema resolve this ambiguity by reducing the return to a single elevation estimate, typically tied to the leading edge of the waveform and attributed to the point of closest approach (POCA) to the satellite (Aublanc et al., 2024; European Space Agency, 2021). While this approach has underpinned decades of operational altimetry, it discards much of the information encoded in the waveform and is particularly susceptible to data loss and degradation in complex terrain, where these simplifying assumptions break down (Bamber, 1994; Simonsen et al., 2017; McMillan et al., 2018).

While current interferometric missions like CryoSat-2 provide enhanced capabilities over complex terrain, the majority of historical altimetric data consists of non-interferometric observations from missions such as ERS-1, ERS-2, and Envisat. Furthermore, the ongoing Sentinel-3 constellation operates in SAR mode only, a configuration that will continue with the planned Sentinel-3C and 3-D satellites (Egido, 2023). Likewise, although the upcoming CRISTAL mission will employ SAR interferometry at Ku-band, it will employ SAR-only at Ka-band (Kern et al., 2020). This highlights the need for improved methods to recover more reliable and more extensive elevation change measurements from non-interferometric data, ensuring

that the multi-decadal altimetry record can be fully exploited for monitoring ice-sheet evolution.

In Chapter 6, I demonstrated that deep learning can successfully extract distributional elevation information from non-interferometric radar waveforms, fundamentally reframing how altimetric measurements can be processed and interpreted. By training an ensemble of neural networks to predict the 5^{th}, 50^{th}, and 95^{th} quantiles of possible topographic profiles within each footprint, I demonstrated how this approach is able to capture both the inherent ambiguity in waveform interpretation (aleatoric uncertainty) and model confidence (epistemic uncertainty). This represents a fundamental shift from conventional altimetry Level-2 processing over ice sheets, offering a pathway to recover elevation information that traditional approaches discard.

In this chapter I build upon this foundational work to explore the practical application of the deep learning framework through two complementary analyses designed to test its real-world utility and generalisability. First, I extend the methodology to Pine Island Glacier, West Antarctica; a region that was withheld from model training and represents one of the most rapidly changing sectors of the Antarctic Ice Sheet (Otosaka et al., 2023; Park et al., 2013; Wingham et al., 2009). Second, I apply the trained models to the Greenland Ice Sheet, testing their ability to generalise across fundamentally different ice sheet environments. In both cases, statistical calibration is assessed via Prediction Interval Coverage Probability (PICP) against contemporaneous ICESat-2 measurements, using model predictions of swath elevation generated over a 6-year period (2019-2024), spanning the ICESat-2 operational lifetime (launched September 2018). I then use these measurements to derive estimates of ice sheet elevation change during that time. I compare these new estimates to equivalent solutions derived from three independent altimetric datasets: CryoSat-2 conventional POCA processing, CryoSat-2 interferometric swath altimetry (EOLIS), and ICESat-2 laser altimetry - the latter serving as the primary reference dataset due to its centimeter-level accuracy. This multi-dataset

comparison allows me to assess the relative performance of these different processing approaches and validate the deep learning framework against established operational products. Finally, by applying the SVD-based topographic characterisation method introduced in Chapter 5 to derive new high-resolution estimates of Greenland Ice Sheet slope and roughness, I investigate how the performance of these different altimetry datasets is impacted by varying levels of topographic complexity.

7.2 Study Sites and Methods

7.2.1 Study Sites

This chapter applies the deep learning framework developed in Chapter 6 to generate and evaluate new estimates of ice sheet elevation change across two contrasting polar environments: Pine Island Glacier (PIG) in West Antarctica and the Greenland Ice Sheet.

7.2.1.1 Pine Island Glacier

PIG is one of the fastest-changing glaciers in Antarctica, thinning at rates exceeding 1 m yr^{-1} close to its terminus, and responsible for $\sim$13% of West Antarctic Ice Sheet mass loss in recent decades (Reed et al., 2024; Rignot et al., 2019). PIG has been the subject of extensive research campaigns in recent years (e.g., iSTAR, BAS airborne surveys, CryoSat-2 analyses), meaning that signals of elevation change across this glacier are well characterised, as a result of its dynamic thinning and grounding-line retreat (Otosaka et al., 2023; Park et al., 2013; Wingham et al., 2009).

For this study, PIG serves as a withheld test site; i.e. no data from this region were included in model training (2012-2021), making it a fully unseen case. This allows me to evaluate how well the model generalises to an unseen and highly dynamic sector of the West Antarctic Ice Sheet. Given PIG's central role in regional mass loss, robust dh/dt analysis here is critical for assessing model skill in the areas most sensitive to changes in climate. Specifically, analysis is conducted over

2019-2024, providing some additional temporal extension beyond the training period (2012-2021).

7.2.1.2 Greenland Ice Sheet

The second test site is the Greenland Ice Sheet. Greenland provides a natural counterpart to the Antarctic test, sharing similarities in scale and climate sensitivity but differing in geometry, accumulation, and melt processes (Carr et al., 2013; Straneo et al., 2013; Enderlin et al., 2014). Current behaviour across the ice sheet also provides high variability, from close to no change ($\sim$0 m yr^{-1}) across the interior, to high thinning rates (often exceeding 1-2 m yr^{-1}) across its marine-terminating outlet glaciers (Rignot et al., 2011; Otosaka et al., 2023; Straneo et al., 2010; Motyka et al., 2011; Straneo et al., 2013; Joughin et al., 2004; Motyka et al., 2011; Joughin et al., 2010). Greenland was entirely absent from training and represents the application of the model to an entirely different ice sheet setting, thus making it an ideal domain for assessing generalisability: can the model produce realistic surface change estimates in a system governed by different glaciological processes on the other side of the globe?

Additionally, producing high-quality dh/dt maps for Greenland remains a major priority for constraining global sea-level budgets, especially given the need to reconcile radar altimetry with laser altimetry, input-output, and gravimetry-based estimates (Otosaka et al., 2023). This inclusion therefore functions as a limit test, probing whether the network has learned a transferable mapping between waveform shape and surface geometry, rather than overfitting to Antarctic-specific characteristics.

7.2.2 Data Acquisition and Processing

In this section, I outline the datasets used in my analysis and the steps taken to process them. Each dataset has already been introduced and described in Chapter 2; and so here I provide the context necessary for understanding their role in this

study.

CryoSat-2 (CS2) is an ESA Earth Explorer mission launched in 2010 to monitor variations in ice sheet and sea ice thickness (Parrinello et al., 2018). The mission's primary instrument, the SIRAL radar altimeter, operates in three distinct acquisition modes: Low Resolution Mode (LRM), Synthetic Aperture Radar (SAR) mode, and Synthetic Aperture Radar Interferometric (SARIn) mode. This analysis focuses exclusively on data acquired within the SARIn mask, which covers grounded ice and adjacent land at the ice sheet margins. Within this region, the Level-1b (L1b) products serve as input data for elevation predictions generated through the deep learning framework. For comparative analysis, conventional Level-2+ point-of-closest-approach (POCA) and swath elevations solutions are also employed. Both of these latter products additionally utilise interferometric phase information. Here, the POCA product uses interferometric phase information to directly locate the POCA, unlike standard SAR-based pipelines where slope correction methods compensate for the lack of interferometric information.

In addition to CryoSat-2 data products, this study incorporates measurements from ICESat-2, a NASA laser altimetry mission launched in 2018 that provides high-precision elevation measurements using photon-counting lidar technology (Markus et al., 2017). The Advanced Topographic Laser Altimeter System (ATLAS) measures surface elevations with centimeter-level precision across six ground tracks, offering detailed profiles of ice sheet surfaces. Although ICESat-2 coverage is spatially more limited than radar altimeters due to narrow track widths and clear-sky requirements, it has become the standard reference for validating satellite radar altimetry measurements and is widely recognised as the most accurate source of satellite-derived polar elevation data currently available (Aublanc et al., 2025b; Helm et al., 2014; Crétaux et al., 2018; Aublanc et al., 2025a; Dawson et al., 2023). Consequently, ICESat-2 data serve as the ground-truth baseline in this study, providing independent validation for both the conventional CryoSat-2 datasets and the deep learning framework.

Specifically, the deep learning framework uses CryoSat-2 Level-1B baseline D waveform data as input to generate elevation predictions (available at `ftp://science-pds.cryosat.esa.int`). As outlined previously, these predictions represent across-track elevation distributions and include both aleatoric uncertainty, linked to inherent ambiguity in the origin of reflections, and epistemic uncertainty, reflecting model variability. To place these model outputs in context, they are compared against three independent datasets that represent different measurement strategies: (1) the Cryo-TEMPO Level-2+ POCA product (baseline D; available at `http://science-pds.cryosat.esa.int/`), which represents the standard interferometric solution and forms the baseline for many CryoSat-2 applications; (2) the Cryo-TEMPO EOLIS point product (version 2; available at `https://cs2eo.org/cryotempo`), which extends this by incorporating interferometric phase information at ranges beyond the POCA, thus providing swath elevations across the satellite track; and (3) the ICESat-2 L3A Land Ice Height product (version 6; available at `https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2670138092-NSIDC_CPRD`), which offers a contrasting approach through laser altimetry, delivering centimeter-level accuracy with dense along-track sampling.

All datasets are filtered to the CryoSat-2 SARIn operational mode mask and restricted to the six-year period 2019-2024, corresponding to the operational lifetime of ICESat-2, which ensures consistent spatial and temporal coverage across all products. The ensemble generates 150 across-track elevation measurements for each CryoSat-2 waveform, with each point comprising estimates of three quantiles (5^{th}, 50^{th}, and 95^{th} percentiles) of the predicted elevation distribution. This approach yields 394,405,653 model-derived points over Pine Island Glacier and 2,003,717,083 points across Greenland. For comparison, the EOLIS interferometric product provides 280,850,310 measurements over Pine Island Glacier and 2,233,867,447 over Greenland, while the conventional Level-2+ POCA approach yields substantially fewer points (4,341,314 for Pine Island Glacier, 34,841,933 for Greenland) due to its single-point-per-waveform methodology. ICESat-2 contributes 233,468,190 high-

precision laser altimetry measurements over Pine Island Glacier and 2,549,046,256 across Greenland.

7.2.2.1 Slope and Roughness from REMA and ArcticDEM

In conjunction with the altimetric datasets used in this study, slope and roughness metrics are also derived from DEM mosaics at two complementary scales: locally, to guide the orientation of predicted across-track profiles, and regionally, to map topographic variability across Greenland. Specifically, to guide across-track orientation, the 1 km resolution mosaics of both REMA (available at `data.pgc.umn.edu/elev/dem/setsm/REMA`) and ArcticDEM (available at `data.pgc.umn.edu/elev/dem/setsm/ArcticDEM`) are used. To map topographic variability across the Greenland Ice Sheet, I use the 100 m ArcticDEM mosaic product, mirroring the slope and roughness maps generated using REMA in Chapter 5, and ensuring consistent analysis across both polar ice sheets. To limit the DEMs to the areas of interest, REMA is clipped to Pine Island Glacier using established grounded ice area and drainage basin definitions (basin 22; (Zwally et al., 2012)), while ArcticDEM is clipped to Greenland using the coastline shapefile developed by (Gerrish, 2020) (version 1; available at `https://ramadda.data.bas.ac.uk/repository/entry/show?entryid=8cecde06-8474-4b58-a9cb-b820fa4c9429`)

At the footprint scale, I employ the coarse 1 km mosaics to resolve the directionality of slope in the across-track direction. This is due to the fact that waveform ambiguity makes the orientation of individual across-track profiles fundamentally indeterminate. To avoid doubling the number of possible solutions, I therefore enforce a consistent orientation at prediction and use local slope to flip predicted profiles into the correct direction post-inference.

In practice, for each CS2 L1b record, DEM values are extracted within a 380 m $\times$ 15 km rectangle oriented according to track heading and interpolated to a 15-point across-track line. Surface slope and roughness are then calculated using the Singular Value Decomposition (SVD) method introduced in Chapter 5 and applied in Chapter

6 to generate training data. For each 15-point profile, coordinates are mean-centred, and SVD of the coordinate matrix yields the unit normal to the best-fit line. Surface slope is derived from the angle of this line relative to the horizontal, while roughness is defined as the peak-to-peak (min-max) range of orthogonal residuals about the fit. Full methodological detail, including the two-dimensional extension used for ice-sheet-wide mapping, is provided in Chapter 5.

As in Chapter 6, the present analysis is restricted to one dimension, since the SAR footprint is overwhelmingly sensitive to across-track topography. Although along-track roughness within the ∼380 m SAR footprint may introduce some waveform variability, its influence is considered negligible relative to across-track surface geometry (∼15 km), which dominates waveform structure over ice sheets. Treating the problem as one-dimensional therefore provides a substantial simplification while retaining the primary topographic control. Although this procedure itself is unchanged from earlier chapters, a coarser DEM input is used here. Whereas Chapter 6 employed 100 m resolution mosaics, the present study uses 1 km resolution products, judged sufficient given the 15 km across-track width and the relative insensitivity of the min-max residual roughness metric to sampling density.

Secondly, for the ice-sheet-wide analyses, I apply the full 3-D SVD method from Chapter 5 using the same 900×900 m sliding window, but with the 100 m ArcticDEM mosaic. DEM coordinates are mean-centered and decomposed to obtain the unit normal vector to the fitted plane, from which slope and roughness are derived in two dimensions. The result is continuous slope and roughness fields across the Greenland ice sheet, providing a direct analogue to the Antarctic products developed in Chapter 5. To my knowledge, this represents the first pan-Greenland surface roughness dataset that has been derived independently of slope.

7.2.2.2 Prediction Interval Coverage Probability

To evaluate my ensemble predictions in an application setting, I compute the Prediction Interval Coverage Probability (PICP) with respect to ICESat-2 (Equation

6.4). PICP measures whether the proportion of the ICESat-2 reference elevations falling within a given prediction interval matches its nominal confidence level (Sluijterman et al., 2024), allowing me to test the statistical validity of the full predictive distributions in cases where direct pointwise comparisons are not possible. For example, coverage of the 5^{th}-95^{th} quantiles (i.e. the proportion of ICESat-2 measurements falling within) should ideally be $\sim$90%. Specifically, I assess four cases: the central 90% interval (5^{th}-95^{th} quantiles), the proportion $\leq 50^{\text{th}}$ quantile, the proportion $\geq 95^{\text{th}}$ quantile, and the proportion $\leq 5^{\text{th}}$ quantile. This mirrors the calibration performed with respect to the REMA mosaic profiles in Chapter 6, but instead with respect to a reference dataset with high absolute accuracy. Here, the purpose shifts from assessing the model's capability to reproduce a REMA topographic surface to evaluating the real-world precision of elevation predictions for monitoring ice sheet change.

7.2.2.3 Ice Sheet Surface Elevation Change

To compute elevation change (dh) and rates of change (dh/dt), I adopted a consistent processing approach across each of the four datasets: (1) the deep learning framework, (2) Cryo-TEMPO POCA, (3) Cryo-TEMPO EOLIS, and (4) ICESat-2. This approach largely follows prior works and represents a well-established methodological standard widely used in elevation-change analyses for over a decade (Flament et al., 2012; Slater et al., 2018; McMillan et al., 2014; McMillan et al., 2016; McMillan et al., 2019).

In all cases, measurements with reported uncertainty greater than 3 m were excluded prior to processing. For the deep learning framework, the 5^{th}, 50^{th}, and 95^{th} quantile quantiles were processed independently, yielding three parallel estimates. This provides not only an inherent uncertainty envelope for the elevation change time series - defined by the maximum, median, and minimum across quantiles at each epoch - but also a means to directly explore the relative performance of elevation change estimates derived from each of the different quantiles. As shown in Chapter

6 the quantiles exhibit differing performance, with the 95$^{\text{th}}$ quantile displaying the greatest consistency across ensemble members, followed by the 50$^{\text{th}}$ and 5$^{\text{th}}$. This pattern reflects the underlying physics of radar altimetry: the leading edge of the waveform (captured by the 95$^{\text{th}}$ quantile) corresponds to the first returns from the highest elevations within the footprint and thus provides less ambiguous information, whereas later parts of the waveform (represented by the 50$^{\text{th}}$ and 5$^{\text{th}}$ quantiles) integrate progressively noisier and more mixed returns (Bamber, 1994; Scott et al., 1994).

To compute surface elevation change from each dataset, I first binned all measurements into spatially defined 5 km grid cells spanning the area covered by Pine Island Glacier and the Greenland Ice Sheet. A 5 km grid spacing was selected to balance spatial resolution against the requirement for a sufficient number of measurements per cell to support robust surface fitting, consistent with prior ice sheet elevation change studies using similar datasets (McMillan et al., 2016; McMillan et al., 2019). All measurements falling within each grid cell were collected, and for the measurements within each cell a polynomial surface model was fit using ordinary least squares. In the case where heading information was available (Cryo-TEMPO POCA only), it was included as an independent variable (Equation 7.1). In all other cases, the heading term was omitted (Equation 7.2). Grid cells were discarded if the measurements spanned less than 50% of the total 6-year temporal observation window (2019-2024), ensuring that only cells with sufficiently long observational records were retained for robust trend fitting. Outliers were removed iteratively by excluding measurements whose residuals exceeded 2σ from the fitted surface, repeating until convergence or until less than 20 points remained. The surface and heading components were then subtracted from the fitted elevations, leaving only the temporal signal (dh) plus residual noise. From these residuals, a linear model was then fitted (Equation 7.3). While more sophisticated spatiotemporal approaches, such as Gaussian process regression (e.g. GPSat) or dynamic-resolution frameworks, may offer advantages for interpolating sparse

observations, the polynomial surface fitting method adopted here is computationally efficient at ice-sheet scale and has been extensively validated in the literature (Flament et al., 2012; McMillan et al., 2016; McMillan et al., 2019). Its established performance also provides a robust and well-understood baseline framework within which to evaluate the new deep learning framework.

$$z(x, y, h, t) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + a_6h + a_7t, \tag{7.1}$$

$$z(x, y, t) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + a_6t. \tag{7.2}$$

Here, z is elevation, x, y are Antarctic (or Arctic) polarstereographic coordinates relative to the grid cell centre, h is satellite heading, and t is time in decimal years relative to the midpoint of the observation period.

$$dh(t) = mt + c \tag{7.3}$$

Finally, the elevation residuals were temporally binned into epochs to form a time series of elevation change. For the deep learning framework, the EOLIS swath dataset, and the ICESat-2 product, I adopted 30-day epochs, whilst for the POCA solution a 60-day epoch was used to account for the lower data density. This provided elevation change with respect to the midpoint of the observation period. The elevation change at the start of the period was then subtracted to attain elevation change with respect to the start of 2019. I then average across the spatial grids by taking the median, attaining an elevation change timeseries for each of the Pine Island Glacier and Greenland study regions, spanning 6 years (2019-2025). Over Pine Island Glacier, 214,253,684 model-derived points (54.3% of the original dataset) successfully passed through the uncertainty filtering and surface fitting process, compared to 211,425,295 EOLIS points (75.3% retention), 123,784,747 ICESat-2 measurements (53.0% retention), and 2,986,843 Cryo-TEMPO POCA points (68.8% retention). Over Greenland, the pattern is similar: 1,121,529,522 model points (56.0% retention), 1,601,602,394 EOLIS points (71.7% retention), 1,370,831,600

ICESat-2 measurements (53.8% retention), and 23,905,590 Cryo-TEMPO POCA points (68.6% retention).

Together, these steps provide a consistent framework for directly comparing the ensemble predictions with conventional altimetry datasets and for assessing their ability to capture established spatial and temporal patterns of ice sheet dynamics and evolution. Specifically, this approach allows for intercomparison of elevation change derivations from the three CryoSat-2 processing methods with respect to ICESat-2 as an independent reference, while also using ICESat-2 for validating the statistical calibration of the probabilistic predictions through PICP analysis.

7.3 Results

7.3.1 Slope and Roughness over Greenland

First, I report the results from applying the 3-D SVD slope and roughness algorithm to the Greenland Ice Sheet, using the 100 m resolution ArcticDEM mosaic (Figure 7.1). To my knowledge, these are the first maps of Greenland surface slope and roughness produced using this technique, and the first Greenland-wide roughness map derived from the high-resolution ArcticDEM product, where roughness is defined as the macro-scale peak-to-peak range of orthogonal residuals from a best-fit plane within a 900 m sliding window. The overall topographic characteristics of the ice sheet are somewhat similar to Antarctica (Chapter 5; Figure 5.1), with low slopes and roughness in the interior increasing toward higher values at the margin. However, the Greenland Ice Sheet exhibits steeper surface slopes overall, with a median slope of 0.434°, which is more than twice that of Antarctica (0.192°).

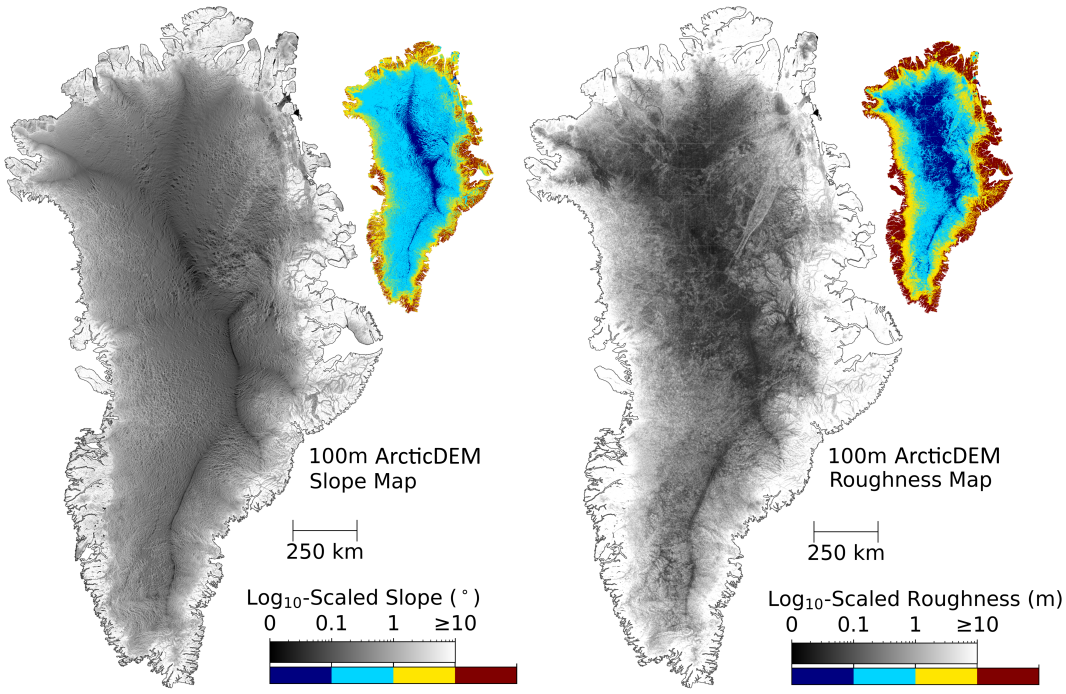

Figure 7.1: 5 km resolution, $\log_{10}$-scaled slope and roughness maps of the Greenland Ice Sheet generated using 100 m resolution ArcticDEM and the Singular Value Decomposition techniques outlined in Section 7.2.2.1. The coloured inset maps show discretely binned versions of these datasets, to aid the visualisation of broader scale patterns.

Similarly, the surface roughness of Greenland is also much greater than its Antarctic counterpart, with a median roughness of the former of 0.855 m, nearly three times higher than Antarctica's 0.297 m (Chapter 5). The spatial variability in these values is also considerably greater, as evidenced by the median absolute deviation of 0.793 m for roughness and 0.323° for slope, compared to Antarctica's values of 0.219 m and 0.134°, respectively.

The relationship between slope and roughness in Greenland (Figure 7.2) shows a strong positive correlation, indicating that steeper terrain is associated with increased surface roughness. This can be observed from the Pearson correlation coefficient between slope and roughness, which at 0.794, is comparable to, but slightly lower than, the value observed for Antarctica (0.808) (Chapter 5). However, the absolute values and ranges of both slope and roughness are substantially

higher in Greenland, reflecting its more dynamic and topographically variable nature compared to the relatively stable interior that dominates Antarctica. This close correlation between the two ice sheets supports the robustness of the slope and roughness algorithm, as the consistency across geographically distinct regions suggests the correlation is intrinsic to topographic characteristics rather than an artifact of the computational method.

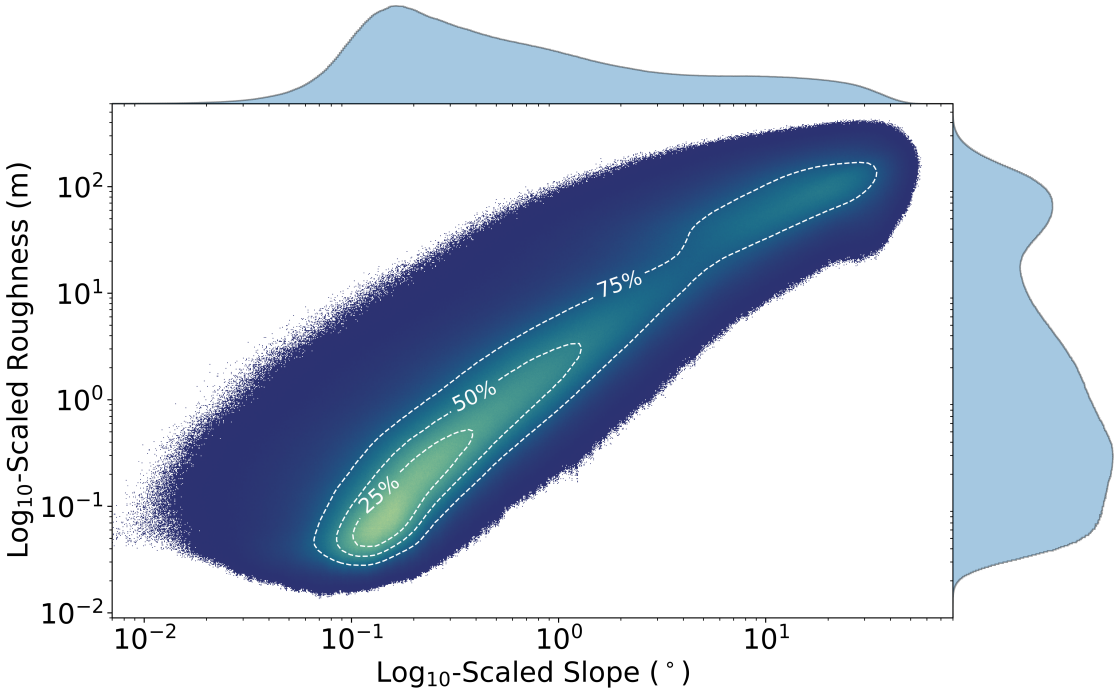

Figure 7.2: Log_{10}-binned joint and marginal distributions of slope and roughness across the Greenland Ice Sheet. The joint distribution is displayed as a density plot, with lighter colours indicating a higher density of measurements, and the contours bounding 25%, 50%, and 75% of the data.

7.3.2 Analysis over Pine Island Glacier

Next, I move on to the analysis of the ensemble in estimating ice sheet elevation and elevation change. I start by assessing performance across the Pine Island Glacier. This region provides a stringent test of the framework, both because it was deliberately withheld from training and because it represents one of the most

rapidly evolving sectors of the West Antarctic Ice Sheet (Reed et al., 2024; Rignot et al., 2019; Otosaka et al., 2023; Park et al., 2013; Wingham et al., 2009). By focusing on PIG, I assess whether probabilistic predictions from the deep learning framework can recover realistic glaciological signals in an unseen region where surface conditions, flow dynamics, and elevation change rates depart markedly from the average found in the Antarctic-wide training domain. Validation against ICESat-2 is particularly important in this context. Whereas the REMA profiles used in Chapter 6 were anchored to a 2015 mean epoch (Howat et al., 2019) and thus subject to temporal mismatch, the 2019-2024 ICESat-2 dataset provides contemporaneous reference elevations aligned with the model outputs. This eliminates artifacts caused by mosaic averaging, ensuring that deviations reflect genuine calibration tendencies rather than temporal offsets, yielding a more complete view of real-world model performance. In the following sections I first present an analysis of PICP elevation statistics, before then proceeding to assess model performance with respect to resolving rates of surface elevation change.

7.3.2.1 Prediction Interval Coverage Probability

First, I assess prediction interval coverage probability (PICP) over Pine Island Glacier using ICESat-2 altimetry from 2019-2024 as a reference. As in Chapter 6, I analyse the deviation of the PICP from the nominal proportion of data expected within each interval, providing a direct measure of statistical calibration across the predicted distribution. More specifically, four cases are evaluated: (1) the central 90% interval (5^{th}-95^{th} quantiles), (2) the $\leq 50^{\text{th}}$ quantile, (3) the $\geq 95^{\text{th}}$ quantile, and (4) the $\leq 5^{\text{th}}$ quantile.

As outlined in Chapter 6, interpreting the extreme intervals ($\geq 95^{\text{th}}$ and $\leq 5^{\text{th}}$) requires care. Negative PICP errors in these cases are capped at -5%, which occurs when the predicted quantile lies entirely above ($\geq 95^{\text{th}}$) or below ($\leq 5^{\text{th}}$) all ICESat-2 points. This bound masks the true degree of under-coverage, meaning that a value of -5% for these two quantiles signals unquantifiable misplacement rather than good

calibration. Positive errors, by contrast, can rise to +95% and more directly reflect miscalibration. Equivalent bounds apply for the $\leq 50^{\text{th}}$ ($\pm 50\%$) and 5^{th}-95^{th} ($+10\%$) intervals, so stratification is necessary to interpret these extremes.

Across PIG, PICP errors are -21.5% for the 5^{th}-95^{th} interval, -19.0% for the $\leq 50^{\text{th}}$ quantile, +21.3% for the $\geq 95^{\text{th}}$ quantile, and -0.25% for the $\leq 5^{\text{th}}$ quantile. In other words, on average, the central interval captures only 68.5% of ICESat-2 points compared with the nominal 90% target. These results mirror the same overall pattern that was observed for the 2023 Antarctic-wide assessment (Chapter 6), which instead used REMA as the reference dataset (-10.4%, -8.3%, +8.5%, and +1.9% respectively), though magnitudes here are roughly doubled. As before, the lower tail ($\leq 5^{\text{th}}$) is best calibrated, the central and median intervals are systematically under-covered, and the upper tail ($\geq 95^{\text{th}}$) is consistently over-covered.

Stratification by time using 3-month bins - chosen to provide sufficient sample density per interval while resolving seasonal cycles - revealed no significant temporal trends across the 2019-2024 period (Figure 7.3a), which is encouraging given the intended use of the data to compute rates of elevation change through time.

Topographic dependencies mirror those found in the Antarctic-wide (REMA-based) analysis, but with larger magnitudes. Central coverage (5^{th}-95^{th}) declines at an increasing rate with slope, starting in the region of -15 to -20%, and reaching -60% beyond 0.9° (Figure 7.3b). The $\leq 5^{\text{th}}$ quantile shows similar trends with respect to slope, albeit starting from 0% at low slope and then increasing to in excess of +50% error for steeper slopes. In contrast, the $\geq 95^{\text{th}}$ quantile remains comparatively stable and even improves slightly with slope. The dependencies with respect to surface roughness also show a similar pattern (Figure 7.3c), which is unsurprising given the high degree of correlation between slope and roughness ($r = 0.794$), identified in the previous analysis (Section 7.3.1).

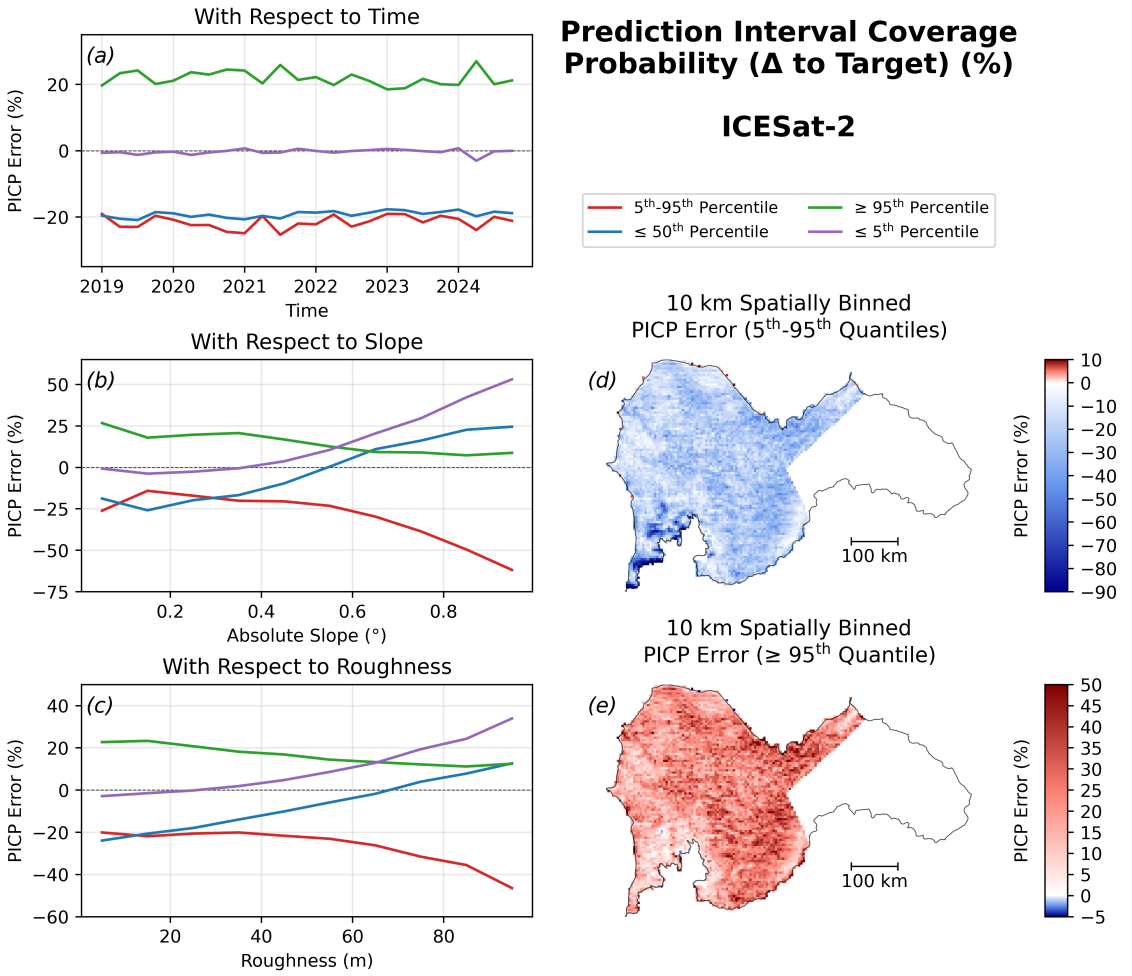

Figure 7.3: Prediction interval coverage probability (PICP) errors across Pine Island for model predictions in comparison to ICESat-2 measurements (6 years; 2019-2024). PICP errors are expressed as deviations from nominal coverage levels. Values are presented for the (1) central 90% interval (5^{th}-95^{th} quantiles), (2) $\leq 50^{th}$ quantile, (3) $\geq 95^{th}$ quantile, and (4) $\leq 5^{th}$ quantile. (a) PICP error as a function of time over all intervals. (b) PICP error as a function of surface slope over all intervals. (c) PICP error as a function of surface roughness over all intervals. (d) Spatial distribution of PICP error for the 5^{th}-95^{th} interval, binned at 10 km resolution. (e) Spatial distribution of PICP error for the $\geq 95^{th}$ quantile, binned at 10 km resolution.

The spatial patterns of PICP error reinforce these findings. Central coverage (5^{th}-95^{th}) is most negative in topographically complex sectors, particularly in the northern part of the catchment close to the Pine Island ice shelf (i.e. the bottom left region of Figure 7.3d). Broad similarities exist between the spatial

patterns of central coverage and the $\geq 95^{\text{th}}$ quantile (Figure 7.3e), but without such pronounced errors in the aforementioned region for the latter. This aligns with the previous observation of the increased robustness of the upper quantile with respect to topographic complexity. Notably, the strong negative anomalies seen in the Antarctic-wide REMA validation are absent, supporting the interpretation that much of the apparent miscalibration in the upper quantile in Chapter 6 arose from temporal mismatch. In other words, in regions of sustained elevation loss, the 2015-averaged REMA mosaic is systematically higher than contemporaneous altimetry, which produced spurious biases into the comparison. In contrast, validation against ICESat-2, which is matched in time, removes this effect, demonstrating more consistent model performance in this dynamic environment. Errors remain, however, indicating that DEM mismatch is not the sole source of miscalibration.

Overall, these results - which compare the ensemble predictions to contemporaneous ICESat-2 reference measurements over Pine Island Glacier - confirm the Antarctic-wide REMA findings reported in Chapter 6. This demonstrates that calibration characteristics transfer robustly to an unseen region, which is important for downstream applications mapping surface elevation change where topography differs from that seen during training. Furthermore, the 95^{th} quantile again proves stable across varying topographic complexity, reflecting its physical anchoring in the leading-edge return. In contrast, the central and median intervals remain systematically under-covered, reflecting predictions that are too narrowly constrained. The larger error magnitudes over PIG can be partly attributed to its prevalence of steep and rough terrain, though differences between ICESat-2 and REMA validation datasets also may contribute.

The temporal alignment of ICESat-2 gives confidence that the biases identified here represent genuine calibration tendencies rather than artefacts of DEM mismatch. At the same time, the agreement in error structure between ICESat-2 (Pine Island) and REMA (Antarctic-wide) validations indicates consistent model behaviour across independent datasets. The Pine Island case therefore illustrates

both the strengths and limitations of the approach: calibration is broadly stable and transferable, but quantiles remain moderately under-calibrated, with errors most pronounced in the lower and central parts of the distribution over increasingly complex terrain.

7.3.2.2 Ice Sheet Surface Elevation Change

Having established the calibration characteristics of the probabilistic predictions over Pine Island Glacier, I now use the complete Pine Island Glacier dataset to compute rates of ice sheet elevation change for the period 2019-2024, and compare these new estimates to equivalent measurements derived from mature data sources. This assessment is critical for demonstrating the practical utility of the deep learning framework: while individual, distributional predictions provide valuable uncertainty information, their ultimate value depends on whether they can produce scientifically meaningful estimates of glaciological processes when aggregated across space and time.

To evaluate this capability, I generate elevation change rate (dh/dt) maps and elevation change time series (dh) using the processing pipeline described in Section 7.2.2.3. My ensemble predictions are compared against three independent altimetric datasets: ICESat-2 (treated as the validation baseline due to its centimetre-level accuracy), Cryo-TEMPO L2+ POCA (representing conventional interferometric processing), and Cryo-TEMPO EOLIS (providing interferometric swath solutions). This multi-dataset comparison allows assessment of both the absolute accuracy of my approach and its performance relative to existing operational products, noting that the approach does not make use of any interferometric phase information.

Before presenting results, an important conceptual point requires clarification. Although the three predicted quantiles (5^{th}, 50^{th}, and 95^{th} quantiles) represent different distributional thresholds of the across-track elevation distribution within each footprint, they should exhibit similar rates of temporal change. This is because the quantiles relate to the same underlying surface, and while their absolute

elevations differ, the temporal evolution of the ice surface should be consistent across the footprint. The quantiles thus provide multiple estimates of the same glaciological signal, with their spread offering an inherent measure of prediction uncertainty. Figure 7.4 presents dh/dt maps of the Pine Island Glacier catchment derived from each dataset, with my ensemble results computed as the mean across the three predicted quantiles. This averaging approach reduces noise while preserving the underlying signal, as the quantiles should converge on similar change rates despite their different absolute elevations.

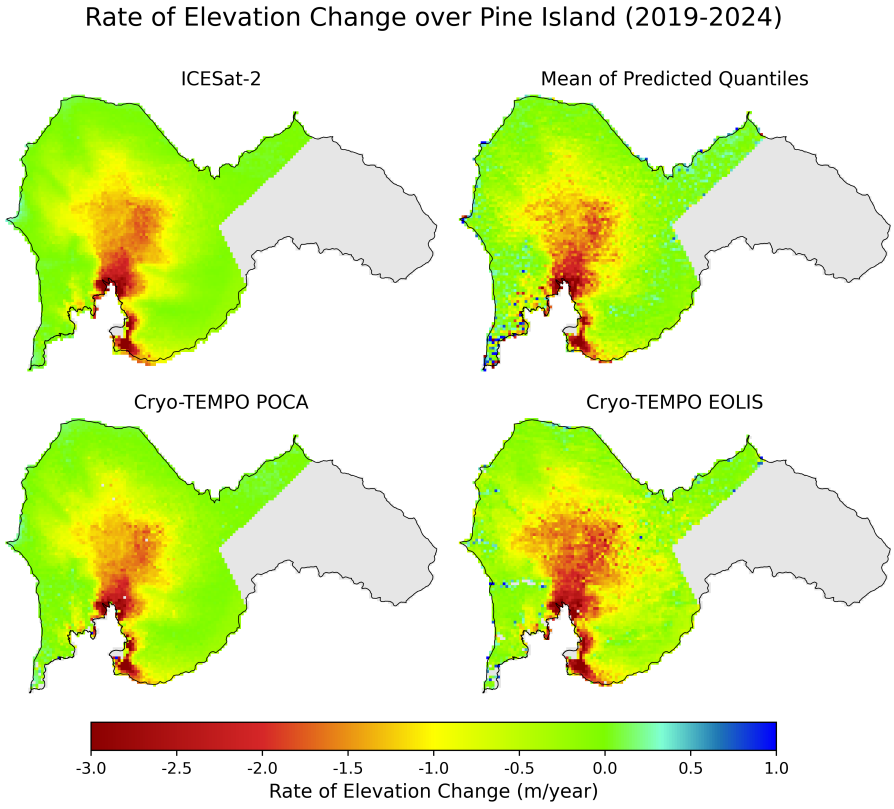

Figure 7.4: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of Pine Island Glacier (2019-2024). Top left: ICESat-2 baseline solution, serving as the reference by which the other solutions are compared to. Top right: Mean of 5^{th}, 50^{th}, and 95^{th} predicted quantiles from my deep learning framework, which excludes any interferometric information. Bottom left: Cryo-TEMPO POCA interferometric measurements. Bottom right: Cryo-TEMPO EOLIS interferometric swath measurements.

The spatial patterns of elevation change are consistent across all four datasets

(Figure 7.4), clearly reproducing the well-established signals of dynamical ice loss across Pine Island Glacier (Reed et al., 2024; Wingham et al., 2009). More specifically, all approaches capture the characteristic pattern of rapid thinning extending from the grounding line inland along the glacier trunk and tributaries, with rates exceeding 2-3 m yr^{-1} in the fastest-changing sectors. The deep learning framework demonstrates methodological stability by consistently producing near-zero elevation change rates in inland regions, in agreement with ICESat-2 observations. This indicates the approach is internally consistent, being appropriately sensitive to changes in the waveforms. The patterns identified are consistent across all methodologies and align well with previously published results that document the impact of dynamical imbalance across this sector of the West Antarctic Ice Sheet (IMBIE Team, 2020; Park et al., 2013; Reed et al., 2024; Wingham et al., 2009).

However, notable differences in data quality are apparent. ICESat-2 and POCA produce smooth, relatively noise-free elevation change fields, whereas EOLIS exhibits a greater number of noisy pixels and small patches of missing data. The ensemble approach likewise appears noisier than ICESat-2 or POCA, particularly in the topographically complex north-western coastal regions, although outside these areas it shows comparable or marginally lower sporadic noise than EOLIS. This pattern may partly reflect differences in measurement approach. POCA is anchored to the waveform leading edge, which generally provides the clearest and least ambiguous component of the radar return. Methods that exploit non-POCA portions of the waveform, including swath-based approaches such as EOLIS and the ensemble method used here, may therefore be more susceptible to noise. ICESat-2, by contrast, benefits from a different measurement geometry and the high ranging precision of photon-counting laser altimetry, which likely contributes to its comparatively smooth fields. The spatial pattern of elevated noise in the ensemble dh/dt estimates closely matches the regions of higher PICP errors identified in Figure 7.3d, indicating that prediction quality degrades in areas of complex surface topography and that this, in turn, affects the derived elevation-change trends.

Despite this, it is important to note that this level of performance is achieved without the use of interferometric information incorporated by both POCA and EOLIS, and relative performance should therefore be interpreted in light of this fundamental difference.

Despite these differences, the large scale pattern and spatial extent of detected elevation change is consistent. My approach, ICESat-2, and POCA all delineate similar boundaries for the region of significant mass loss, while EOLIS appears to encompass a slightly larger area, particularly toward the central trunk where the most rapid thinning occurs.

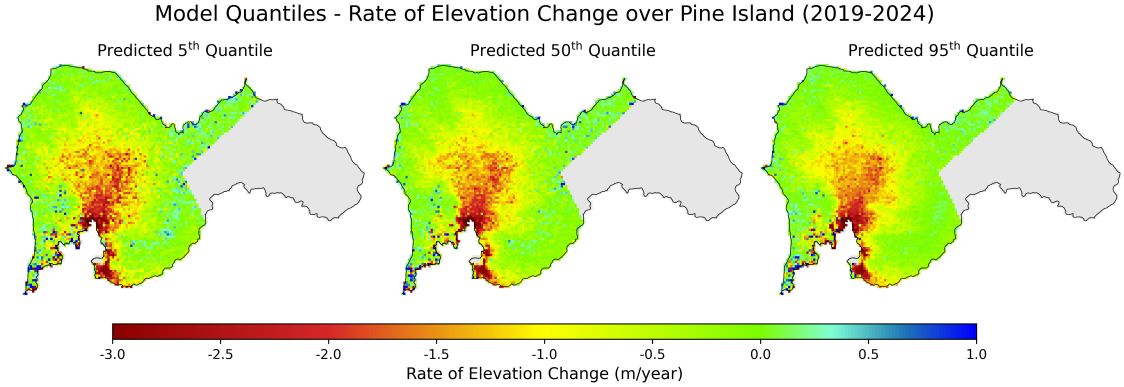

Figure 7.5: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of Pine Island Glacier (2019-2024) for individual 5^{th}, 50^{th}, and 95^{th} predicted quantiles from my deep learning framework.

Figure 7.5 examines the individual performance of each predicted quantile, prior to the averaging presented in Figure 7.4, revealing systematic differences that align with the physical basis of radar altimetry. Spatial noise decreases progressively from the 5^{th}, to the 50^{th}, and to the 95^{th} quantile, with the latter producing the smoothest elevation change field. This pattern reinforces the finding from Chapter 6 that the 95^{th} quantile, anchored to the leading edge of the waveform, provides the most reliable and least ambiguous elevation information. Importantly, this noise is not entirely randomly distributed but concentrates in specific geographic areas. The previously-identified north-western coastal region shows degraded performance

across all quantiles, while other noise appears as isolated pixels scattered throughout the domain. This suggests that the systematic errors are largely driven by more challenging surface conditions.

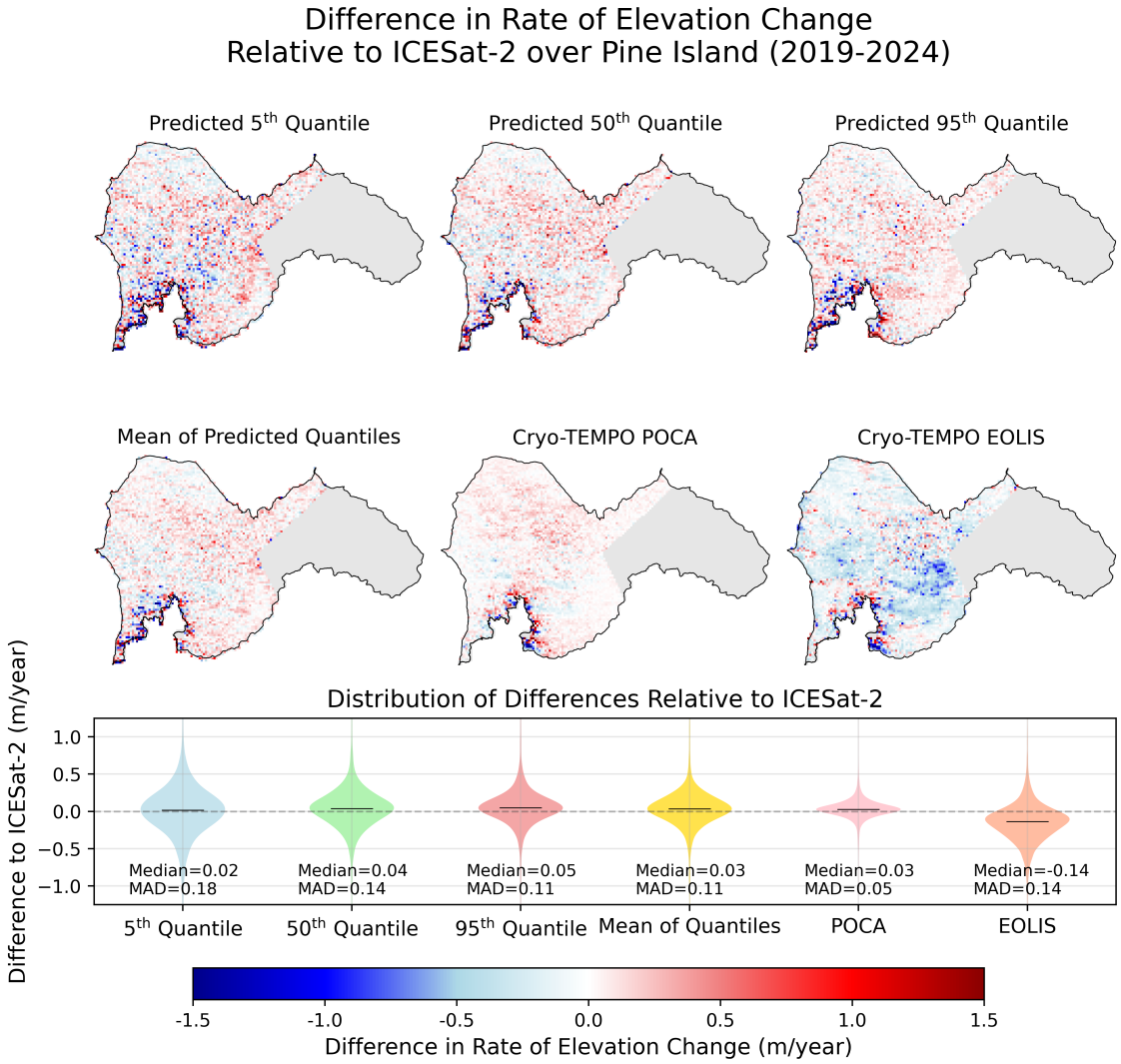

Figure 7.6: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of Pine Island Glacier (2019-2024), differenced to the ICESat-2 baseline. Top row: Individual 5^{th}, 50^{th}, and 95^{th} quantiles from my deep learning framework. Middle row: Mean of quantiles from my deep learning framework (left), Cryo-TEMPO POCA solution (centre), and Cryo-TEMPO EOLIS solution (right). Bottom row: The overall distributions of the elevation rate differences, relative to ICESat-2.

Figure 7.6 quantifies these qualitative observations by mapping the differences

between each approach and the reference ICESat-2 baseline dataset. Firstly, intercomparing the individual quantile solutions confirms the previous visual assessment; namely (1) that prediction quality improves systematically from the 5^{th} to the 95^{th} quantile, (2) that taking the mean across quantiles produces the best overall performance, and (3) that the greatest disagreement occurs in the previously identified north-western coastal area.

The statistical distributions of these differences (Figure 7.6, bottom row) provide additional insight into relative performance. The median absolute differences (MADs) decrease progressively with quantile rank, from the 5^{th} through to the 95^{th} quantile. Interestingly, the median differences show the opposite trend, performing better for the lower quantiles. Taking the average across quantiles yields the best combination of both metrics, suggesting that ensemble averaging effectively balances both the systematic biases present in individual quantiles and the random noise inherent in each solution.

Comparing my new approach to the other more established radar altimetry methods that, additionally, utilise interferometric phase information, I find that whilst the POCA solution achieves the best overall agreement with ICESat-2, exhibiting both a very low MAD and a median difference, my method outperforms the EOLIS solution in terms of its agreement with respect to ICESat-2. Specifically, EOLIS shows both a larger MAD in comparison to my ensemble average (comparable to the 50^{th} quantile alone) and a greater magnitude bias of -0.14 m yr^{-1}. This systematic offset is visible both in the statistical distribution and also in the spatial difference patterns (Figure 7.6). By further quantifying the difference between the EOLIS and quantile-averaged solutions, I find that my framework reduces the bias relative to ICESat-2 by 78.6% and the MAD by 21.4%.

Figure 7.7 presents the elevation change time series for Pine Island Glacier, spatially integrated across the entire domain to show cumulative change since 2019. These series are constructed using 30-day epochs for the deep learning framework, the ICESat-2 solution, and the EOLIS product; while the POCA solution employs

60-day epochs due to its lower data density. All series are normalised to begin at zero elevation change, providing a clear comparison of temporal evolution across this glaciological basin.

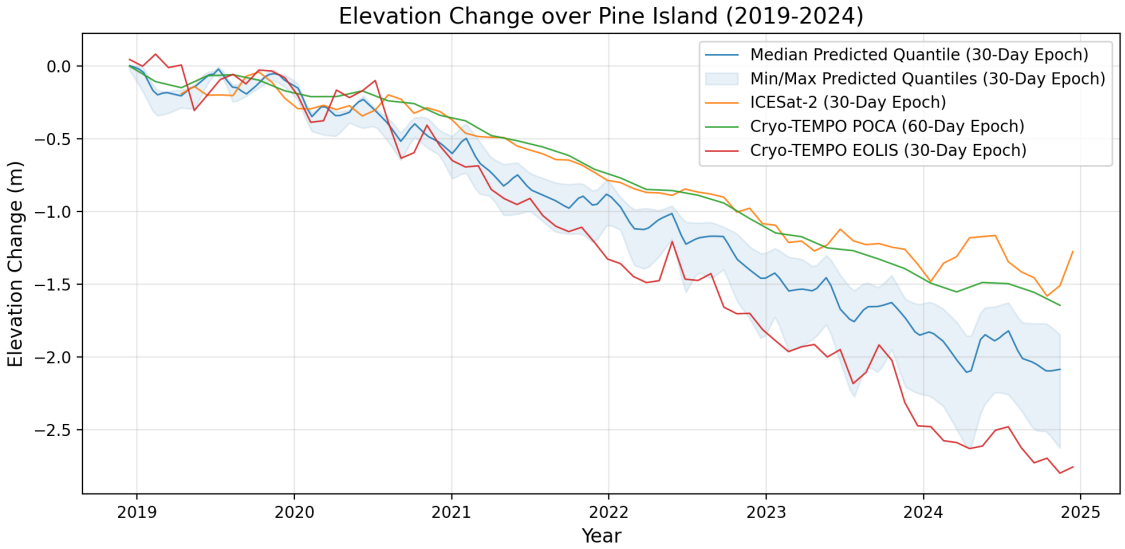

Figure 7.7: Elevation change time series over the CryoSat-2 SARIn zone of Pine Island Glacier (2019-2024), normalised to start at zero. Blue line shows the median across ensemble quantiles with min/max uncertainty envelope (shaded). Green: Cryo-TEMPO POCA (60-day epochs). Red: Cryo-TEMPO EOLIS (30-day epochs). Orange: ICESat-2 (30-day epochs). Each point represents the spatial median across 5 km grid cells for that epoch.

For my ensemble predictions, I plot the minimum, median, and maximum elevation change across the three quantiles at each epoch, creating an uncertainty envelope that reflects both the spread in quantile predictions and their temporal evolution. The ordering of quantiles is not explicitly fixed through time, as their relative performance can vary depending on surface conditions and waveform characteristics at different epochs. However, in practice, the three quantiles were found to maintain a consistent hierarchy throughout the observation period, with the 5th quantile predicting the steepest elevation loss rates (bottom line), the 50th quantile showing intermediate rates (middle line), and the 95th quantile predicting the shallowest losses (top line). This temporal analysis reveals several important patterns. ICESat-2 and POCA show strong agreement throughout the

observation period, with POCA appearing somewhat smoothed due to its longer temporal binning period. Notably, the first three months of ICESat-2 data required removal due to excessive noise that was orders of magnitude larger than subsequent measurements, likely due to satellite commissioning activities early in the mission.

The ensemble predictions show systematically more negative absolute elevation changes compared to ICESat-2 and POCA, indicating a slightly higher region-wide estimate of thinning rates. The uncertainty envelope formed by the quantile spread grows progressively through time, suggesting that the quantiles themselves develop slightly different estimates of the overall elevation change rate. Despite this systematic offset, the temporal variability closely matches the ICESat-2 pattern, particularly towards the end of the period when seasonal peaks and troughs align.

EOLIS produces the most negative elevation changes, indicating the highest absolute estimate of regional thinning rates. Throughout the time series, EOLIS tracks the temporal variability seen in other datasets (especially towards the end of the period) but maintains a consistent, increasingly negative offset relative to both ICESat-2 and POCA. Particularly noteworthy is the similarity in temporal variability between my predictions and EOLIS, with both datasets capturing comparable seasonal and interannual fluctuations throughout the observation period. This correspondence is notable given that my approach relies solely on power waveform information while EOLIS incorporates additional interferometric phase data to resolve across-track positioning.

The elevation change analysis demonstrates several key capabilities of my deep learning framework. First, the approach successfully recovers the established spatial and temporal patterns of Pine Island Glacier dynamics without any prior training on this region. The consistency of these patterns relative to multiple independent validation datasets provides strong evidence that the probabilistic predictions contain meaningful glaciological information relating to ice sheet elevation and, crucially, elevation change. This suggests that, when aggregated across many observations, the ambiguity inherent in individual probabilistic predictions reduces,

such that estimates closely match point-based operational products, demonstrating that sufficient information is retained to resolve glaciological signals at relevant monitoring scales.

Second, the systematic differences between quantiles reveal important insights into the physical basis of the predictions. The improved performance of the 95^{th} quantile aligns with radar altimetry theory: the leading edge of the waveform corresponds to the first returns from the highest elevations within the footprint, providing the clearest and least ambiguous signal (Bamber, 1994; Scott et al., 1994). This is visible in Figure 7.5, where the 95^{th} quantile produces smoother spatial patterns of elevation change, and in Figure 7.6, where it achieves progressively lower MAD values relative to ICESat-2. Lower quantiles integrate progressively more complex and mixed returns from across the footprint, leading to increased uncertainty but still preserving the essence of the temporal trend in elevation through time.

The comparison with EOLIS is particularly encouraging. Despite EOLIS having access to interferometric phase information, which is intentionally withheld from the deep learning model, my ensemble predictions show comparable or slightly better agreement with the ICESat-2 baseline in terms of both spatial patterns and temporal evolution. This demonstrates the potential of my probabilistic framework, which can respond to subtle changes in waveform characteristics that correlate with sub-meter annual elevation changes, despite predicting distributions often spanning tens of meters in elevation.

These results establish the practical utility of the deep learning framework for ice sheet elevation change monitoring. While predictions remain noisier than mature operational products, particularly in topographically complex regions, the approach demonstrates clear capability for recovering scientifically meaningful glaciological signals, especially considering that no interferometric phase information is utilised. When applied to estimate elevation change, performance is comparable with interferometric processing despite using only power waveform information,

suggesting potential for enhancing the scientific value of both historical and future non-interferometric altimetric observations.

7.3.3 Analysis over Greenland Ice Sheet

Having demonstrated the framework's capabilities over the Pine Island Glacier, I finally turn to evaluating its performance across the Greenland Ice Sheet, a fundamentally different glaciological environment that provides an extremely stringent test of model generalisability. Unlike the Antarctic training domain, Greenland is characterised by steeper topographic gradients, narrower outlet glaciers, and extensive seasonal surface melting that creates distinct radar scattering conditions. These differences pose significant challenges for transferring learned relationships between waveforms and surface topography to a completely unseen domain. It is important to note that this analysis applies the model trained exclusively on Antarctic Ice Sheet data directly to Greenland without any retraining, which differs from how one would deploy the framework operationally where training would incorporate Greenland-specific data. The Greenland validation therefore serves as a theoretical assessment of out-of-distribution generalisability rather than an ultimate measure of operational performance over Greenland. Within this analysis, I therefore apply the same evaluation methodology used for Pine Island Glacier, assessing PICP against ICESat-2 elevation measurements, and then comparing derived elevation change estimates with operational altimetric datasets to quantify both statistical calibration and investigate broader glaciological utility in this contrasting environment.

7.3.3.1 Prediction Interval Coverage Probability

Across the Greenland Ice Sheet, region-wide PICP errors are -22.9% for the 5^{th}-95^{th} interval, +4.3% for the $\leq 50^{\text{th}}$ quantile, +5.8% for the $\geq 95^{\text{th}}$ quantile, and +15.6 for the $\leq 5^{\text{th}}$ quantile. At face value, these differ substantially from those reported for Pine Island Glacier (-21.5%, -19.0%, +21.3%, -0.25%) and from the Antarctic-wide

REMA assessment (-10.4%, -8.3%, +8.5%, +1.9%). Once stratified by slope and roughness however, the underlying calibration structure aligns closely with findings from both Antarctica and Pine Island Glacier.

Over low slopes, PICP errors approach values comparable to the Antarctic-wide REMA case (Figure 7.8b). As surface complexity increases, central coverage (5^{th}-95^{th}) declines sharply, reaching magnitudes exceeding -60% for slopes steeper than 0.9°, matching the level of under-coverage observed in Pine Island's steepest terrain. This decline is driven primarily by systematic under-placement of the 5^{th} quantile, which shows errors exceeding +50% in the most complex topography, while the 95^{th} quantile maintains stability across the full range of surface conditions, as observed with both the GIS and AIS cases. Similar patterns emerge when stratifying by surface roughness (Figure 7.8c). Smooth surfaces show PICP errors comparable to those found with Antarctic, while much rougher surfaces exhibit degradation magnitudes similar to those observed in Pine Island's upper bounds. The consistency of these slope and roughness dependencies across all three study regions demonstrates that the differences in region-wide statistics reflect Greenland's topographic composition rather than a fundamental breakdown in model generalisability. As established in Section 7.3.1, Greenland exhibits median slopes and roughness values over twice and nearly three times higher than Antarctica respectively, naturally shifting the ice sheet toward regimes where model calibration is known to degrade.

Temporal analysis reveals no evidence of long-term bias or drift in PICP across the 2019-2024 observation period, but does show a pronounced seasonal cycle affecting all quantiles (Figure 7.8a). The central interval (5^{th}-95^{th}), median ($\leq 50^{\text{th}}$), and lower bound ($\leq 5^{\text{th}}$) all exhibit systematic lowering in coverage errors during summer months (May-September) and raising during winter months (November-March), with peak-to-trough variations reaching 5-10%. The upper bound ($\geq 95^{\text{th}}$ quantile) demonstrates the inverse behaviour, with raised coverage error during summer and lowered during winter. These seasonal variations align closely with

expected patterns of surface elevation change driven by accumulation and ablation cycles (Slater et al., 2021). The consistent nature of these oscillations suggests that they reflect genuine seasonal elevation changes rather than calibration instabilities, although further more detailed analysis would be needed to fully understand their origin.

Spatially, the most severe central coverage errors (approaching -90% in many locations) cluster around the ice sheet margins (Figure 7.8d). These patterns mirror the behaviour observed in topographically complex Antarctic regions such as the Transantarctic Mountains and Antarctic Peninsula, but cover a substantially larger fraction of the ice sheet due to Greenland's more constrained topographic setting and steeper peripheral zones. Interior regions, particularly across the central plateau, show near-zero or slightly positive central coverage errors, indicating distributions that are appropriately calibrated or slightly over-conservative. It is also worth reiterating that these marginal regions represent out-of-distribution conditions for the Antarctic-trained model, which may contribute to any observed performance degradation.

For the upper bound ($\geq 95^{\text{th}}$ quantile), spatial patterns show less dramatic variation, in-line with those observed for Pine Island and the AIS (Figure 7.8e). Flatter interior regions where central coverage is near-optimal often exhibit the maximum negative error (-5%), indicating that the 95^{th} quantile exceeds all ICESat-2 observations, though the actual scale of this misplacement is indeterminate. Unlike the Antarctic REMA case, these spatial patterns show no clear correlation with regions of documented high mass loss, such as the major outlet glacier systems of Jakobshavn Isbræ, Helheim Glacier, or Kangerdlugssuaq Glacier, confirming the suggested hypothesis that temporal mismatches between mosaic and altimetric data were a significant source of apparent miscalibration in the earlier analyses with REMA. Much of the remaining spatial structure appears as unresolved variability without clear topographic or glaciological associations, consistent with findings across all validation regions.

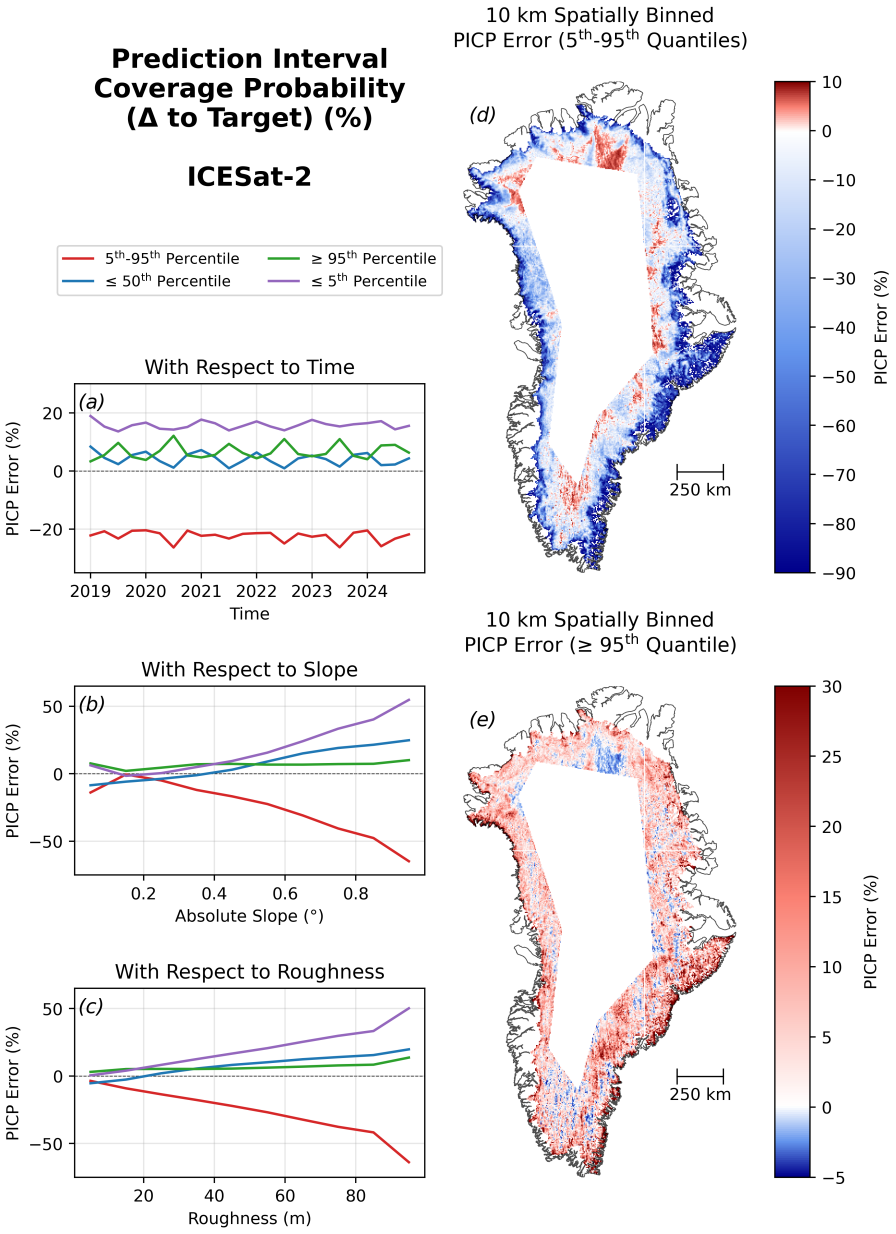

Figure 7.8: Prediction interval coverage probability (PICP) errors across the CryoSat-2 SARIn zone of the Greenland Ice Sheet for model predictions that are compared to ICESat-2 measurements (6 years; 2019-2024). PICP errors are expressed as deviations from nominal coverage levels. Values are presented for the (1) central 90% interval (5^{th}-95^{th} quantiles), (2) $\leq 50^{\text{th}}$ quantile, (3) $\geq 95^{\text{th}}$ quantile, and (4) $\leq 5^{\text{th}}$ quantile. (a) PICP error as a function of time over all intervals. (b) PICP error as a function of surface slope over all intervals. (c) PICP error as a function of surface roughness over all intervals. (d) Spatial distribution of PICP error for the 5^{th}-95^{th} interval, binned at 10 km resolution. (e) Spatial distribution of PICP error for the $\geq 95^{\text{th}}$ quantile, binned at 10 km resolution.

The replication of fundamental calibration characteristics across Greenland - broad under-coverage that degrades with topographic complexity, robustness of the 95th quantile, and spatial patterns controlled by surface slope and roughness - provides strong evidence that the deep learning framework has successfully generalised beyond its Antarctic training domain. The seasonal cycles unique to Greenland's more temperate environment (Slater et al., 2021) offer additional insights into model behaviour under conditions of active surface mass balance, while the overall consistency of topographic dependencies demonstrates that the underlying physical relationships between waveform characteristics and surface geometry have been successfully captured by the ensemble approach.

7.3.3.2 Ice Sheet Surface Elevation Change

I now assess the framework's ability to detect elevation change patterns across Greenland's distinct glaciological environment. The evaluation follows the same analytical framework employed for Pine Island Glacier, generating elevation change rate (dh/dt) maps and time series (dh) using the processing pipeline described in Section 7.2.2.3. My ensemble predictions are compared against equivalent ICESat-2, Cryo-TEMPO L2+ POCA, and Cryo-TEMPO EOLIS derived estimates across the 2019-2024 observation period. As with Pine Island, the quantiles are processed independently and their mean calculated to provide the primary ensemble estimate, with the quantile spread serving as an inherent uncertainty measure.

Figure 7.9 presents ice-sheet-wide dh/dt maps derived from each dataset, revealing the established spatial patterns of Greenland mass loss documented extensively in the literature (Slater et al., 2020; Otosaka et al., 2023; IMBIE Team, 2020; Broeke et al., 2016; Hanna et al., 2020; Simonsen et al., 2021; Nilsson et al., 2024). All four approaches, including my ensemble, successfully capture the characteristic zones of rapid thinning around the ice sheet periphery, where major outlet glacier systems that are currently in a state of dynamic imbalance drain the interior ice mass toward the ocean. The southern coastal regions exhibit the most

widespread thinning rates, commonly exceeding 1-2 m yr^{-1}, while individual outlet glaciers such as Jakobshavn Isbræ, Helheim Glacier, and Kangerdlugssuaq Glacier show localized thinning rates approaching 3 m yr^{-1}. The interior accumulation zone remains relatively stable, consistent with expected patterns of mass balance across the ice sheet. As with Pine Island, the framework shows stability by producing near-zero inland elevation change rates, consistent with ICESat-2, indicating internal consistency and proper sensitivity to waveform changes.

Despite this fundamental agreement in large-scale patterns, some differences in data quality are apparent across the datasets. As expected, ICESat-2 produces the smoothest and most coherent elevation change field, reflecting both its centimeter-level measurement precision and mature processing algorithms. CryoSat-2 POCA and EOLIS exhibit comparable smoothness across most regions, although with some localised, pixel-level anomalies in topographically complex coastal areas.

The ensemble approach successfully reproduces the expected glaciological signals and shows strong agreement with the operational datasets in identifying the major zones of mass loss. However, the predictions exhibit substantially more noise than the comparison datasets, particularly around the ice sheet margins and most prominently in the southern and eastern regions, approaching the coast. This noise, which often manifests as apparent elevation gain signals of several meters per year, correlates strongly with areas of high topographic complexity identified in the slope and roughness analysis of Section 7.3.1. The spatial distribution of this noise closely resembles the degraded performance observed in the southwestern coastal regions of Pine Island Glacier, suggesting similar underlying causes related to challenging surface conditions that exceed the framework's optimal operating regime.

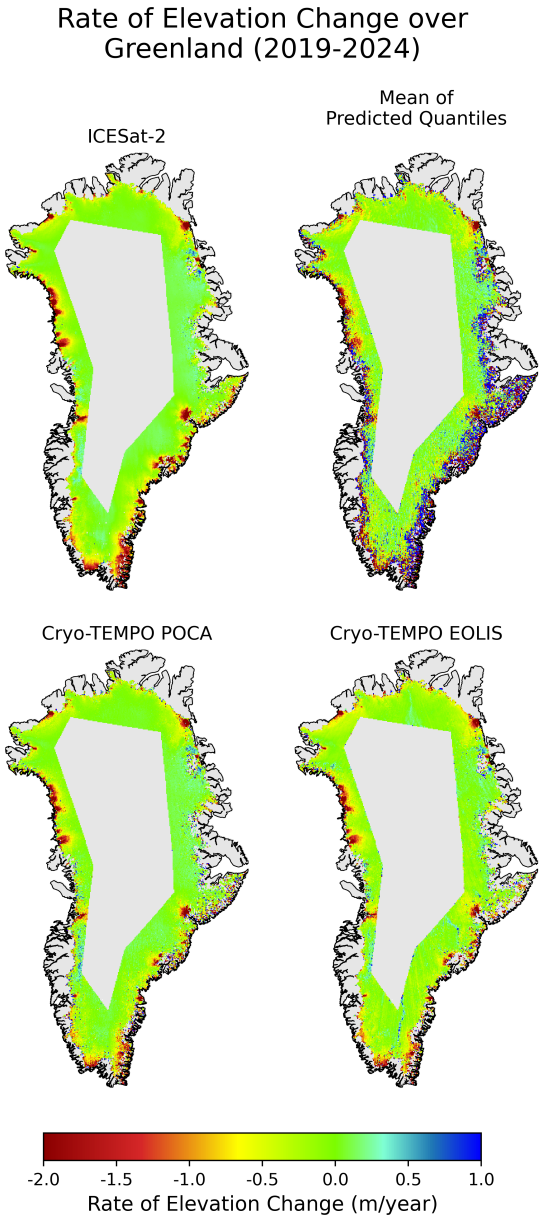

Figure 7.9: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of the Greenland Ice Sheet (2019-2024). Top left: ICESat-2 baseline solution, serving as the reference by which the other solutions are compared to. Top right: Mean of 5^{th}, 50^{th}, and 95^{th} predicted quantiles from my deep learning framework, which excludes any interferometric information. Bottom left: Cryo-TEMPO POCA interferometric measurements. Bottom right: Cryo-TEMPO EOLIS interferometric swath measurements.

The quantile-specific performance (Figure 7.10) reveals patterns broadly consistent with those observed for Pine Island (Section 7.3.2.2), though the relationships

are less pronounced due to the much larger spatial scale and greater topographic diversity of Greenland. The 5th quantile appears to exhibit the highest levels of noise, while the 95th quantile produces the smoothest elevation change field, although this overall distinction is not as clear as with the Pine Island counterpart. All three quantiles show agreement in the spatial distribution of problematic regions, with the systematic positive elevation change anomalies appearing consistently across the 5th, 50th, and 95th percentile predictions.

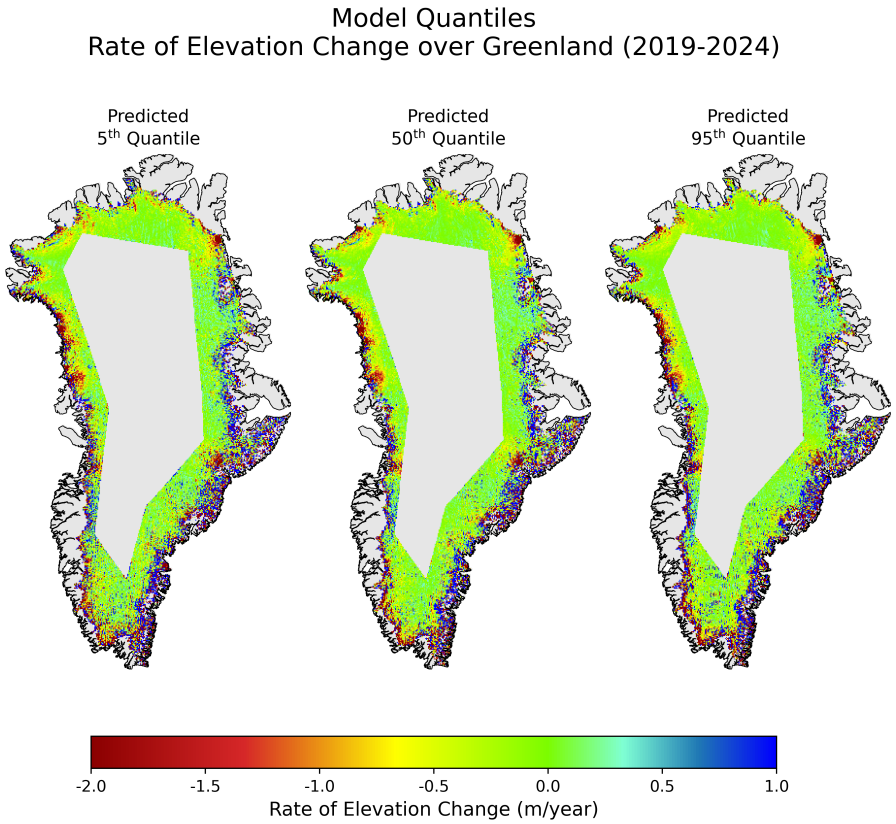

Figure 7.10: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of the Greenland Ice Sheet (2019-2024) for individual 5th, 50th, and 95th predicted quantiles from my deep learning framework.

The difference maps relative to ICESat-2 (Figure 7.11) quantify these qualitative observations and provide statistical measures of relative performance. The ensemble quantiles achieve good agreement with ICESat-2 across large portions of the study region, with performance comparable to other solutions in many areas. The main

disagreements are concentrated close to the margins, and particularly in the southern and eastern regions, where the steep topography is expected to degrade model performance, with differences often exceeding ±3 m yr^{-1} in these most affected areas. Despite these localised disagreements, the quantile-specific statistics reveal minimal systematic bias in the overall regional estimates, with median differences of -0.03, -0.03, and -0.02 m yr^{-1} for the 5^{th}, 50^{th}, and 95^{th} percentiles respectively. The MADs are larger than what was seen in Pine Island, but similarly decrease progressively with quantile rank (0.29, 0.22, 0.21 m yr^{-1} for the 5^{th}, 50^{th} and 95^{th} quantiles, respectively).

Taking the mean across quantiles achieves the same median difference and MAD as the 95^{th} percentile alone (-0.02 m yr^{-1} median, 0.21 m yr^{-1} MAD), suggesting that the averaging approach provides limited additional benefit over Greenland as compared to Pine Island Glacier. This may reflect the more challenging surface conditions encountered across much of Greenland's periphery, where systematic errors in individual quantiles persist rather than averaging out. Comparing the different solutions, Cryo-TEMPO POCA demonstrates the closest agreement with ICESat-2, exhibiting a median difference of -0.01 m yr^{-1} and MAD of 0.08 m yr^{-1}. In contrast, EOLIS shows a systematic negative bias (-0.09 m yr^{-1} median difference) that is greater than the bias apparent in my averaged quantile solution, similar to the behaviour observed for Pine Island Glacier. However, it achieves a lower MAD (0.11 m yr^{-1}) than any of the ensemble quantiles, indicating more spatially consistent performance despite the overall bias. In this case, the average of the predicted quantiles represents a 50.0% reduction in bias relative to ICESat-2 when compared to EOLIS, but increases the MAD by 71.4%.

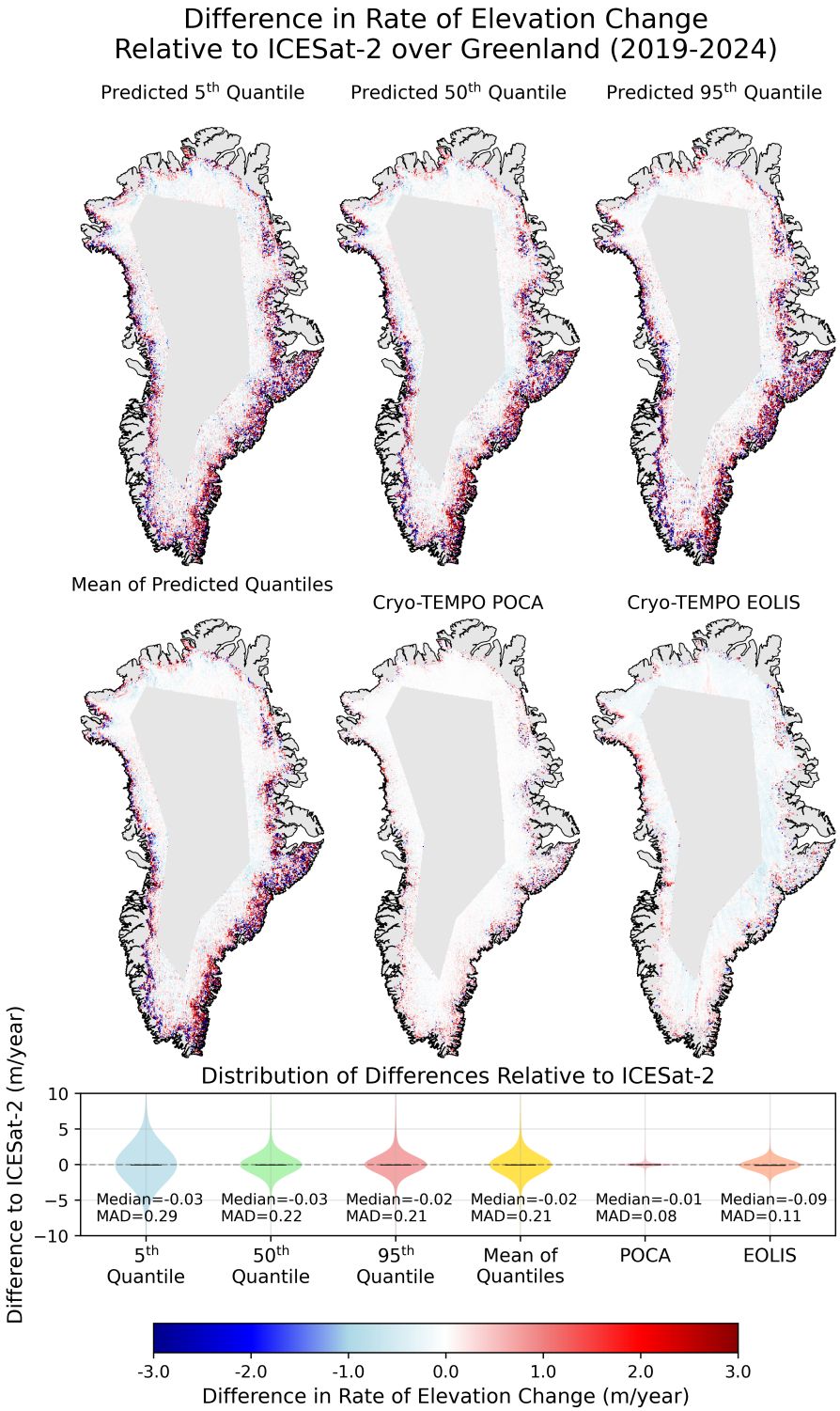

Figure 7.11: 5 km-gridded rates of elevation change (dh/dt) over the CryoSat-2 SARIn zone of the Greenland Ice Sheet (2019-2024), differenced to ICESat-2 baseline. Top row: Individual 5^{th}, 50^{th}, and 95^{th}.

The temporal analysis (Figure 7.12) reveals patterns that mirror those observed for Pine Island (Section 7.3.2.2) while providing additional insights into model behaviour at ice-sheet scales. The ICESat-2 and POCA solutions again show strong temporal agreement throughout the observation period. Encouragingly, my ensemble prediction also captures the overall downward trend seen in these datasets, with the uncertainty envelope formed by the quantile minimum and maximum encompassing both ICESat-2 and POCA measurements for nearly the entire time series. This represents increased temporal agreement than observed for Pine Island Glacier, and suggests that although I observe regions of noise in some sectors, my model is able to accurately resolve the net volume change of the ice sheet. As with Pine Island Glacier, the three quantiles maintain a consistent hierarchy throughout the observation period, with the 5th, 50th, and 95th quantiles predicting successively shallower elevation loss rates.

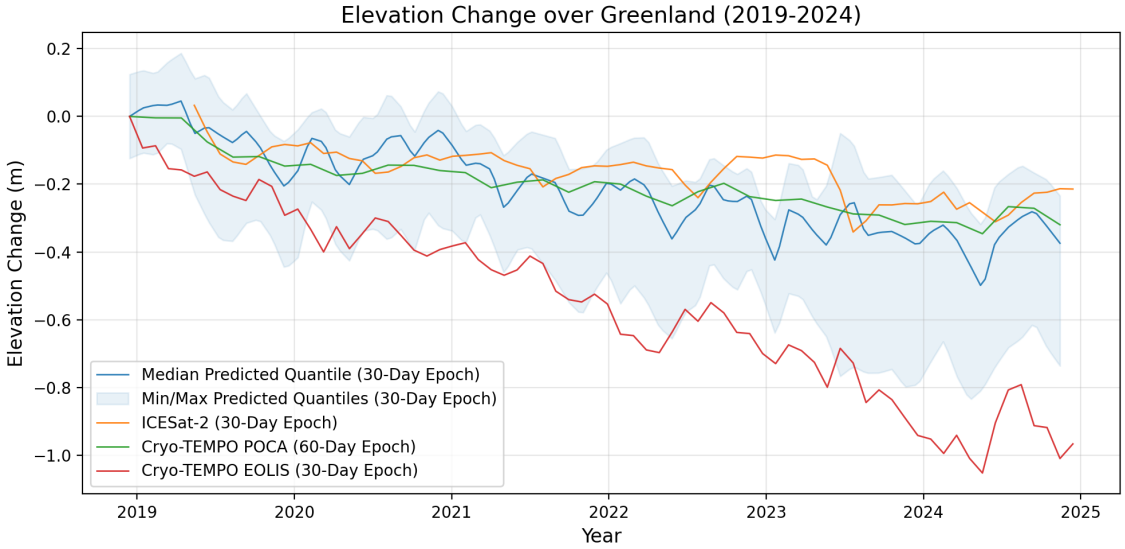

Figure 7.12: Elevation change time series over the CryoSat-2 SARIn zone of the Greenland Ice Sheet (2019-2024), normalised to zero at start. Blue line shows median across ensemble quantiles with min/max uncertainty envelope (shaded). Green: Cryo-TEMPO POCA solution (60-day epochs). Red: Cryo-TEMPO EOLIS solution (30-day epochs). Orange: ICESat-2 solution (30-day epochs). Each point represents the median across 5 km grid cells for that epoch.

As was observed for Pine Island Glacier, EOLIS again exhibits the most negative

regional elevation change estimates, maintaining a systematic bias that places it outside the ensemble uncertainty bounds for most of the observation period after mid-2021. The consistency of this EOLIS offset across both Pine Island Glacier and Greenland likely suggests a fundamental difference in measurement philosophy or processing approach rather than region-specific calibration issues.

The regional integration of elevation change across Greenland (Figure 7.12) supports the key finding from Pine Island Glacier (Section 7.3.2.2); namely, that distributional ambiguity reduces into coherent signals of elevation change when aggregated across space and time. The agreement between my ensemble mean and both ICESat-2 and POCA demonstrates successful convergence despite substantially higher noise in individual grid cells, indicating that essential physical relationships between waveform characteristics and surface evolution are preserved even in this fundamentally different ice sheet environment. Also consistent with Pine Island results (Section 7.3.2.2), my ensemble demonstrates closer temporal correspondence with ICESat-2 than with EOLIS, despite using solely power waveform data compared to EOLIS's additional interferometric phase information. This reinforces the framework's promising sensitivity to subtle waveform changes correlating with sub-meter elevation variations, despite predicting distributions spanning tens of meters - a capability that I demonstrate here transfers between polar environments.

However, Greenland reveals new limitations not evident over Pine Island. Systematic elevation gain anomalies in southern and eastern Greenland represent significant data quality degradation in areas of extremely high topographic complexity. Their correlation with surface complexity provides guidance for identifying regions that require cautious interpretation and further model refinement. This likely also reflects the fact that, although Pine Island was withheld from model training, it still belongs to the Antarctic Ice Sheet (AIS) and therefore remains more consistent with the training distribution, whereas Greenland represents cross-polar generalisation to a fundamentally different ice-sheet environment. That said, explicitly training future operational models on this presently out-of-distribution

region could reduce such artifacts. Despite these topographic limitations, the overall Greenland performance demonstrates good generalisability for an Antarctic-trained framework. The successful reproduction of established glaciological patterns and consistent calibration characteristics across fundamentally different environments provide compelling evidence that the ensemble has learned transferable physical relationships rather than region-specific artifacts. This transferability, established through both Pine Island and Greenland-wide validation, represents a crucial step toward operational applications that could utilise decades of underutilised altimetric data for enhanced ice-sheet monitoring.

7.4 Conclusion

This chapter has developed and assessed new estimates of ice sheet elevation and elevation change which have been derived, for the first time, using a probabilistic deep learning framework. Specifically, results were generated and evaluated over Pine Island Glacier in West Antarctica, and the Greenland Ice Sheet - regions withheld from the training process. The results demonstrate that the framework can extract meaningful glaciological signals from non-interferometric radar waveforms across fundamentally different polar environments, while also highlighting important operational constraints. To my knowledge, this is the first time that swath processing of non-interferometric waveforms has been successfully demonstrated.

Prediction interval coverage probability (PICP) analysis against ICESat-2 showed that the deep learning framework, which was trained on Antarctic data, exhibited consistent calibration behaviour across both Pine Island Glacier and the Greenland Ice Sheet: systematic under-coverage that increases with topographic complexity, alongside stability of the 95^{th} quantile. This consistency provides strong evidence that the ensemble has learned transferable physical relationships between radar waveforms and surface geometry, rather than overfitting to the Antarctic-specific conditions during training.

The framework's capacity to resolve spatial elevation change rates (dh/dt) over 2019-2024 was further validated against three independent altimetric datasets; namely dh/dt estimates that were derived from ICESat-2, Cryo-TEMPO L2+ POCA, and Cryo-TEMPO EOLIS. In this regard, my model clearly reproduced well-established patterns of ice sheet thinning, including 2-3 m yr^{-1} losses propagating inland from Pine Island Glacier's grounding line (Reed et al., 2024; Wingham et al., 2009; Park et al., 2013) and major outlet glacier losses across Greenland (Slater et al., 2021; IMBIE Team, 2020), showing good agreement with these other datasets. Aggregated elevation change time series (dh) also tracked seasonal variability in agreement with all three datasets, while the predicted 5^{th}, 50^{th}, and 95^{th} quantiles provided a natural uncertainty envelope capturing both model confidence and the plausible range of change. It was particularly encouraging to observe that the temporal patterns matched those from interferometric solutions despite my framework not being given access to phase information. Crucially, the ability to detect sub-meter annual elevation changes from probability distributions spanning tens of meters (1) demonstrates the ensemble's capacity to resolve subtle waveform variations and translate them into meaningful elevation signals, and (2) validates the central premise that waveform ambiguity constitutes exploitable information rather than a processing obstacle.

This chapter also contributes the first comprehensive roughness characterisation of the Greenland Ice Sheet, that is independent of slope, derived using Singular Value Decomposition techniques applied to the 100 m ArcticDEM mosaic. The approach directly parallels the analysis presented in Chapter 5 for Antarctica, which used the 100 m REMA DEM. The resulting 100 m-resolution slope and roughness maps reveal Greenland's markedly steeper (median slope 0.434° vs 0.192°) and rougher (median roughness 0.855 m vs 0.297 m) surface compared to Antarctica.

The competitive performance of my new deep learning approach relative to interferometric processing, achieved using only power waveform information, carries significant implications for historical, current, and future data utilisation. The

pre-2010 record from ERS-1, ERS-2, and Envisat missions consists solely of non-interferometric, Low Resolution Mode (LRM) altimetry. These represent decades of potentially underexploited observations that could extend elevation change records back to the early 1990s, though adaptation would be required for LRM rather than SAR processing. Similarly, operational Sentinel-3 satellites and the planned Sentinel-3 Next Generation (S3NG) satellites provide SAR coverage but lack interferometric capability (Egido, 2023), while the upcoming CRISTAL mission will only apply SARIn at Ku-band, not at Ka-band (Kern et al., 2020). All thus present further avenues for enhanced information extraction.

More broadly, this chapter demonstrates the value and potential of fundamentally alternative data-driven approaches to exploiting non-interferometric altimetry across polar regions. While the POCA approach continues to provide the most precise individual elevation measurements through its focus on the unambiguous leading-edge return, the deep learning framework offers a complementary and distinct advantage: a substantially higher density of elevation estimates across the full across-track swath, together with explicit uncertainty quantification. The practical value of this is most evident in data-sparse conditions, or for historical non-interferometric missions where POCA alone provides insufficient spatial coverage to resolve fine-scale glaciological processes. While limitations in complex terrain remain an obstacle to universal deployment, the demonstrated ability to recover elevation change signals from waveforms alone represents a significant methodological advance. As such, this new framework offers a promising route towards improved ice-sheet monitoring and, ultimately, more robust sea-level projections to support climate science and policy.

Chapter 8

Synthesis and Conclusions

This final chapter synthesises the principal findings from the preceding studies, discusses limitations of the methodologies developed, and outlines priorities for future research. It then concludes by reflecting on the contribution of this work toward unlocking new capabilities from non-interferometric SAR radar altimetry observations for ice sheet monitoring.

8.1 Overview

Satellite radar altimeters have provided a near-continuous record of polar ice sheets since 1991, revolutionising understanding of ice sheet behaviour and contributions to sea level rise (IMBIE Team, 2018; Sandberg Sørensen et al., 2018; Nilsson et al., 2022; Otosaka et al., 2023). The complex topography of the ice margins, however, presents several deep-seated challenges for conventional processing; namely, difficulty in tracking the ice surface, and fundamental ambiguity in determining the origin of surface reflections. While both challenges limit measurement capabilities, the later represents a fundamental limitation inherent in non-interferometric systems, with traditional approaches relying on dimensionality reduction to single-point estimates, discarding much of the waveform's information content (Brooks et al., 1975; Ridley et al., 1988; Davis et al., 1993; Bamber, 1994). While state-of-the-art methods have

refined these approaches, contemporary processing continues to rely on reduced-dimensionality representations, limiting the ability to fully exploit the information encoded within waveforms - a limitation quantified in Chapter 5 and directly addressed by the distributional framework developed and assessed in Chapters 6 and 7.

This thesis develops methods to better understand and overcome these deficiencies through novel deep learning methodologies and high-resolution Digital Elevation Models. To achieve this, four interconnected objectives were pursued: (1) systematic assessment of current operational systems, (2) the introduction of new methods for topographic characterisation, (3) the design of probabilistic frameworks that quantify waveform ambiguity as an information source, and (4) demonstration of practical utility through application to untrained regions and comparisons to establish approaches.

This thesis demonstrates that deep learning can successfully quantify surface ambiguity in non-interferometric instruments across polar ice sheets, significantly increasing measurement density and offering improved capabilities for long-term monitoring of ice sheet surface elevation. These advances address longstanding limitations in SAR altimetry, impacting past, current, and future missions and, ultimately, our understanding of ice sheet contributions to sea level rise.

The specific objectives of this thesis, as outlined in Chapter 4, were:

1. **Assess current operational capabilities**: Evaluate Sentinel-3 SAR altimeter performance over the Antarctic Ice Sheet using high-resolution topographic datasets, developing new methodologies for understanding instrument limitations and topographic controls on altimetry processing (Chapter 5).

2. **Develop probabilistic deep learning framework**: Create and validate a novel approach for extracting distributional elevation information directly from non-interferometric SAR radar waveforms, explicitly modelling waveform ambiguity (Chapter 6).

3. **Demonstrate transferability and utility**: Apply the deep learning framework to regions entirely absent from training data and compare performance against established altimetric approaches to demonstrate practical utility for operational ice sheet elevation change monitoring (Chapter 7).

8.2 Summary of Key Findings

This thesis directly addresses the aims outlined above through three interconnected studies that each contribute to enhanced understanding of satellite radar altimetry capabilities and limitations over polar ice sheets.

8.2.1 Operational Performance Assessment and Topographic Characterisation

Chapter 5 provided the first comprehensive, ice-sheet-scale assessment of Sentinel-3 SAR altimeter performance over Antarctica using high-resolution topographic datasets. This work built upon previous studies that accessed accuracy and precision of Sentinel-3 elevation measurements (McMillan et al., 2019; McMillan et al., 2021)), established new benchmarks for understanding altimeter capabilities, and introduced methodological approaches with implications extending beyond immediate technical assessment.

The study developed new continent-wide surface slope and roughness datasets by applying Singular Value Decomposition (SVD) to the Reference Elevation Model of Antarctica (REMA). This approach algorithmically separates slope and roughness components, offering improvements over previous methods that often conflated these distinct topographic properties (Yi et al., 2005; Riley et al., 1999; Wilson et al., 2007)). This conflation is problematic because slope and roughness represent distinct topographic characteristics that influence different aspects of ice sheet behaviour, making their separation essential for accurate interpretation. The analysis revealed that 76.9% of the Antarctic Ice Sheet exhibits slopes below 0.5°, while 62.4% has

surface roughness below 0.5 m, with a strong positive correlation between slope and roughness (Pearson correlation coefficient of 0.808). Application to the Greenland Ice Sheet in Chapter 7, mirroring the Antarctica analysis, revealed that Greenland exhibits more complex surface conditions: over twice the median surface slopes (0.434° vs 0.192°) and nearly three times the median roughness (0.855 m vs 0.297 m) compared to Antarctica.

Over Antarctica, systematic evaluation of the fundamental assumption in conventional altimetric processing that the Point of Closest Approach (POCA) corresponds to the leading edge of the waveform revealed that 94.1% of Sentinel-3 acquisitions successfully capture the POCA within the range window. This represents a 4.6% improvement over the previous non-thematic product. However, the remaining 5.9% of cases - occurring predominantly in areas of high topographic complexity - demonstrate a clear breakdown of this core assumption.

Further analysis demonstrated that, 57.4% of the ice sheet exhibits greater topographic variance within the beam footprint than can be fully captured by the current 60 m range window, demonstrating limitations in this instrumental choice. Doubling the range window size from 60 m to 120 m could increase mean topographic capture from 90.6% to 96.9%.

It was also demonstrated that optimal range window placement achieves a median of 89.2% of total possible topographic capture, declining systematically with increasing slope and roughness. This demonstrates that even under conditions where full capture is theoretically possible, significant improvements can still be achieved.

Chapter 5 also quantified how surface topography affects along-track waveform correlation over Antarctica, revealing that inter-record correlation remains high (median 0.944) but decorrelates more rapidly over complex terrain. Surface slope and roughness explain 38.2% and 22.2% of ice-sheet-wide variance in 50-record correlation decay, respectively, establishing quantifiable relationships between topography and radar scattering behaviour.

These findings support the range window size chosen for the upcoming CRISTAL

mission (256 m range window with interferometric capability), validating design choices for next-generation altimetry missions (Kern et al., 2020). Importantly, the implications of this analysis extend beyond Antarctica: given Greenland's steeper surface slopes and higher roughness, the results presented in Chapter 5 are likely to prove even more critical for interpreting altimeter performance in this region.

8.2.2 Probabilistic Deep Learning Framework Development

Chapter 6 introduced a probabilistic deep learning approach that reframes waveform ambiguity as an information source rather than a processing limitation. This represents the first demonstration that meaningful distributional elevation information can be extracted from across the entire swath using solely power waveforms alone without reliance on interferometric phase data as shown in previous studies (Gray et al., 2013; Gourmelen et al., 2018). The framework employs an ensemble of 16 ResNet-RS models trained with Pinball loss for quantile regression, predicting 5^{th}, 50^{th}, and 95^{th} elevation quantiles across 150 points spanning a 15 km across-track swath. This approach provides an alternative to conventional single-point processing while explicitly quantifying both aleatoric uncertainty (arising from fundamental ambiguity in possible surface configurations) and epistemic uncertainty (reflecting confidence in model predictions).

To ensure high-quality training data, the study developed a new, simplified waveform simulator using ray-casting techniques, which proved robust to highly irregular surface geometry whilst being computationally efficient. This simulator enabled filtering of mismatched waveform-topography pairs using Wasserstein distance thresholds to maintain geometric consistency in the training data.

Evaluation of the deep learning framework against an independent Antarctic test set (2023, ice-sheet wide) demonstrated robust topographic profile recovery, with the 50^{th} quantile showing a median elevation offset of only -2.97 $\pm$ 2.96 m relative to REMA across the entire ice sheet. Prediction Interval Coverage Probability (PICP) analysis revealed systematic calibration patterns: the framework captured 79.6% of

REMA points within the 5^{th}-95^{th} quantile range (target: 90%) when taken ice-sheet wide, with under-coverage increasing substantially in complex terrain. Notably, the 95^{th} quantile maintained good robustness across all surface conditions, reflecting fundamental radar physics (Bamber, 1994; Scott et al., 1994), where the 95^{th} quantile is anchored to leading-edge returns that provide the clearest elevation signal with minimal ambiguity, while lower quantiles capture progressively more ambiguous later returns from multiple scattering surfaces.

Aleatoric uncertainty, captured through quantile spread, reflects the fundamental physical ambiguity in possible surface configurations and varies systematically with terrain complexity. Distributional widths (5^{th}-95^{th} quantile coverage) increase from 23.5 m at 0° slope to 60.8 m at 1° slope, and from 21.9 m to 59.8 m as roughness increases from 0 m to 100 m, reflecting the more complex waveforms and multiple scattering sources in topographically challenging terrain. Epistemic uncertainty also increased with topographic complexity and showed systematic patterns with the 95^{th} quantile exhibiting the lowest uncertainty (median 1.48 m), followed by the 50^{th} (2.17 m) and 5^{th} quantiles (3.19 m). This hierarchy again reflects the upper quantiles physical anchoring in leading-edge returns that provide the clearest elevation signal.

8.2.3 Cross-Domain Validation and Transferability

Chapter 7 established the transferability and practical utility of the deep learning framework through application to Pine Island Glacier and the Greenland Ice Sheet - regions entirely absent from Antarctic training data. This validation demonstrated transferability while revealing both capabilities and constraints for operational deployment.

The framework showed consistent calibration behaviour across all three regions (Antarctica, Pine Island Glacier, Greenland), providing evidence that the ensemble learned transferable physical relationships between radar scattering and surface geometry rather than region-specific characteristics. PICP patterns replicated systematically when stratified by topographic complexity across all domains, with

validation conducted against ICESat-2 observations.

Established glaciological signals were successfully reproduced, capturing 2-3 m yr^{-1} thinning at Pine Island Glacier (Nilsson et al., 2022; Reed et al., 2024; Wingham et al., 2009) and major outlet glacier losses across Greenland (Simonsen et al., 2021; Nilsson et al., 2024). Validation against three independent datasets - ICESat-2, Cryo-TEMPO Level-2+ POCA, and Cryo-TEMPO EOLIS, demonstrated good agreement in both spatial elevation change (dh/dt) patterns and region-wide temporal elevation change (dh) time series. The fact the framework achieved comparable spatiotemporal performance to interferometric swath processing, despite using only power waveforms, is particularly notable. The ability to detect sub-meter annual elevation changes from probability distributions spanning tens of meters validated that waveform ambiguity constitutes exploitable information rather than a processing obstacle. However, systematic elevation gain noise appeared in Greenland's most complex coastal regions, correlating clearly with topographic complexity. Looking forward, it will be important to optimise performance in these regions, since they are experiencing the greatest mass loss (IMBIE Team, 2018), though explicitly training future operational models on data from this region, which is presently out-of-distribution, could reduce such artifacts. Section 8.4 outlines potential avenues for future improvements in this regard.

8.3 Synthesis of Key Contributions

This thesis advances satellite radar altimetry by better characterising key limitations and enhancing both processing methodology and topographic characterisation. Several key themes emerge that collectively represent progress in how non-interferometric altimetry processing is approached over polar ice sheets.

8.3.1 Reconceptualising Measurement Ambiguity as Distributional Information

Traditional altimetry processing treats waveform ambiguity as an obstacle requiring elimination through dimensionality reduction to single-point estimates. This thesis demonstrates that this ambiguity can be quantified, revealing valuable distributional information in across-track surface configurations. The key insight is that power waveforms contain sufficient information to predict full elevation distributions rather than singular estimates - fundamentally challenging three decades of deterministic processing philosophy.

The framework, developed in Chapter 6 and further explored in Chapter 7, extracts 150 elevation estimates per waveform where conventional methods yield one, revealing that information content is systematically discarded in standard processing. Furthermore, the systematic hierarchy in quantile performance, with the 95^{th} percentile maintaining stability while the 5^{th} percentile shows greatest uncertainty, validates the fundamental assumptions underlying conventional POCA approaches, while demonstrating exploitable structure in waveform ambiguity.

This reconceptualisation proves particularly powerful for ice sheet monitoring, where sub-footprint surface variability becomes accessible to non-interferometric altimetric observation. Previously undetectable features such as crevasse field morphology, sastrugi patterns, and accumulation heterogeneity could potentially be characterised through distributional analysis. This capability opens pathways for better understanding ice sheet stability, as surface texture changes often precede larger-scale dynamical responses, while offering potential new constraints on firn processes and surface mass balance that could bridge observational gaps between point measurements and regional assessments.

8.3.2 Assessing Conventional Assumptions and Quantifying Performance

The comprehensive evaluation of Sentinel-3 SAR performance over Antarctica (Chapter 5) reveals that fundamental assumptions underlying conventional altimetry processing fail systematically in topographically complex regions. The analysis demonstrates that the core assumption - that the POCA lies within both the illuminated footprint and range window - breaks down predictably as surface complexity increases, with POCA capture success degrading from near 100% at low slopes to ∼30% at 5°. This systematic failure exposes critical limitations in current processing paradigms that assume universal applicability of simplified geometric models. The quantification of these breakdown thresholds provides empirical foundations for understanding where conventional processing approaches become unreliable.

The assessment of topographic capture capabilities reveals that, for Sentinel-3, 57.4% of observations cannot achieve complete surface representation within current range window constraints. This analysis demonstrates the importance of range window size.

Furthermore, even compared to the theoretical optimum, the current window placement captures a median of only 89.2% of what could maximally be recorded within the 60 m window, with this proportion declining further with increasing slope and roughness.

8.3.3 Validating Slope and Roughness as Global Performance Drivers

The consistent topographic-performance relationships observed across Antarctica, Pine Island, and Greenland establish surface geometry as a universal control on altimetric measurement capability, extending beyond regional differences in ice dynamics, climate conditions, and processing methodologies. From the analyses

in Chapter 5 through the probabilistic deep learning framework in Chapter 6 to the cross-domain validation in Chapter 7, all approaches demonstrate identical dependencies on slope and roughness parameters. This universality is significant because it reveals that the physical interaction between radar signals and surface geometry governs altimetric performance independent of glaciological context, establishing a transferable framework for predicting measurement limitations and optimising processing strategies across all polar ice sheet environments.

8.3.4 Turning Methodology into Monitoring Capability

The cross-domain validation in Chapter 7 demonstrates that Antarctic-derived models can predict behaviour in different polar environments. Compared to Antarctica, the Greenland Ice Sheet experiences a warmer, wetter climate and elevation change signals are dominated by surface melt and dynamic thinning from outlet glaciers (Broeke et al., 2017; Slater et al., 2021; Otosaka et al., 2023). The Antarctic Ice Sheet, by contrast, is much colder and drier, with negligible surface melt; elevation change signals are driven primarily by dynamic thinning (IMBIE Team, 2018; Shepherd et al., 2019; Rignot et al., 2019; Otosaka et al., 2023). The replication of calibration patterns and degradation thresholds across diverse settings validates the physical foundation of the deep learning framework developed in Chapter 6, providing confidence for broader operational deployment and historical data exploitation. These generalisability capabilities are especially important for downstream applications mapping surface elevation change where topography differs from that seen during training. This universality also suggests well-characterised regions can inform processing approaches for less well-studied areas, dramatically expanding the effective scope of altimetric observations. Performance across the Greenland Ice Sheet is especially encouraging given the understanding that future models could explicitly train on this region, likely yielding further improvements.

Alongside spatial transferability, comparative performance with ICESat-2 demonstrates the framework's precision, achieving median elevation change differences of

-0.02 to -0.03 m yr^{-1} while successfully capturing known glaciological signals across all regions. This ability to detect sub-meter annual changes validates the approach's capacity to resolve subtle waveform variations into realistic elevation signals while revealing previously inaccessible distributional information in existing observations.

Although the framework was developed using CryoSat-2 SARIn data - selected for its substantially larger range window (∼240 m versus ∼60 m for Sentinel-3) and longer observational record - the underlying methodology is, in principle, directly transferable to Sentinel-3 SAR waveforms and, more generally, to SAR altimetry missions operating with comparable waveform geometry and sampling characteristics. The smaller Sentinel-3 range window would limit the topographic surface captured within each waveform (Chapter 5), reducing the across-track swath that could be recovered, but the distributional processing approach itself remains applicable. Extension to Sentinel-3 therefore represents a clear operational priority, particularly given the mission's ongoing and planned continuity through Sentinel-3C and 3D.

The framework also provides a foundation for operational implementation that could extend elevation change records to the early 1990s through application to ERS-1, ERS-2, and Envisat observations, though adaptation would be required for Low Resolution Mode (LRM) rather than SAR processing. This temporal extension, combined with enhanced spatial coverage from distributional processing, offers potential for improved understanding of processes that drive ice sheet mass loss throughout the satellite era. Advancing understanding of these processes is critical for improving ice-sheet models , which in turn strengthens predictions of future mass loss and sea-level contributions (Levermann et al., 2020; Bamber et al., 2022).

8.4 Limitations and Future Directions

While this thesis provides advancements in satellite radar altimetry capabilities, several limitations must be acknowledged. These limitations stem from data constraints, methodological challenges, and broader issues related to operational deployment.

8.4.1 Training Data Dependencies and Reference Dataset Quality

The deep learning framework's performance remains constrained by the quality of training data derived from REMA. While REMA represents an accurate dataset, the primary limitation arises from temporal mismatch between the REMA mosaic (spanning multiple years) and individual altimetric observations, introducing substantial uncertainties into the training process. Nonetheless, REMA currently represents the most comprehensive and reliable dataset available for this purpose.

The requirement for waveform-topography matching through radar simulation resulted in substantial data filtering, with 88.4% of potential training pairs discarded. This massive filtering suggests underlying noise in the dataset and may create systematic biases toward particular topographic regimes. The simplified waveform simulator, while computationally efficient, may itself introduce biases toward certain surface types.

Enhanced waveform simulators capable of generating precisely matched surface-waveform pairs could provide cleaner training datasets. This approach could involve taking REMA profiles representing known topographic conditions and using improved simulators to produce high-quality surface-waveform pairs. Additionally, integrating complementary datasets could provide additional constraints to resolve predicted surface ambiguity during post-processing, thus enhancing operational capabilities.

8.4.2 Model Calibration and Complex Terrain Performance

Systematic calibration issues persist, particularly for lower quantiles in topographically complex regions. The tendency toward over-constrained distribution predictions affects all quantiles but is most pronounced for the 5^{th} and 50^{th} percentiles, with under-coverage worsening systematically as terrain complexity increases. Performance degradation in highly complex topography represents a significant challenge, with the deep learning framework often performing worse than conventional approaches in these conditions. Notably, the systematic appearance of spurious elevation gain noise in Greenland's most complex coastal regions highlights areas requires further investigation. Though this is likely improved by training on data from this region explicitly.

Improved calibration through enhanced loss functions, potentially incorporating adversarial training approaches or physics-informed constraints, could provide more robust predictions. Investigation of alternative neural network architectures, including transformer-based approaches or hybrid frameworks combining physics-based constraints with data-driven learning, may also offer pathways to improved performance in complex terrain. Insights gained from these developments could also support the extension of processing to other glaciated regions, such as ice caps and mountain glaciers, and more broadly to other geophysical environments such as sea ice.

8.4.3 Computational and Operational Deployment Challenges

The ensemble approach requires running 16 independent models at inference, increasing computational demands compared to conventional processing methods. While the deep learning framework utilises GPU processing for efficient inference, comprehensive benchmarking against conventional processing speeds remains necessary for operational assessment. The requirement for a priori slope information to resolve profile orientation ambiguity introduces additional data dependencies, though this constraint is shared with current slope correction approaches. As

demonstrated in Chapter 7, this post-processing correction can be accomplished using much lower resolution DEMs (e.g., 1 km resolution).

Model compression techniques and architectural optimisation could reduce computational requirements while maintaining distributional prediction capabilities. Additionally, further explorations into the minimum number of ensemble members required for optimal epistemic uncertainty quantification could provide improvements. Development of real-time processing capabilities would enable operational deployment for monitoring rapid ice sheet changes.

8.4.4 Integration with Existing Processing Chains

The probabilistic framework represents a departure from conventional altimetry processing expectations. Many downstream applications expect single elevation estimates rather than distributional predictions, requiring development of appropriate synthesis approaches that extract relevant information while maintaining enhanced uncertainty characterisation.

Standardisation of uncertainty reporting and distribution characterisation within the altimetry community requires development of guidelines and best practices for extracting application-specific information from distributional predictions. Integration with existing processing chains would require careful consideration of backward compatibility.

An additional avenue for future work is the fusion of complementary altimetric datasets - for example, combining CryoSat-2 SARIn swath products with Sentinel-3 SAR coverage - to exploit the respective strengths of each mission's spatial coverage, repeat period, and measurement geometry. Such multi-mission approaches could further enhance the density and accuracy of elevation change estimates, particularly over the ice sheet margins where measurement gaps remain most consequential.

8.4.5 Physical Interpretation and Model Explainability

While the framework demonstrates consistent physical behaviour and transferability across diverse polar environments, understanding why specific waveform features lead to particular topographic predictions remains limited. The relationship between learned waveform features and physical scattering processes, while validated through transferability and calibration consistency, lacks the explicit interpretability that characterises conventional processing approaches.

Investigation of interpretable machine learning approaches that provide insight into learned waveform-topography relationships could enhance operational confidence and facilitate model refinement. Integration of physics-informed constraints, or hybrid approaches that combine interpretable physical models with data-driven components, could provide better balance between performance and explainability.

8.5 Concluding Remarks

This thesis advances satellite radar altimetry by characterising key limitations in current systems, developing innovative processing methodologies, and improving topographic characterisation. The systematic progression from comprehensive performance assessment through algorithmic development to cross-domain validation establishes foundations for enhanced ice sheet monitoring.

The demonstration that waveform ambiguity can be quantified and exploited as a valuable information source rather than eliminated as a processing limitation represents the central conceptual contribution of this work. By reframing the fundamental challenge of non-interferometric altimetry from obstacle to opportunity, this research opens pathways for enhanced information extraction from both current and historical radar altimetry observations. The ability to provide up to 150 elevation estimates across the swath where traditional methods return only one, while maintaining comprehensive distributional uncertainty characterisation, demonstrates untapped potential in existing altimetric observations.

The competitive performance achieved by the deep learning framework, using only power waveforms, relative to swath interferometric processing validates the central premise that conventional altimetry processing discards valuable information that can be recovered through sophisticated machine learning approaches. The ability to detect sub-meter annual elevation changes from probability distributions spanning tens of meters, while maintaining physically consistent uncertainty estimates, validates the presence of exploitable signal in observations traditionally considered too ambiguous for reliable processing.

The comprehensive assessment of Sentinel-3 performance in Chapter 5 provided operational insights through novel topographic characterisation methods and systematic evaluation of instrument limitations. The SVD-based approach for separating slope and roughness components offers methodological advances applicable beyond altimetry applications, while the quantification of performance with respect to range window placement and high-resolution topographic datasets provides concrete guidance for current and future mission optimisation.

The transferability demonstrated across Antarctica, Pine Island Glacier, and Greenland provides evidence that the deep learning framework learned fundamental physical relationships governing radar-surface interactions rather than region-specific characteristics. This transferability, combined with the systematic replication of calibration patterns and topographic dependencies across diverse polar environments, establishes the approach as a pathway for enhanced ice sheet monitoring with broad applicability. Performance here is especially encouraging given the understanding that future models could explicitly train on these withheld regions, likely yielding further improvements. Given the global application of satellite altimetry, the methodological contributions presented here extend beyond glaciological applications to potentially inform altimetric observations across diverse terrestrial environments.

Although performance limitations persist, particularly in the most topographically complex regions, the demonstrated capabilities represent meaningful advances

using power waveforms alone. The enhanced capabilities demonstrated for current missions like Sentinel-3, combined with the guidance provided for future missions including CRISTAL and Sentinel-3 Next Generation Topography, position this work to contribute to the next generation of ice sheet monitoring capabilities.

The operational implications are promising. The potential to apply these approaches to historical missions such as ERS-1, ERS-2, and Envisat could extend elevation change records back to the early 1990s, though adaptation for Low Resolution Mode processing will be required. The enhanced measurement density and distributional uncertainty quantification demonstrated here contribute directly to addressing monitoring challenges while providing tools for more effective utilisation of both current and future altimetry missions.

As ice sheet mass loss continues to accelerate under contemporary climate warming, the need for robust, scalable monitoring approaches becomes increasingly critical for constraining sea level projections. The research presented here establishes probabilistic distributional altimetry as a viable approach for polar science applications. By demonstrating that ambiguity can be transformed from limitation to information source, this work provides methodological foundations that can unlock substantially more scientific value from satellite radar altimetry observations. Although operational deployment will require addressing the limitations identified, the fundamental advance in reframing waveform ambiguity opens new pathways for extracting maximum information content from observations of Earth's changing ice sheets.

In conclusion, this thesis demonstrates that the intersection of advanced machine learning with non-interferometric SAR radar altimetry can overcome longstanding limitations and unlock new capabilities for monitoring polar ice sheet evolution. Through advancements in characterising altimeter performance, improved topographic representation, and development of novel probabilistic frameworks, the methods and insights presented here provide a foundation for improved altimetry processing approaches, offering enhanced tools for understanding ice sheet dynamics

and improving projections of future sea level rise in support of climate science and policy applications.

References

Abdalla, Saleh, Abdolnabi Abdeh Kolahchi, Michaël Ablain, Susheel Adusumilli, Suchandra Aich Bhowmick, Eva Alou-Font, Laiba Amarouche, Ole Baltazar Andersen, Helena Antich, Lotfi Aouf, Brian Arbic, Thomas Armitage, Sabine Arnault, Camila Artana, Giuseppe Aulicino, Nadia Ayoub, Sergei Badulin, Steven Baker, Chris Banks, Lifeng Bao, Silvia Barbetta, Bàrbara Barceló-Llull, François Barlier, Sujit Basu, Peter Bauer-Gottwein, Matthias Becker, Brian Beckley, Nicole Bellefond, Tatyana Belonenko, Mounir Benkiran, Touati Benkouider, Ralf Bennartz, Jérôme Benveniste, Nicolas Bercher, Muriel Berge-Nguyen, Joao Bettencourt, Fabien Blarel, Alejandro Blazquez, Denis Blumstein, Pascal Bonnefond, Franck Borde, Jérôme Bouffard, François Boy, Jean-Paul Boy, Cédric Brachet, Pierre Brasseur, Alexander Braun, Luca Brocca, David Brockley, Laurent Brodeau, Shannon Brown, Sean Bruinsma, Anna Bulczak, Sammie Buzzard, Madeleine Cahill, Stéphane Calmant, Michel Calzas, Stefania Camici, Mathilde Cancet, Hugues Capdeville, Claudia Cristina Carabajal, Loren Carrere, Anny Cazenave, Eric P. Chassignet, Prakash Chauhan, Selma Cherchali, Teresa Chereskin, Cecile Cheymol, Daniele Ciani, Paolo Cipollini, Francesca Cirillo, Emmanuel Cosme, Steve Coss, Yuri Cotroneo, David Cotton, Alexandre Couhert, Sophie Coutin-Faye, Jean-François Crétaux, Frederic Cyr, Francesco d'Ovidio, José Darrozes, Cedric David, Nadim Dayoub, Danielle De Staerke, Xiaoli Deng, Shailen Desai, Jean-Damien Desjonqueres, Denise Dettmering, Alessandro Di Bella, Lara Díaz-Barroso, Gerald Dibarboure, Habib Boubacar Dieng, Salvatore Dinardo, Henryk Dobslaw, Guillaume Dodet, Andrea Doglioli, Alessio

Domeneghetti, David Donahue, Shenfu Dong, Craig Donlon, Joël Dorandeu, Christine Drezen, Mark Drinkwater, Yves Du Penhoat, Brian Dushaw, Alejandro Egido, Svetlana Erofeeva, Philippe Escudier, Saskia Esselborn, Pierre Exertier, Ronan Fablet, Cédric Falco, Sinead Louise Farrell, Yannice Faugere, Pierre Femenias, Luciana Fenoglio, Joana Fernandes, Juan Gabriel Fernández, Pascale Ferrage, Ramiro Ferrari, Lionel Fichen, Paolo Filippucci, Stylianos Flampouris, Sara Fleury, Marco Fornari, Rene Forsberg, Frédéric Frappart, Marie-laure Frery, Pablo Garcia, Albert Garcia-Mondejar, Julia Gaudelli, Lucile Gaultier, Augusto Getirana, Ferran Gibert, Artur Gil, Lin Gilbert, Sarah Gille, Luisella Giulicchi, Jesús Gómez-Enri, Laura Gómez-Navarro, Christine Gommenginger, Lionel Gourdeau, David Griffin, Andreas Groh, Alexandre Guerin, Raul Guerrero, Thierry Guinle, Praveen Gupta, Benjamin D. Gutknecht, Mathieu Hamon, Guoqi Han, Danièle Hauser, Veit Helm, Stefan Hendricks, Fabrice Hernandez, Anna Hogg, Martin Horwath, Martina Idžanović, Peter Janssen, Eric Jeansou, Yongjun Jia, Yuanyuan Jia, Liguang Jiang, Johnny A. Johannessen, Masafumi Kamachi, Svetlana Karimova, Kathryn Kelly, Sung Yong Kim, Robert King, Cecile M.M. Kittel, Patrice Klein, Anna Klos, Per Knudsen, Rolf Koenig, Andrey Kostianoy, Alexei Kouraev, Raj Kumar, Sylvie Labroue, Loreley Selene Lago, Juliette Lambin, Léa Lasson, Olivier Laurain, Rémi Laxenaire, Clara Lázaro, Sophie Le Gac, Julien Le Sommer, Pierre-Yves Le Traon, Sergey Lebedev, Fabien Léger, Benoıˆt Legresy, Frank Lemoine, Luc Lenain, Eric Leuliette, Marina Levy, John Lillibridge, Jianqiang Liu, William Llovel, Florent Lyard, Claire Macintosh, Eduard Makhoul Varona, Cécile Manfredi, Frédéric Marin, Evan Mason, Christian Massari, Constantin Mavrocordatos, Nikolai Maximenko, Malcolm McMillan, Thierry Medina, Angelique Melet, Marco Meloni, Stelios Mertikas, Sammy Metref, Benoit Meyssignac, Jean-François Minster, Thomas Moreau, Daniel Moreira, Yves Morel, Rosemary Morrow, John Moyard, Sandrine Mulet, Marc Naeije, Robert Steven Nerem, Hans Ngodock, Karina Nielsen, Jan Even Øie Nilsen, Fernando Niño, Carolina Nogueira Loddo, Camille Noûs,

Estelle Obligis, Inès Otosaka, Michiel Otten, Berguzar Oztunali Ozbahceci, Roshin P. Raj, Rodrigo Paiva, Guillermina Paniagua, Fernando Paolo, Adrien Paris, Ananda Pascual, Marcello Passaro, Stephan Paul, Tamlin Pavelsky, Christopher Pearson, Thierry Penduff, Fukai Peng, Felix Perosanz, Nicolas Picot, Fanny Piras, Valerio Poggiali, Étienne Poirier, Sonia Ponce de León, Sergey Prants, Catherine Prigent, Christine Provost, M-Isabelle Pujol, Bo Qiu, Yves Quilfen, Ali Rami, R. Keith Raney, Matthias Raynal, Elisabeth Remy, Frédérique Rémy, Marco Restano, Annie Richardson, Donald Richardson, Robert Ricker, Martina Ricko, Eero Rinne, Stine Kildegaard Rose, Vinca Rosmorduc, Sergei Rudenko, Simón Ruiz, Barbara J. Ryan, Corinne Salaün, Antonio Sanchez-Roman, Louise Sandberg Sørensen, David Sandwell, Martin Saraceno, Michele Scagliola, Philippe Schaeffer, Martin G. Scharffenberg, Remko Scharroo, Andreas Schiller, Raphael Schneider, Christian Schwatke, Andrea Scozzari, Enrico Ser-giacomi, Frederique Seyler, Rashmi Shah, Rashmi Sharma, Andrew Shaw, Andrew Shepherd, Jay Shriver, C.K. Shum, Wim Simons, Sebatian B. Simonsen, Thomas Slater, Walter Smith, Saulo Soares, Mikhail Sokolovskiy, Laurent Soudarin, Ciprian Spatar, Sabrina Speich, Margaret Srinivasan, Meric Srokosz, Emil Stanev, Joanna Staneva, Nathalie Steunou, Julienne Stroeve, Bob Su, Yohanes Budi Sulistioadi, Debadatta Swain, Annick Sylvestre-baron, Nicolas Taburet, Rémi Tailleux, Katsumi Takayama, Byron Tapley, Angelica Tarpanelli, Gilles Tavernier, Laurent Testut, Praveen K. Thakur, Pierre Thibaut, LuAnne Thompson, Joaquín Tintoré, Céline Tison, Cédric Tourain, Jean Tournadre, Bill Townsend, Ngan Tran, Sébastien Trilles, Michel Tsamados, Kuo-Hsin Tseng, Clément Ubelmann, Bernd Uebbing, Oscar Vergara, Jacques Verron, Telmo Vieira, Stefano Vignudelli, Nadya Vinogradova Shiffer, Pieter Visser, Frederic Vivier, Denis Volkov, Karina von Schuckmann, Valerii Vuglinskii, Pierrik Vuilleumier, Blake Walter, Jida Wang, Chao Wang, Christopher Watson, John Wilkin, Josh Willis, Hilary Wilson, Philip Woodworth, Kehan Yang, Fangfang Yao, Raymond Zaharia, Elena Zakharova, Edward D. Zaron, Yongsheng Zhang,

Zhongxiang Zhao, Vadim Zinchenko, and Victor Zlotnicki (2021). "Altimetry for the future: Building on 25 years of progress". In: *Advances in Space Research* 68.2, pp. 319–363.

Abdel-Hamid, Ossama, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu (Oct. 2014). "Convolutional neural networks for speech recognition". In: *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 22.10, pp. 1533–1545.

Abdelsattar, Montaser, Ahmed AbdelMoety, and Ahmed Emad-Eldeen (July 2025). "ResNet-based image processing approach for precise detection of cracks in photovoltaic panels". In: *Scientific Reports* 15.1, p. 24356.

Adusumilli, Susheel, Helen Amanda Fricker, Brooke Medley, Laurie Padman, and Matthew R. Siegfried (Sept. 2020). "Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves". In: *Nature Geoscience* 13.9, pp. 616–620.

Adusumilli, Susheel, Helen Amanda Fricker, Matthew R. Siegfried, Laurie Padman, Fernando S. Paolo, and Stefan R. M. Ligtenberg (2018). "Variable Basal Melt Rates of Antarctic Peninsula Ice Shelves, 1994–2016". In: *Geophysical Research Letters* 45.9, pp. 4086–4095.

Agbehadji, Israel Edem and Ibidun Christiana Obagbuwa (2024). "Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction". In: *Atmosphere* 15.11.

Amini, Alexander, Wilko Schwarting, Ava Soleimany, and Daniela Rus (2020). *Deep Evidential Regression*.

Amini, Alexander, Ava Soleimany, Sertac Karaman, and Daniela Rus (2019). *Spatial Uncertainty Sampling for End-to-End Control*.

Andernach, M., M.-L. Kapsch, and U. Mikolajewicz (2025). "Impact of Greenland Ice Sheet disintegration on atmosphere and ocean disentangled". In: *Earth System Dynamics* 16.2, pp. 451–474.

Andersen, Ole Baltazar (June 2025). *Cryo-TEMPO Product Handbook*. Tech. rep. LU_ESA_Cryo-TEMPO_PH. Version 4.1. Checked by Malcolm McMillan. ESA.

Andersson, Tom R., J. Scott Hosking, María Pérez-Ortiz, Brooks Paige, Andrew Elliott, Christopher Russell, Stephen Law, Daniel C. Jones, Jeremy Wilkinson, Tony Phillips, James Byrne, Steffen Tietsche, Beena Balan Sarojini, Edward Blanchard-Wrigglesworth, Yevgeny Aksenov, Richard Downie, and Emily Shuckburgh (Aug. 2021). "Seasonal Arctic sea ice forecasting with probabilistic deep learning". In: *Nature Communications* 12.1, p. 5124.

Aublanc, J., F. Boy, F. Borde, and P. Féménias (2025a). "A facet-based numerical model to retrieve ice sheet topography from Sentinel-3 altimetry". In: *The Cryosphere* 19.5, pp. 1937–1954.

Aublanc, J., P. Féménias, F. Catapano, and A. Chamayou (2024). *Sentinel-3 SRAL/MWR Land User Handbook*. Technical Report S3MPC-STM_RP_0038. Issue 1.2 – 08/01/2024. European Space Agency (ESA): S3 Mission Performance Centre (S3MPC).

Aublanc, J., T. Moreau, P. Thibaut, F. Boy, F. Rémy, and N. Picot (2018). "Evaluation of SAR altimetry over the antarctic ice sheet from CryoSat-2 acquisitions". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1307–1323.

Aublanc, Jérémie, Julien Renou, Fanny Piras, Karina Nielsen, Stine K. Rose, Sebastian B. Simonsen, Sara Fleury, Stefan Hendricks, Nicolas Taburet, Giovanni D'Apice, Anouk Chamayou, Pierre Féménias, Filomena Catapano, and Marco Restano (2025b). "Sentinel-3 Altimetry Thematic Products for Hydrology, Sea Ice and Land Ice". In: *Scientific Data* 12.1, p. 714.

Aublanc, Jérémie, Pierre Thibaut, Amandine Guillot, François Boy, and Nicolas Picot (2021). "Ice Sheet Topography from a New CryoSat-2 SARIn Processing Chain, and Assessment by Comparison to ICESat-2 over Antarctica". In: *Remote Sensing* 13.22.

AVISO (2019). *Delay-Doppler / SAR altimetry*. `https://www.aviso.altimetry.fr/en/techniques/altimetry/principle/delay-doppler-/-sar-altimetry.html`. Accessed: 2025-09-26. AVISO - Archiving, Validation and Interpretation of Satellite Oceanographic data.

Ball, James G. C., Katerina Petrova, David A. Coomes, and Seth Flaxman (2022). "Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation". In: *Methods in Ecology and Evolution* 13.11, pp. 2622–2634.

Bamber, J. L. (1994). "Ice sheet altimeter processing scheme". In: *International Journal of Remote Sensing* 15.4, pp. 925–938.

Bamber, J. L., M. Oppenheimer, R. E. Kopp, W. P. Aspinall, and Roger M. Cooke (2022). "Ice Sheet and Climate Processes Driving the Uncertainty in Projections of Future Sea Level Rise: Findings From a Structured Expert Judgement Approach". In: *Earth's Future* 10.10. e2022EF002772 2022EF002772, e2022EF002772.

Bello, Irwan, William Fedus, Xianzhi Du, Ekin D. Cubuk, Aravind Srinivas, Tsung-Yi Lin, Jonathon Shlens, and Barret Zoph (2021). *Revisiting ResNets: Improved Training and Scaling Strategies*.

Bengio, Y., P. Simard, and P. Frasconi (1994). "Learning long-term dependencies with gradient descent is difficult". In: *IEEE Transactions on Neural Networks* 5.2, pp. 157–166.

Bergstra, James and Yoshua Bengio (Feb. 2012). "Random search for hyper-parameter optimization". In: *J. Mach. Learn. Res.* 13.null, pp. 281–305.

Birkett, C. M. and B. Beckley (2010). "Investigating the Performance of the Jason-2/OSTM Radar Altimeter over Lakes and Reservoirs". In: *Marine Geodesy* 33.sup1, pp. 204–238.

Bishop, Christopher M (Nov. 1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

Blarel, Fabien, Frédéric Frappart, Benoıˆt Legrésy, Denis Blumstein, Frédérique Rémy, Christophe Fatras, Eric Mougin, Fabrice Papa, Catherine Prigent, Fernando Niño, et al. (2015). "Altimetry backscattering signatures at Ku and S bands over land and ice sheets". In: *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII*. Vol. 9637. SPIE, pp. 315–324.

Boehme, Lars and Isabella Rosso (2021). "Classifying Oceanographic Structures in the Amundsen Sea, Antarctica". In: *Geophysical Research Letters* 48.5. e2020GL089412 2020GL089412, e2020GL089412.

Bre, Facundo, Juan M. Gimenez, and Víctor D. Fachinotti (2018). "Prediction of wind pressure coefficients on building surfaces using artificial neural networks". In: *Energy and Buildings* 158, pp. 1429–1441.

Breen, C. M., W. R. Currier, C. Vuyovich, Z. Miao, and L. R. Prugh (2024). "Snow Depth Extraction From Time-Lapse Imagery Using a Keypoint Deep Learning Model". In: *Water Resources Research* 60.7. e2023WR036682 2023WR036682, e2023WR036682.

Brenner, A. C., R. A. Blndschadler, R. H. Thomas, and H. J. Zwally (1983). "Slope-induced errors in radar altimetry over continental ice sheets". In: *Journal of Geophysical Research: Oceans* 88.C3, pp. 1617–1623.

Brenner, Anita C., John P. DiMarzio, and H. Jay Zwally (2007). "Precision and Accuracy of Satellite Radar and Laser Altimeter Data Over the Continental Ice Sheets". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.2, pp. 321–331.

Broeke, M. R. van den, E. M. Enderlin, I. M. Howat, P. Kuipers Munneke, B. P. Y. Noël, W. J. van de Berg, E. van Meijgaard, and B. Wouters (2016). "On the recent contribution of the Greenland ice sheet to sea level change". In: *The Cryosphere* 10.5, pp. 1933–1946.

Broeke, Michiel van den, Jonathan Bamber, Janneke Ettema, Eric Rignot, Ernst Schrama, Willem Jan van de Berg, Erik van Meijgaard, Isabella Velicogna, and

Bert Wouters (2009). "Partitioning Recent Greenland Mass Loss". In: *Science* 326.5955, pp. 984–986.

Broeke, Michiel van den, Jason Box, Xavier Fettweis, Edward Hanna, Brice Noël, Marco Tedesco, Dirk van As, Willem Jan van de Berg, and Leo van Kampenhout (Dec. 2017). "Greenland Ice Sheet Surface Mass Loss: Recent Developments in Observation and Modeling". In: *Current Climate Change Reports* 3.4, pp. 345–356.

Brooks, W. L. and R. P. Dooley (Apr. 1975). *Technical guidance and analytic services in support of SEASAT-A*. NASA Contractor Report TSC-W3-20. NASA-CR-141399. Silver Spring, MD, United States: Technology Service Corp., Silver Spring, MD.

Brown, G. (1977). "The average impulse response of a rough surface and its applications". In: *IEEE Transactions on Antennas and Propagation* 25.1, pp. 67–74.

Bu, Yuanpeng, Jinxuan Hu, Cheng Chen, Songhang Bai, Zuohui Chen, Tianyu Hu, Guwen Zhang, Na Liu, Chang Cai, Yuhao Li, Qi Xuan, Ye Wang, Zhongjing Su, Yun Xiang, and Yaming Gong (Jan. 2024). "ResNet incorporating the fusion data of RGB & hyperspectral images improves classification accuracy of vegetable soybean freshness". In: *Scientific Reports* 14.1, p. 2568.

Campos-Taberner, Manuel, Francisco Javier García-Haro, Beatriz Martínez, Emma Izquierdo-Verdiguier, Clement Atzberger, Gustau Camps-Valls, and María Amparo Gilabert (Oct. 2020). "Understanding deep learning in land use classification based on Sentinel-2 time series". In: *Scientific Reports* 10.1, p. 17188.

Carr, J. Rachel, Andreas Vieli, and Chris Stokes (2013). "Influence of sea ice decline, atmospheric warming, and glacier width on marine-terminating outlet glacier behavior in northwest Greenland at seasonal to interannual timescales". In: *Journal of Geophysical Research: Earth Surface* 118.3, pp. 1210–1226.

Chartrand, A. M. and I. M. Howat (2020). "Basal Channel Evolution on the Getz Ice Shelf, West Antarctica". In: *Journal of Geophysical Research: Earth Surface* 125.9. e2019JF005293 2019JF005293, e2019JF005293.

Chauché, N., A. Hubbard, J.-C. Gascard, J. E. Box, R. Bates, M. Koppes, A. Sole, P. Christoffersen, and H. Patton (2014). "Ice–ocean interaction and calving front morphology at two west Greenland tidewater outlet glaciers". In: *The Cryosphere* 8.4, pp. 1457–1468.

Chen, J. L., C. R. Wilson, D. D. Blankenship, and B. D. Tapley (2006). "Antarctic mass rates from GRACE". In: *Geophysical Research Letters* 33.11.

Chen, X., M. Patel, F. J. Pena Cantu, J. Park, J. Noa Turnes, L. Xu, K. A. Scott, and D. A. Clausi (2024). "MMSeaIce: a collection of techniques for improving sea ice mapping with a multi-task model". In: *The Cryosphere* 18.4, pp. 1621–1632.

Cheung, Mark, John Shi, Oren Wright, Lavendar Y. Jiang, Xujin Liu, and Jose M. F. Moura (Nov. 2020). "Graph Signal Processing and Deep Learning: Convolution, Pooling, and Topology". In: *IEEE Signal Processing Magazine* 37.6, pp. 139–149.

Chollet, François (2017). "Xception: Deep Learning with Depthwise Separable Convolutions". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807.

Chu, X., X. Yao, H. Duan, C. Chen, J. Li, and W. Pang (2022). "Glacier extraction based on high-spatial-resolution remote-sensing images using a deep-learning approach with attention mechanism". In: *The Cryosphere* 16.10, pp. 4273–4289.

Chuter, S. J. and J. L. Bamber (2015). "Antarctic ice shelf thickness from CryoSat-2 radar altimetry". In: *Geophysical Research Letters* 42.24, pp. 10, 721–10, 729.

Chuter, S. J., A. Zammit-Mangion, J. Rougier, G. Dawson, and J. L. Bamber (2022). "Mass evolution of the Antarctic Peninsula over the last 2 decades from a joint Bayesian inversion". In: *The Cryosphere* 16.4, pp. 1349–1367.

Clerc, Sébastien, Craig Donlon, Franck Borde, Nicolas Lamquin, Samuel E Hunt, Dave Smith, Malcolm McMillan, Jonathan Mittaz, Emma Woolliams, Matthew Hammond, et al. (2020). "Benefits and lessons learned from the Sentinel-3 tandem phase". In: *Remote Sensing* 12.17, p. 2668.

Cook, A. J. and D. G. Vaughan (2010). "Overview of areal changes of the ice shelves on the Antarctic Peninsula over the past 50 years". In: *The Cryosphere* 4.1, pp. 77–98.

Cooper, A.P.R. (1989). "Slope Correction By Relocation For Satellite Radar Altimetry". In: *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium,* vol. 4, pp. 2730–2733.

Core Writing Team, R.K. Pachauri, and L.A. Meyer, eds. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: Intergovernmental Panel on Climate Change, p. 151.

Crétaux, Jean-François, Muriel Bergé-Nguyen, Stephane Calmant, Nurzat Jamangulova, Rysbek Satylkanov, Florent Lyard, Felix Perosanz, Jacques Verron, Amanda Samine Montazem, Gianfranco Le Guilcher, Delphine Leroux, Joel Barrie, Philippe Maisongrande, and Pascal Bonnefond (2018). "Absolute Calibration or Validation of the Altimeters on the Sentinel-3A and the Jason-3 over Lake Issykkul (Kyrgyzstan)". In: *Remote Sensing* 10.11.

Curlander, John C and Robert N McDonough (1991). *Synthetic aperture radar*. Vol. 11. Wiley, New York.

Dastour, Hatef and Quazi K. Hassan (2023). "A Comparison of Deep Transfer Learning Methods for Land Use and Land Cover Classification". In: *Sustainability* 15.10.

Davis, C.H. (1997). "A robust threshold retracking algorithm for measuring ice-sheet surface elevation change from satellite radar altimeters". In: *IEEE Transactions on Geoscience and Remote Sensing* 35.4, pp. 974–979.

Davis, Curt H., Yonghong Li, Joseph R. McConnell, Markus M. Frey, and Edward Hanna (2005). "Snowfall-Driven Growth in East Antarctic Ice Sheet Mitigates Recent Sea-Level Rise". In: *Science* 308.5730, pp. 1898–1901.

Davis, Curt H. and Richard K. Moore (1993). "A combined surface-and volume-scattering model for ice-sheet radar altimetry". In: *Journal of Glaciology* 39.133, pp. 675–686.

Dawson, G. J. and J. L. Bamber (2017). "Antarctic Grounding Line Mapping From CryoSat-2 Radar Altimetry". In: *Geophysical Research Letters* 44.23, pp. 11, 886–11, 893.

Dawson, G. J. and J. C. Landy (2023). "Comparing elevation and backscatter retrievals from CryoSat-2 and ICESat-2 over Arctic summer sea ice". In: *The Cryosphere* 17.9, pp. 4165–4178.

de Rezende, Edmar R.S., Guilherme C.S. Ruppert, Antônio Theóphilo, Eric K. Tokuda, and Tiago Carvalho (2018). "Exposing computer generated images by using deep convolutional neural networks". In: *Signal Processing: Image Communication* 66, pp. 113–126.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255.

Detlefsen, Nicki S., Martin Jørgensen, and Søren Hauberg (2019). *Reliable training and estimation of variance networks*.

Dinniman, Michael S, Xylar S Asay-Davis, Benjamin K Galton-Fenzi, Paul R Holland, Adrian Jenkins, and Ralph Timmermann (2016). "Modeling ice shelf/ocean interaction in Antarctica: A review". In: *Oceanography* 29.4, pp. 144–153.

Donlon, Craig, Bruno Berruti, A Buongiorno, M-H Ferreira, P Féménias, J Frerick, P Goryl, U Klein, H Laur, C Mavrocordatos, et al. (2012a). "The global monitoring for environment and security (GMES) sentinel-3 mission". In: *Remote sensing of Environment* 120, pp. 37–57.

Donlon, Craig, Bruno Berruti, S Mecklenberg, Jens Nieke, Helge Rebhan, Ulf Klein, Alessandra Buongiorno, Constantin Mavrocordatos, Johannes Frerick, Bernd Seitz, et al. (2012b). "The sentinel-3 mission: Overview and status". In:

2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1711–1714.

Drinkwater, Mark, Richard Francis, Guy Ratier, and Duncan Wingham (June 2004). "The European Space Agency's Earth Explorer Mission CryoSat: Measuring variability in the cryosphere". In: *Annals of Glaciology* 39, pp. 313–320.

Edwards, Tamsin L., Sophie Nowicki, Ben Marzeion, Regine Hock, Heiko Goelzer, Hélène Seroussi, Nicolas C. Jourdain, Donald A. Slater, Fiona E. Turner, Christopher J. Smith, Christine M. McKenna, Erika Simon, Ayako Abe-Ouchi, Jonathan M. Gregory, Eric Larour, William H. Lipscomb, Antony J. Payne, Andrew Shepherd, Cécile Agosta, Patrick Alexander, Torsten Albrecht, Brian Anderson, Xylar Asay-Davis, Andy Aschwanden, Alice Barthel, Andrew Bliss, Reinhard Calov, Christopher Chambers, Nicolas Champollion, Youngmin Choi, Richard Cullather, Joshua Cuzzone, Christophe Dumas, Denis Felikson, Xavier Fettweis, Koji Fujita, Benjamin K. Galton-Fenzi, Rupert Gladstone, Nicholas R. Golledge, Ralf Greve, Tore Hattermann, Matthew J. Hoffman, Angelika Humbert, Matthias Huss, Philippe Huybrechts, Walter Immerzeel, Thomas Kleiner, Philip Kraaijenbrink, Sébastien Le clec'h, Victoria Lee, Gunter R. Leguy, Christopher M. Little, Daniel P. Lowry, Jan-Hendrik Malles, Daniel F. Martin, Fabien Maussion, Mathieu Morlighem, James F. O'Neill, Isabel Nias, Frank Pattyn, Tyler Pelle, Stephen F. Price, Aurélien Quiquet, Valentina Radić, Ronja Reese, David R. Rounce, Martin Rückamp, Akiko Sakai, Courtney Shafer, Nicole-Jeanne Schlegel, Sarah Shannon, Robin S. Smith, Fiammetta Straneo, Sainan Sun, Lev Tarasov, Luke D. Trusel, Jonas Van Breedam, Roderik van de Wal, Michiel van den Broeke, Ricarda Winkelmann, Harry Zekollari, Chen Zhao, Tong Zhang, and Thomas Zwinger (May 2021). "Projected land ice contributions to twenty-first-century sea level rise". In: *Nature* 593.7857, pp. 74–82.

Egido, Alejandro (Nov. 2023). "The Sentinel-3 Next Generation Topography (S3NG-TOPO) Mission; Enhancing Continuity, Performance and Hydrology Capabili-

ties". In: *2023 Ocean Surface Topography Science Team Meeting*. id.137. Ocean Surface Topography Science Team. Online.

Elabd, Emad, Hany Mohamed Hamouda, M. A. Mohamed Ali, and Yasser Fouad (May 2025). "Climate change prediction in Saudi Arabia using a CNN GRU LSTM hybrid deep learning model in al Qassim region". In: *Scientific Reports* 15.1, p. 16275.

Enderlin, Ellyn M., Ian M. Howat, Seongsu Jeong, Myoung-Jong Noh, Jan H. van Angelen, and Michiel R. van den Broeke (2014). "An improved mass budget for the Greenland ice sheet". In: *Geophysical Research Letters* 41.3, pp. 866–872.

EUMETSAT (Dec. 2017). *Sentinel-3 SRAL Marine User Handbook*. Doc.No. : EUM/OPS-SEN3/MAN/17/920901, Version v1A, e-signed, 12 December 2017. Eumetsat-Allee 1, D-64295 Darmstadt, Germany.

European Space Agency (2021). *CryoSat-2 Product Handbook*. Technical Report C2-LI-ACS-ESL-5319. Version Baseline E 1.0 – Draft C. European Space Agency.

— (2024). *Sentinel-3 Satellite and Payloads*. `https://sentiwiki.copernicus.eu/web/s3-mission`. Figure 2.

Everett, Alistair, Tavi Murray, Nick Selmes, David Holland, and Dominic Reeve (2021). "The Impacts of a Subglacial Discharge Plume On Calving, Submarine Melting, and Mélange Mass Loss At Helheim Glacier, South East Greenland". In: *Journal of Geophysical Research: Earth Surface* 126.3, e2020JF005910.

Fallah, B., M. Rostami, E. Russo, P. Harder, C. Menz, P. Hoffmann, I. Didovets, and F. F. Hattermann (2025). "Climate model downscaling in central Asia: a dynamical and a neural network approach". In: *Geoscientific Model Development* 18.1, pp. 161–180.

Fettweis, X., S. Hofer, U. Krebs-Kanzow, C. Amory, T. Aoki, C. J. Berends, A. Born, J. E. Box, A. Delhasse, K. Fujita, P. Gierz, H. Goelzer, E. Hanna, A. Hashimoto, P. Huybrechts, M.-L. Kapsch, M. D. King, C. Kittel, C. Lang, P. L. Langen, J. T. M. Lenaerts, G. E. Liston, G. Lohmann, S. H. Mernild, U. Mikolajewicz, K. Modali, R. H. Mottram, M. Niwano, B. Noël, J. C. Ryan, A. Smith, J. Streffing,

M. Tedesco, W. J. van de Berg, M. van den Broeke, R. S. W. van de Wal, L. van Kampenhout, D. Wilton, B. Wouters, F. Ziemen, and T. Zolles (2020). "GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet". In: *The Cryosphere* 14.11, pp. 3935–3958.

Flament, Thomas and Frédérique Rémy (2012). "Dynamic thinning of Antarctic glaciers from along-track repeat radar altimetry". In: *Journal of Glaciology* 58.211, pp. 830–840.

Folgoc, Loic Le, Vasileios Baltatzis, Sujal Desai, Anand Devaraj, Sam Ellis, Octavio E. Martinez Manzanera, Arjun Nair, Huaqi Qiu, Julia Schnabel, and Ben Glocker (2021). *Is MC Dropout Bayesian?*

Fox-Kemper, B., H.T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J.-B. Sallée, A.B.A. Slangen, and Y. Yu (2021). "Ocean, Cryosphere and Sea Level Change". In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 1211–1362.

Frajka-Williams, E., N. Foukal, and G. Danabasoglu (2023). "Should AMOC observations continue: how and why?" In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 381.2262, p. 20220195.

Frederikse, Thomas, Maya K Buchanan, Erwin Lambert, Robert E Kopp, Michael Oppenheimer, DJ Rasmussen, and Roderik SW van de Wal (2020). "Antarctic Ice Sheet and emission scenario controls on 21st-century extreme sea-level changes". In: *Nature communications* 11.1, p. 390.

Fretwell, Peter, Hamish D Pritchard, David G Vaughan, Jonathan L Bamber, Nicholas E Barrand, Robin Bell, Cesidio Bianchi, RG Bingham, Donald D Blankenship, Gino Casassa, et al. (2013). "Bedmap2: improved ice bed, surface and thickness datasets for Antarctica". In: *The cryosphere* 7.1, pp. 375–393.

Fricker, Helen Amanda and Laurie Padman (2012). "Thirty years of elevation change on Antarctic Peninsula ice shelves from multimission satellite radar altimetry". In: *Journal of Geophysical Research: Oceans* 117.C2.

Fukushima, Kunihiko (Apr. 1980). "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4, pp. 193–202.

Gal, Yarin and Zoubin Ghahramani (2016). *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*.

Garcia-Cardona, Cristina, Yen Ting Lin, and Tanmoy Bhattacharya (Jan. 2021). "UNCERTAINTY QUANTIFICATION FOR DEEP LEARNING REGRESSION MODELS IN THE LOW DATA LIMIT". In: pp. 379–393.

Gerrish, L. (2020). *The coastline of Kalaallit Nunaat/Greenland available as a shapefile and geopackage, covering the main land and islands, with glacier fronts updated as of 2017*. Dataset. Version 1.0.

Glen, E., A. Leeson, A. F. Banwell, J. Maddalena, D. Corr, O. Atkins, B. Noël, and M. McMillan (2025). "A comparison of supraglacial meltwater features throughout contrasting melt seasons: southwest Greenland". In: *The Cryosphere* 19.3, pp. 1047–1066.

Glorot, Xavier and Yoshua Bengio (13–15 May 2010). "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, pp. 249–256.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio (Nov. 2011). "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International*

Conference on Artificial Intelligence and Statistics. Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, pp. 315–323.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press.

Google (2023). *Deep Learning Tuning Playbook*. Accessed: 2025-09-27.

Gourmelen, N., M.J. Escorihuela, A. Shepherd, L. Foresta, A. Muir, A. Garcia-Mondéjar, M. Roca, S.G. Baker, and M.R. Drinkwater (2018). "CryoSat-2 swath interferometric altimetry for mapping ice elevation and elevation change". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1226–1242.

Gourmelen, Noel, Dan N. Goldberg, Kate Snow, Sian F. Henley, Robert G. Bingham, Satoshi Kimura, Anna E. Hogg, Andrew Shepherd, Jeremie Mouginot, Jan T. M. Lenaerts, Stefan R. M. Ligtenberg, and Willem Jan van de Berg (2017). "Channelized Melting Drives Thinning Under a Rapidly Melting Antarctic Ice Shelf". In: *Geophysical Research Letters* 44.19, pp. 9796–9804.

Goyal, Priya, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He (2018). *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour*.

Gray, L., D. Burgess, L. Copland, R. Cullen, N. Galin, R. Hawley, and V. Helm (2013). "Interferometric swath processing of Cryosat data for glacial ice topography". In: *The Cryosphere* 7.6, pp. 1857–1867.

Gray, L., D. Burgess, L. Copland, T. Dunse, K. Langley, and G. Moholdt (2017). "A revised calibration of the interferometric mode of the CryoSat-2 radar altimeter improves ice height and height change measurements in western Greenland". In: *The Cryosphere* 11.3, pp. 1041–1058.

Griggs, J.A. and J.L. Bamber (2011). "Antarctic ice-shelf thickness from satellite radar altimetry". In: *Journal of Glaciology* 57.203, pp. 485–498.

Gudmundsson, G. Hilmar, Fernando S. Paolo, Susheel Adusumilli, and Helen A. Fricker (2019). "Instantaneous Antarctic ice sheet mass loss driven by thinning ice shelves". In: *Geophysical Research Letters* 46.23, pp. 13903–13909.

Hahn, L., C. C. Ummenhofer, and Y.-O. Kwon (2018). "North Atlantic Natural Variability Modulates Emergence of Widespread Greenland Melt in a Warming Climate". In: *Geophysical Research Letters* 45.17, pp. 9171–9178.

Han, Wei, Zhengdong Zhang, Yu Zhang, Jiahui Yu, Chung-Cheng Chiu, James Qin, Anmol Gulati, Ruoming Pang, and Yonghui Wu (2020). *ContextNet: Improving Convolutional Neural Networks for Automatic Speech Recognition with Global Context*.

Hanna, Edward, John Cappelen, Xavier Fettweis, Sebastian H Mernild, Thomas L Mote, Ruth Mottram, Konrad Steffen, Thomas J Ballinger, and Richard J Hall (2021). "Greenland surface air temperature changes from 1981 to 2019 and implications for ice-sheet melt and mass-balance change". In: *International Journal of Climatology* 41, E1336–E1352.

Hanna, Edward, Thomas E Cropper, Richard J Hall, and John Cappelen (2016). "Greenland Blocking Index 1851–2015: a regional climate change signal". In: *International Journal of Climatology* 36.15, pp. 4847–4861.

Hanna, Edward, Philippe Huybrechts, Konrad Steffen, John Cappelen, Russell Huff, Christopher Shuman, Tristram Irvine-Fynn, Stephen Wise, and Michael Griffiths (2008). "Increased runoff from melt from the Greenland Ice Sheet: a response to global warming". In: *Journal of Climate* 21.2, pp. 331–341.

Hanna, Edward, Frank Pattyn, Francisco Navarro, Vincent Favier, Heiko Goelzer, Michiel R. van den Broeke, Miren Vizcaino, Pippa L. Whitehouse, Catherine Ritz, Kevin Bulthuis, and Ben Smith (2020). "Mass balance of the ice sheets and glaciers – Progress since AR5 and challenges". In: *Earth-Science Reviews* 201, p. 102976.

Hastie, Trevor (2020). "Ridge Regularization: An Essential Concept in Data Science". In: *Technometrics* 62.4. PMID: 36033922, pp. 426–433.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2015). *Deep Residual Learning for Image Recognition*.

He, Tong, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li (2018). *Bag of Tricks for Image Classification with Convolutional Neural Networks*.

Helm, V., A. Humbert, and H. Miller (2014). "Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2". In: *The Cryosphere* 8.4, pp. 1539–1559.

Hinton, Geoffrey, Nitish Srivastava, and Kevin Swersky (2012). *Overview of mini-batch gradient descent*. Lecture 6a, Neural Networks for Machine Learning.

Hochreiter, Sepp (Apr. 1991). "Untersuchungen zu dynamischen neuronalen Netzen". In.

Hogg, Anna E., Andrew Shepherd, Lin Gilbert, Alan Muir, and Mark R. Drinkwater (2018). "Mapping ice sheet grounding lines with CryoSat-2". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1191–1202.

Horn, Berthold K.P. and Brian G. Schunck (1981). "Determining optical flow". In: *Artificial Intelligence* 17.1, pp. 185–203.

Howat, I. M., C. Porter, B. E. Smith, M.-J. Noh, and P. Morin (2019). "The Reference Elevation Model of Antarctica". In: *The Cryosphere* 13.2, pp. 665–674.

Howat, Ian, Claire Porter, Myoung-Jon Noh, Erik Husby, Samuel Khuvis, Evan Danish, Karen Tomko, Judith Gardiner, Adelaide Negrete, Bidhyananda Yadav, James Klassen, Cole Kelleher, Michael Cloutier, Jesse Bakker, Jeremy Enos, Galen Arnold, Greg Bauer, and Paul Morin (2022). *The Reference Elevation Model of Antarctica - Mosaics, Version 2*. Version V1.

Hu, Jie, Li Shen, Samuel Albanie, Gang Sun, and Andrea Vedaldi (2019). *Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks*.

Hu, Jie, Li Shen, and Gang Sun (2018). "Squeeze-and-Excitation Networks". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141.

Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger (2017). "Densely Connected Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269.

Huang, Gao, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q. Weinberger (2016). "Deep Networks with Stochastic Depth". In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling. Cham: Springer International Publishing, pp. 646–661.

Huang, Jiahao, Liutao Yang, Fanwen Wang, Yinzhe Wu, Yang Nan, Weiwen Wu, Chengyan Wang, Kuangyu Shi, Angelica I. Aviles-Rivero, Carola-Bibiane Schönlieb, Daoqiang Zhang, and Guang Yang (2025a). "Enhancing global sensitivity and uncertainty quantification in medical image reconstruction with Monte Carlo arbitrary-masked mamba". In: *Medical Image Analysis* 99, p. 103334.

Huang, Qi, Jennifer Maddalena, and Malcolm McMillan (2025b). "Mapping Earth's cryosphere and inland waters with satellite radar altimetry". In: *The Innovation Geoscience* 3.2, p. 100134.

Huang, Qi, Malcolm McMillan, Alan Muir, Joe Phillips, and Thomas Slater (2024). "Multipeak retracking of radar altimetry waveforms over ice sheets". In: *Remote Sensing of Environment* 303, p. 114020.

Hurkmans, R. T. W. L., J. L. Bamber, C. H. Davis, I. R. Joughin, K. S. Khvorostovsky, B. S. Smith, and N. Schoen (2014). "Time-evolving mass loss of the Greenland Ice Sheet from satellite altimetry". In: *The Cryosphere* 8.5, pp. 1725–1740.

Ignéczi, Ádám, Andrew J. Sole, Stephen J. Livingstone, Amber A. Leeson, Xavier Fettweis, Nick Selmes, Noel Gourmelen, and Kate Briggs (2016). "Northeast sector of the Greenland Ice Sheet to undergo the greatest inland expansion

of supraglacial lakes during the 21st century". In: *Geophysical Research Letters* 43.18, pp. 9729–9738.

IMBIE Team (June 2018). "Mass balance of the Antarctic Ice Sheet from 1992 to 2017". In: *Nature* 558.7709, pp. 219–222.

— (2020). "Mass balance of the Greenland Ice Sheet from 1992 to 2018". In: *Nature* 579.7798, pp. 233–239.

Ioffe, Sergey and Christian Szegedy (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*.

IPF (May 2024). *Sentinel-3 Level 2 SRAL SM2 - LI Algorithm Theoretical Baseline Definition*. Tech. rep. S3MPC.ATBD.LI, Issue 4.3. Available from the European Space Agency (ESA). Sentinel-3 Mission Performance Centre (S3MPC).

isardSAT (n.d.). *Satellite Image*. Available via EUMESAT. Credit: isardSAT. Accessed: 2025-09-26.

Jackson, Rebecca H., Fiammetta Straneo, and David A. Sutherland (July 2014). "Externally forced fluctuations in ocean temperature at Greenland glaciers in non-summer months". In: *Nature Geoscience* 7.7, pp. 503–508.

Jacovi, Alon, Oren Sar Shalom, and Yoav Goldberg (2020). *Understanding Convolutional Neural Networks for Text Classification*.

Joughin, Ian, Waleed Abdalati, and Mark Fahnestock (Dec. 2004). "Large fluctuations in speed on Greenland's Jakobshavn Isbræ glacier". In: *Nature* 432.7017, pp. 608–610.

Joughin, Ian, Ben E. Smith, Ian M. Howat, Ted Scambos, and Twila Moon (2010). "Greenland flow variability from ice-sheet-wide velocity mapping". In: *Journal of Glaciology* 56.197, pp. 415–430.

Joughin, Ian, Dale Winebrenner, Mark Fahnestock, Ron Kwok, and William Krabill (1996). "Measurement of ice-sheet topography using satellite-radar interferometry". In: *Journal of Glaciology* 42.140, pp. 10–22.

Kashiparekh, Kathan, Jyoti Narwariya, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff (2019). *ConvTimeNet: A Pre-trained Deep Convolutional Neural Network for Time Series Classification*.

Kendall, Alex and Yarin Gal (2017). *What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?*

Kern, M., R. Cullen, B. Berruti, J. Bouffard, T. Casal, M. R. Drinkwater, A. Gabriele, A. Lecuyot, M. Ludwig, R. Midthassel, I. Navas Traver, T. Parrinello, G. Ressler, E. Andersson, C. Martin-Puig, O. Andersen, A. Bartsch, S. Farrell, S. Fleury, S. Gascoin, A. Guillot, A. Humbert, E. Rinne, A. Shepherd, M. R. van den Broeke, and J. Yackel (2020). "The Copernicus Polar Ice and Snow Topography Altimeter (CRISTAL) high-priority candidate mission". In: *The Cryosphere* 14.7, pp. 2235–2251.

Keskar, Nitish Shirish, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang (2017). *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*.

Kim, Yoon (2014). "Convolutional Neural Networks for Sentence Classification". In: *CoRR* abs/1408.5882.

Kimura, Satoshi, Adrian Jenkins, Heather Regan, Paul R. Holland, Karen M. Assmann, Daniel B. Whitt, Melchoir Van Wessem, Willem Jan van de Berg, Carleen H. Reijmer, and Pierre Dutrieux (2017). "Oceanographic Controls on the Variability of Ice-Shelf Basal Melting and Circulation of Glacial Meltwater in the Amundsen Sea Embayment, Antarctica". In: *Journal of Geophysical Research: Oceans* 122.12, pp. 10131–10155.

Koenker, Roger and Gilbert Bassett (1978). "Regression Quantiles". In: *Econometrica* 46.1, pp. 33–50.

Kokhanovsky, Alexander, Maxim Lamare, Olaf Danne, Carsten Brockmann, Marie Dumont, Ghislain Picard, Laurent Arnaud, Vincent Favier, Bruno Jourdain, Emmanuel Le Meur, Biagio Di Mauro, Teruo Aoki, Masashi Niwano, Vladimir Rozanov, Sergey Korkin, Sepp Kipfstuhl, Johannes Freitag, Maria Hoerhold,

Alexandra Zuhr, Diana Vladimirova, Anne-Katrine Faber, Hans Christian Steen-Larsen, Sonja Wahl, Jonas K. Andersen, Baptiste Vandecrux, Dirk van As, Kenneth D. Mankoff, Michael Kern, Eleonora Zege, and Jason E. Box (2019). "Retrieval of Snow Properties from the Sentinel-3 Ocean and Land Colour Instrument". In: *Remote Sensing* 11.19.

Konrad, Hannes, Andrew Shepherd, Lin Gilbert, Anna E Hogg, Malcolm McMillan, Alan Muir, and Thomas Slater (Apr. 2018). "Net retreat of Antarctic glacier grounding lines". en. In: *Nat. Geosci.* 11.4, pp. 258–262.

Kovaly, John J (1976). "Synthetic aperture radar". In: *Dedham*.

Krogh, Anders and John A. Hertz (1991). "A simple weight decay can improve generalization". In: *Proceedings of the 5th International Conference on Neural Information Processing Systems*. NIPS'91. Denver, Colorado: Morgan Kaufmann Publishers Inc., pp. 950–957.

Kvanum, A. F., C. Palerme, M. Müller, J. Rabault, and N. Hughes (2024). "Developing a deep learning forecasting system for short-term and high-resolution prediction of sea ice concentration". In: *EGUsphere* 2024, pp. 1–26.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell (2017). *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*.

Landy, Jack C., Geoffrey J. Dawson, Michel Tsamados, Mitchell Bushuk, Julienne C. Stroeve, Stephen E. L. Howell, Thomas Krumpen, David G. Babb, Alexander S. Komarov, Harry D. B. S. Heorton, H. Jakob Belter, and Yevgeny Aksenov (Sept. 2022). "A year-round satellite sea-ice thickness record from CryoSat-2". In: *Nature* 609.7927, pp. 517–522.

Landy, Jack C., Michel Tsamados, and Randall Kenneth Scharien (2019). "A Facet-Based Numerical Model for Simulating SAR Altimeter Echoes From Heterogeneous Sea Ice Surfaces". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.7, pp. 4164–4180.

Lawrence, Isobel R., Thomas W.K. Armitage, Michel C. Tsamados, Julienne C. Stroeve, Salvatore Dinardo, Andy L. Ridout, Alan Muir, Rachel L. Tilling, and

Andrew Shepherd (2021). "Extending the Arctic sea ice freeboard and sea level record with the Sentinel-3 radar altimeters". In: *Advances in Space Research* 68.2. 25 Years of Progress in Radar Altimetry, pp. 711–723.

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444.

LeCun, Yann A., Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller (2012). "Efficient BackProp". In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 9–48.

Lee, Jungkyu, Taeryun Won, Tae Kwan Lee, Hyemin Lee, Geonmo Gu, and Kiho Hong (2020). *Compounding the Performance Improvements of Assembled Techniques in a Convolutional Neural Network*.

Leeson, AA, Andrew Shepherd, Kate Briggs, Ian Howat, Xavier Fettweis, Mathieu Morlighem, and E. Rignot (Dec. 2014). "Supraglacial lakes on the Greenland ice sheet advance inland under warming climate". In: *Nature Climate Change* 5, pp. 51–55.

Legresy, Benoit, Fabrice Papa, Frederique Remy, Gaetan Vinay, Mathias van den Bosch, and Ouan-Zan Zanife (2005). "ENVISAT radar altimeter measurements over continental surfaces and ice caps using the ICE-2 retracking algorithm". In: *Remote Sensing of Environment* 95.2, pp. 150–163.

Lenaerts, Jan T. M., Brooke Medley, Michiel R. van den Broeke, and Bert Wouters (2019). "Observing and Modeling Ice Sheet Surface Mass Balance". In: *Reviews of Geophysics* 57.2, pp. 376–420.

Levermann, A. and R. Winkelmann (2016). "A simple equation for the melt elevation feedback of ice sheets". In: *The Cryosphere* 10.4, pp. 1799–1807.

Levermann, A., R. Winkelmann, T. Albrecht, H. Goelzer, N. R. Golledge, R. Greve, P. Huybrechts, J. Jordan, G. Leguy, D. Martin, M. Morlighem, F. Pattyn, D. Pollard, A. Quiquet, C. Rodehacke, H. Seroussi, J. Sutter, T. Zhang, J. Van Breedam, R. Calov, R. DeConto, C. Dumas, J. Garbe, G. H. Gudmundsson, M. J. Hoffman, A. Humbert, T. Kleiner, W. H. Lipscomb, M. Meinshausen, E. Ng, S. M. J. Nowicki, M. Perego, S. F. Price, F. Saito, N.-J. Schlegel, S. Sun, and R. S. W. van de Wal (2020). "Projecting Antarctica's contribution to future sea level rise from basal ice shelf melt using linear response functions of 16 ice sheet models (LARMIP-2)". In: *Earth System Dynamics* 11.1, pp. 35–76.

Levinsen, Joanna Fredenslund, Sebastian B. Simonsen, Louise Sandberg Sørensen, and René Forsberg (2016). "The Impact of DEM Resolution on Relocating Radar Altimetry Data Over Ice Sheets". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.7, pp. 3158–3163.

Li, Rui, Shunyi Zheng, Ce Zhang, Chenxi Duan, Libo Wang, and Peter M. Atkinson (Nov. 2021). "ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 181, pp. 84–98.

Li, T., G. J. Dawson, S. J. Chuter, and J. L. Bamber (2023). "Grounding line retreat and tide-modulated ocean channels at Moscow University and Totten Glacier ice shelves, East Antarctica". In: *The Cryosphere* 17.2, pp. 1003–1022.

Li, Weiran, Cornelis Slobbe, and Stef Lhermitte (2022). "A leading-edge-based method for correction of slope-induced errors in ice-sheet heights derived from radar altimetry". In: *The Cryosphere* 16.6, pp. 2225–2243.

Li, Xin, Eric Rignot, Jeremie Mouginot, and Bernd Scheuchl (2016). "Ice flow dynamics and mass loss of Totten Glacier, East Antarctica, from 1989 to 2015". In: *Geophysical Research Letters* 43.12, pp. 6366–6373.

Li, Zhiqiang, Jia Li, Xuyan Ma, Lei Guo, Long Li, Jiahao Dian, Lingshuai Kong, and Huiguo Ye (2025). "A Glacier Ice Thickness Estimation Method Based on Deep Convolutional Neural Networks". In: *Geosciences* 15.7.

Lin, Min, Qiang Chen, and Shuicheng Yan (2014). *Network In Network*.

Liu, Lin (2021). "A Review of Deep Learning for Cryospheric Studies". In: *Deep Learning for the Earth Sciences*. John Wiley & Sons, Ltd. Chap. 17, pp. 258–268.

Liu, Mingliang, Zemin Wang, Baojun Zhang, Shuang Wu, and Jiachun An (2023). "Extraction and Analysis of the Antarctic Ice Shelf Basal Channel". In: *IEEE Geoscience and Remote Sensing Letters* 20, pp. 1–5.

Liu, Yehao, Matteo Pagliardini, Tatjana Chavdarova, and Sebastian U. Stich (2021). *The Peril of Popular Deep Learning Uncertainty Estimation Methods*.

Loomis, Bryant D., Denis Felikson, Terence J. Sabaka, and Brooke Medley (2021). "High-Spatial-Resolution Mass Rates From GRACE and GRACE-FO: Global and Ice Sheet Analyses". In: *Journal of Geophysical Research: Solid Earth* 126.12. e2021JB023024 2021JB023024, e2021JB023024.

Loshchilov, Ilya and Frank Hutter (2017). *SGDR: Stochastic Gradient Descent with Warm Restarts*.

Love, A. (1985). "In memory of Carl A. Wiley". In: *IEEE Antennas and Propagation Society Newsletter* 27.3, pp. 17–18.

Luthcke, S. B., H. J. Zwally, W. Abdalati, D. D. Rowlands, R. D. Ray, R. S. Nerem, F. G. Lemoine, J. J. McCarthy, and D. S. Chinn (2006). "Recent Greenland Ice Mass Loss by Drainage System from Satellite Gravity Observations". In: *Science* 314.5803, pp. 1286–1289.

Magruder, Lori, Thomas Neumann, and Nathan Kurtz (2021). "ICESat-2 Early Mission Synopsis and Observatory Performance". In: *Earth and Space Science* 8.5. e2020EA001555 2020EA001555, e2020EA001555.

Mandawade, Tejas, Divyansh Shukla, Vaibhav Singh, and Davis Naidu (Jan. 2025). *Floodhawk: Satellite-based Flood Detection Using ResNet-50*. SSRN. Date Written: November 21, 2024.

Markus, Thorsten, Tom Neumann, Anthony Martino, Waleed Abdalati, Kelly Brunt, Beata Csatho, Sinead Farrell, Helen Fricker, Alex Gardner, David Harding, Michael Jasinski, Ron Kwok, Lori Magruder, Dan Lubin, Scott Luthcke, James

Morison, Ross Nelson, Amy Neuenschwander, Stephen Palm, Sorin Popescu, CK Shum, Bob E. Schutz, Benjamin Smith, Yuekui Yang, and Jay Zwally (2017). "The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation". In: *Remote Sensing of Environment* 190, pp. 260–273.

Martin, Thomas V., H. Jay Zwally, Anita C. Brenner, and Robert A. Bindschadler (1983). "Analysis and retracking of continental ice sheet radar altimeter waveforms". In: *Journal of Geophysical Research: Oceans* 88.C3, pp. 1608–1616.

Martins, Luana, Eber Pinto, and Philippe Maillard (Mar. 2023). *MartinsPintoMaillard2023 - Evaluating the performance of four satellite altimetry missions through spatial and temporal interpolation of in situ data.*

Maslov, Konstantin A., Claudio Persello, Thomas Schellenberger, and Alfred Stein (Jan. 2025). "Globally scalable glacier mapping by deep learning matches expert delineation accuracy". In: *Nature Communications* 16.1, p. 43.

McMillan, M., A. Muir, A. Shepherd, R. Escolà, M. Roca, J. Aublanc, P. Thibaut, M. Restano, A. Ambrozio, and J. Benveniste (2019). "Sentinel-3 Delay-Doppler altimetry over Antarctica". In: *The Cryosphere* 13.2, pp. 709–722.

McMillan, Malcolm, Hugh Corr, Andrew Shepherd, Andrew Ridout, Seymour Laxon, and Robert Cullen (2013). "Three-dimensional mapping by CryoSat-2 of subglacial lake volume changes". In: *Geophysical Research Letters* 40.16, pp. 4321–4327.

McMillan, Malcolm, Amber Leeson, Andrew Shepherd, Kate Briggs, Thomas W. K. Armitage, Anna Hogg, Peter Kuipers Munneke, Michiel van den Broeke, Brice Noël, Willem Jan van de Berg, Stefan Ligtenberg, Martin Horwath, Andreas Groh, Alan Muir, and Lin Gilbert (2016). "A high-resolution record of Greenland mass balance". In: *Geophysical Research Letters* 43.13, pp. 7002–7010.

McMillan, Malcolm, Alan Muir, and Craig Donlon (2021). "Brief communication: Ice sheet elevation measurements from the Sentinel-3A and Sentinel-3B tandem phase". In: *The Cryosphere* 15.7, pp. 3129–3134.

McMillan, Malcolm, Andrew Shepherd, Alan Muir, Julia Gaudelli, Anna E. Hogg, and Robert Cullen (2018). "Assessment of CryoSat-2 interferometric and non-interferometric SAR altimetry over ice sheets". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1281–1291.

McMillan, Malcolm, Andrew Shepherd, Aud Sundal, Kate Briggs, Alan Muir, Andrew Ridout, Anna Hogg, and Duncan Wingham (2014). "Increased ice losses from Antarctica detected by CryoSat-2". In: *Geophysical Research Letters* 41.11, pp. 3899–3905.

Meinert, Nis, Jakob Gawlikowski, and Alexander Lavin (June 2023). "The Unreasonable Effectiveness of Deep Evidential Regression". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 37.8, pp. 9134–9142.

Meinert, Nis and Alexander Lavin (2022). *Multivariate Deep Evidential Regression*.

Mertikas, Stelios, Achilleas Tripolitsiotis, Craig Donlon, Constantin Mavrocordatos, Pierre Féménias, Franck Borde, Xenophon Frantzis, Costas Kokolakis, Thierry Guinle, George Vergos, et al. (2020). "The ESA Permanent Facility for altimetry calibration: Monitoring performance of radar altimeters for Sentinel-3A, Sentinel-3B and Jason-3 using transponder and sea-surface calibrations with FRM standards". In: *Remote Sensing* 12.16, p. 2642.

Morlighem, Mathieu, Chris N Williams, Eric Rignot, Lu An, Jan Erik Arndt, Jonathan L Bamber, Ginny Catania, Nolwenn Chauché, Julian A Dowdeswell, Boris Dorschel, et al. (2017). "BedMachine v3: Complete bed topography and ocean bathymetry mapping of Greenland from multibeam echo sounding combined with mass conservation". In: *Geophysical research letters* 44.21, pp. 11–051.

Motyka, Roman J., Martin Truffer, Mark Fahnestock, John Mortensen, Søren Rysgaard, and Ian Howat (2011). "Submarine melting of the 1985 Jakobshavn Isbræ floating tongue and the triggering of the current retreat". In: *Journal of Geophysical Research: Earth Surface* 116.F1.

Mouginot, J., E. Rignot, and B. Scheuchl (2014). "Sustained increase in ice discharge from the Amundsen Sea Embayment, West Antarctica, from 1973 to 2013". In: *Geophysical Research Letters* 41.5, pp. 1576–1584.

Mouginot, Jeremie, Eric Rignot, Bernd Scheuchl, and Romain Millan (2017). "Comprehensive Annual Ice Sheet Velocity Mapping Using Landsat-8, Sentinel-1, and RADARSAT-2 Data". In: *Remote Sensing* 9.4.

Mouginot, Jérémie, Eric Rignot, Anders A. Bjørk, Michiel van den Broeke, Romain Millan, Mathieu Morlighem, Brice Noël, Bernd Scheuchl, and Michael Wood (2019). "Forty-six years of Greenland Ice Sheet mass balance from 1972 to 2018". In: *Proceedings of the National Academy of Sciences* 116.19, pp. 9239–9244.

Mugunthan, Jaya Sree, Claude R. Duguay, and Elena Zakharova (2023). "Machine learning based classification of lake ice and open water from Sentinel-3 SAR altimetry waveforms". In: *Remote Sensing of Environment* 299, p. 113891.

Muir, Alan (Sept. 19, 2024). *Cryo-TEMPO Algorithm Theoretical Basis Document Land Ice*. Tech. rep. LU_ESA_Cryo-TEMPO_ATBD_LI. Version 4.2. Checked by: Malcolm McMillan. ESA.

Murphy, Kevin P. (2023). *Probabilistic Machine Learning: Advanced Topics*. MIT Press.

Nair, Vinod and Geoffrey E. Hinton (2010). "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, pp. 807–814.

Nakayama, Yoshihiro, Georgy Manucharyan, Hong Zhang, Pierre Dutrieux, Hector S. Torres, Patrice Klein, Helene Seroussi, Michael Schodlok, Eric Rignot, and Dimitris Menemenlis (Nov. 2019). "Pathways of ocean heat towards Pine Island and Thwaites grounding lines". In: *Scientific Reports* 9.1, p. 16649.

Nias, Isabel J., Sophie Nowicki, Denis Felikson, and Bryant Loomis (2023). "Modeling the Greenland Ice Sheet's Committed Contribution to Sea Level

During the 21st Century". In: *Journal of Geophysical Research: Earth Surface* 128.2. e2022JF006914 2022JF006914, e2022JF006914.

Niehaus, H., L. Istomina, M. Nicolaus, R. Tao, A. Malinka, E. Zege, and G. Spreen (2024). "Melt pond fractions on Arctic summer sea ice retrieved from Sentinel-3 satellite data with a constrained physical forward model". In: *The Cryosphere* 18.2, pp. 933–956.

Nielsen, Karina, Lars Stenseng, Ole Baltazar Andersen, and Per Knudsen (2017). "The Performance and Potentials of the CryoSat-2 SAR and SARIn Modes for Lake Level Estimation". In: *Water* 9.6.

Nilsson, J., A. Gardner, L. Sandberg Sørensen, and R. Forsberg (2016). "Improved retrieval of land ice topography from CryoSat-2 data and its impact for volume-change estimation of the Greenland Ice Sheet". In: *The Cryosphere* 10.6, pp. 2953–2969.

Nilsson, J. and A. S. Gardner (2024). "Elevation Change of the Greenland Ice Sheet and its Peripheral Glaciers: 1992–2023". In: *Earth System Science Data Discussions* 2024, pp. 1–28.

Nilsson, J., A. S. Gardner, and F. S. Paolo (2022). "Elevation change of the Antarctic Ice Sheet: 1985 to 2020". In: *Earth System Science Data* 14.8, pp. 3573–3598.

Nilsson, Johan, Paul Vallelonga, Sebastian B. Simonsen, Louise Sandberg Sørensen, René Forsberg, Dorthe Dahl-Jensen, Motohiro Hirabayashi, Kumiko Goto-Azuma, Christine S. Hvidberg, Helle A. Kjær, and Kazuhide Satow (2015). "Greenland 2012 melt event effects on CryoSat-2 radar altimetry". In: *Geophysical Research Letters* 42.10, pp. 3919–3926.

Noël, Brice, Jan T. M. Lenaerts, William H. Lipscomb, Katherine Thayer-Calder, and Michiel R. van den Broeke (Nov. 2022). "Peak refreezing in the Greenland firn layer under future warming scenarios". In: *Nature Communications* 13.1, p. 6870.

Noor, Nabilah Abdullah (2018). "The optimal coastal retracked sea levels from saral/altika satellite altimetry over the southeast asia". In.

Nwbeeson (2024). *Pine Island Glacier*. `https://commons.wikimedia.org/wiki/File:PineIslandGlacier.svg`. SVG image. Own work. Licensed under CC0 1.0 via Wikimedia Commons.

Oerlemans, J. (1982). "Response of the antarctic ice sheet to a climatic warming: A model study". In: *Journal of Climatology* 2.1, pp. 1–11.

Oerlemans, J. and N.C. Hoogendoorn (1989). "Mass-Balance Gradients and Climatic Change". In: *Journal of Glaciology* 35.121, pp. 399–405.

Otosaka, I. N., A. Shepherd, E. R. Ivins, N.-J. Schlegel, C. Amory, M. R. van den Broeke, M. Horwath, I. Joughin, M. D. King, G. Krinner, S. Nowicki, A. J. Payne, E. Rignot, T. Scambos, K. M. Simon, B. E. Smith, L. S. Sørensen, I. Velicogna, P. L. Whitehouse, G. A, C. Agosta, A. P. Ahlstrøm, A. Blazquez, W. Colgan, M. E. Engdahl, X. Fettweis, R. Forsberg, H. Gallée, A. Gardner, L. Gilbert, N. Gourmelen, A. Groh, B. C. Gunter, C. Harig, V. Helm, S. A. Khan, C. Kittel, H. Konrad, P. L. Langen, B. S. Lecavalier, C.-C. Liang, B. D. Loomis, M. McMillan, D. Melini, S. H. Mernild, R. Mottram, J. Mouginot, J. Nilsson, B. Noël, M. E. Pattle, W. R. Peltier, N. Pie, M. Roca, I. Sasgen, H. V. Save, K.-W. Seo, B. Scheuchl, E. J. O. Schrama, L. Schröder, S. B. Simonsen, T. Slater, G. Spada, T. C. Sutterley, B. D. Vishwakarma, J. M. van Wessem, D. Wiese, W. van der Wal, and B. Wouters (2023). "Mass balance of the Greenland and Antarctic ice sheets from 1992 to 2020". In: *Earth System Science Data* 15.4, pp. 1597–1616.

Ovadia, Yaniv, Emily Fertig, Jie Ren, Zachary Nado, D Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek (2019). *Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift*.

Pan, Xiang, Yinghui Lu, Kun Zhao, Hao Huang, Mingjun Wang, and Haonan Chen (2021). "Improving Nowcasting of Convective Development by Incorporating Polarimetric Radar Variables Into a Deep-Learning Model". In: *Geophysical Research Letters* 48.21. e2021GL095302 2021GL095302, e2021GL095302.

Panaretos, Victor M. and Yoav Zemel (Mar. 2019). "Statistical Aspects of Wasserstein Distances". In: *Annual Review of Statistics and Its Application* 6.1, pp. 405–431.

Park, J. W., N. Gourmelen, A. Shepherd, S. W. Kim, D. G. Vaughan, and D. J. Wingham (2013). "Sustained retreat of the Pine Island Glacier". In: *Geophysical Research Letters* 40.10, pp. 2137–2142.

Park, Taewook, Yoshihiro Nakayama, and SungHyun Nam (Apr. 2024). "Amundsen Sea circulation controls bottom upwelling and Antarctic Pine Island and Thwaites ice shelf melting". In: *Nature Communications* 15.1, p. 2946.

Parmiggiani, F., M. Moctezuma-Flores, L. Guerrieri, and M. L. Battagliere (2018). "SAR analysis of the Larsen-C A-68 iceberg displacements". In: *International Journal of Remote Sensing* 39.18, pp. 5850–5858.

Parrinello, T., A. Shepherd, J. Bouffard, S. Badessi, T. Casal, M. Davidson, M. Fornari, E. Maestroni, and M. Scagliola (2018). "CryoSat: ESA's ice mission – Eight years in space". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1178–1190.

Passaro, Marcello, Laura Rautiainen, Denise Dettmering, Marco Restano, Michael G. Hart-Davis, Florian Schlembach, Jani Särkkä, Felix L. Müller, Christian Schwatke, and Jérôme Benveniste (2022). "Validation of an Empirical Subwaveform Retracking Strategy for SAR Altimetry". In: *Remote Sensing* 14.16.

Phillips, J. and M. McMillan (2026). "Assessment of Sentinel-3 altimeter performance over Antarctica using high resolution digital elevation models". In: *The Cryosphere* 20.3, pp. 1745–1769.

Polyak, B.T. (1964). "Some methods of speeding up the convergence of iteration methods". In: *USSR Computational Mathematics and Mathematical Physics* 4.5, pp. 1–17.

Porter, Claire, Ian Howat, Myoung-Jon Noh, Erik Husby, Samuel Khuvis, Evan Danish, Karen Tomko, Judith Gardiner, Adelaide Negrete, Bidhyananda Yadav, James Klassen, Cole Kelleher, Michael Cloutier, Jesse Bakker, Jeremy Enos,

Galen Arnold, Greg Bauer, and Paul Morin (2023). *ArcticDEM - Mosaics, Version 4.1*. Version V1.

Price, Stephen F., Antony J. Payne, Ian M. Howat, and Benjamin E. Smith (2011). "Committed sea-level rise for the next century from Greenland ice sheet dynamics during the past decade". In: *Proceedings of the National Academy of Sciences* 108.22, pp. 8978–8983.

Pritchard, HD, Stefan RM Ligtenberg, Helen A Fricker, David G Vaughan, Michiel R van den Broeke, and Laurence Padman (2012). "Antarctic ice-sheet loss driven by basal melting of ice shelves". In: *Nature* 484.7395, pp. 502–505.

Pronk, Maarten, Marieke Eleveld, and Hugo Ledoux (2024). "Assessing Vertical Accuracy and Spatial Coverage of ICESat-2 and GEDI Spaceborne Lidar for Creating Global Terrain Models". In: *Remote Sensing* 16.13.

Quartly, Graham D., Francesco Nencioli, Matthias Raynal, Pascal Bonnefond, Pablo Nilo Garcia, Albert Garcia-Mondéjar, Adrián Flores de la Cruz, Jean-Francois Crétaux, Nicolas Taburet, Marie-Laure Frery, Mathilde Cancet, Alan Muir, David Brockley, Malcolm McMillan, Saleh Abdalla, Sara Fleury, Emeline Cadier, Qi Gao, Maria Jose Escorihuela, Mònica Roca, Muriel Bergé-Nguyen, Olivier Laurain, Jérôme Bruniquel, Pierre Féménias, and Bruno Lucas (2020). "The Roles of the S3MPC: Monitoring, Validation and Evolution of Sentinel-3 Altimetry Observations". In: *Remote Sensing* 12.11.

Quartly, Graham D., Eero Rinne, Marcello Passaro, Ole B. Andersen, Salvatore Dinardo, Sara Fleury, Amandine Guillot, Stefan Hendricks, Andrey A. Kurekin, Felix L. Müller, Robert Ricker, Henriette Skourup, and Michel Tsamados (2019). "Retrieving Sea Level and Freeboard in the Arctic: A Review of Current Radar Altimetry Methodologies and Future Perspectives". In: *Remote Sensing* 11.7.

Raney, R. K. (2021). *Altimetry waveform illustration*. Johns Hopkins University Applied Physics Laboratory. URL: `https://www.aviso.altimetry.fr/en/techniques/altimetry/principle/delay-doppler-/-sar-altimetry.html` (visited on 09/26/2025).

Raney, R.K. (1998). "The delay/Doppler radar altimeter". In: *IEEE Transactions on Geoscience and Remote Sensing* 36.5, pp. 1578–1588.

Rapley, Chris, H Griffiths, V Squire, M Lefebvre, A Birks, Anita Brenner, C Brossier, L Clifford, Alexander Cooper, A Cowan, David Drewry, M Gorman, H Huckle, P Lamb, T Martin, N McIntyre, K Milne, E Novotny, G Peckham, and John Vesecky (Jan. 1983). *A Study of Satellite Radar Altimeter Operations Over Ice-covered Surfaces*.

Rasp, Stephan and Nils Thuerey (2021). "Data-Driven Medium-Range Weather Prediction With a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench". In: *Journal of Advances in Modeling Earth Systems* 13.2. e2020MS002405 2020MS002405, e2020MS002405.

Ravinder, Nitin, Andrew Shepherd, Inès Otosaka, Thomas Slater, Alan Muir, and Lin Gilbert (2024). "Greenland Ice Sheet Elevation Change From CryoSat-2 and ICESat-2". In: *Geophysical Research Letters* 51.24. e2024GL110822 2024GL110822, e2024GL110822.

Ravuri, Suman, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed (Sept. 2021). "Skilful precipitation nowcasting using deep generative models of radar". In: *Nature* 597.7878, pp. 672–677.

Rawat, Waseem and Zenghui Wang (Sept. 2017). "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review". In: *Neural Computation* 29.9, pp. 2352–2449.

Ray, Chris, Cristina Martin-Puig, Maria Paola Clarizia, Giulio Ruffini, Salvatore Dinardo, Christine Gommenginger, and Jérôme Benveniste (2015). "SAR Altimeter Backscattered Waveform Model". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.2, pp. 911–919.

Reed, Brad, J. A. Mattias Green, Adrian Jenkins, and G. Hilmar Gudmundsson (Jan. 2024). "Recent irreversible retreat phase of Pine Island Glacier". In: *Nature Climate Change* 14.1, pp. 75–81.

Remy, F., P. Mazzega, S. Houry, C. Brossier, and J.F. Minster (1989). "Mapping of the Topography of Continental Ice by Inversion of Satellite-altimeter Data". In: *Journal of Glaciology* 35.119, pp. 98–107.

Rémy, Frédérique and Soazig Parouty (2009). "Antarctic Ice Sheet and Radar Altimetry: A Review". In: *Remote Sensing* 1.4, pp. 1212–1239.

Ridley, J. K. and K. C. Partington (1988). "A model of satellite radar altimeter return from ice sheets". In: *International Journal of Remote Sensing* 9.4, pp. 601–624.

Rignot, E., G. Casassa, P. Gogineni, W. Krabill, A. Rivera, and R. Thomas (2004). "Accelerated ice discharge from the Antarctic Peninsula following the collapse of Larsen B ice shelf". In: *Geophysical Research Letters* 31.18.

Rignot, E., I. Velicogna, M. R. van den Broeke, A. Monaghan, and J. T. M. Lenaerts (2011). "Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise". In: *Geophysical Research Letters* 38.5.

Rignot, Eric, Jérémie 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.

Riley, Shawn, Stephen Degloria, and S.D. Elliot (Jan. 1999). "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity". In: *Internation Journal of Science* 5, pp. 23–27.

Roemer, S., B. Legrésy, M. Horwath, and R. Dietrich (2007). "Refined analysis of radar altimetry data applied to the region of the subglacial Lake Vostok/Antarctica". In: *Remote Sensing of Environment* 106.3, pp. 269–284.

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.

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

Russell, Rebecca L. and Christopher Reale (2021). *Multivariate Uncertainty in Deep Learning*.

Sandberg Sørensen, Louise, Sebastian B. Simonsen, René Forsberg, Kirill Khvorostovsky, Rakia Meister, and Marcus E. Engdahl (2018). "25 years of elevation changes of the Greenland Ice Sheet from ERS, Envisat, and CryoSat-2 radar altimetry". In: *Earth and Planetary Science Letters* 495, pp. 234–241.

Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen (2018). "MobileNetV2: Inverted Residuals and Linear Bottlenecks". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520.

Sappington, Jayson M., Kathleen M. Longshore, and David B. Thompson (2007). "Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert". In: *Journal of Wildlife Management* 71.5, pp. 1419–1426.

Schlegel, N.-J., D. N. Wiese, E. Y. Larour, M. M. Watkins, J. E. Box, X. Fettweis, and M. R. van den Broeke (2016). "Application of GRACE to the assessment of model-based estimates of monthly Greenland Ice Sheet mass balance (2003–2012)". In: *The Cryosphere* 10.5, pp. 1965–1989.

Schröder, L., M. Horwath, R. Dietrich, V. Helm, M. R. van den Broeke, and S. R. M. Ligtenberg (2019). "Four decades of Antarctic surface elevation changes from multi-mission satellite altimetry". In: *The Cryosphere* 13.2, pp. 427–449.

Scott, R. F., S. G. Baker, C. M. Birkett, W. Cudlip, S. W. Laxon, D. R. Mantripp, J. A. Mansley, J. G. Morley, C. G. Rapley, J. K. Ridley, F. Strawbridge, and D. J.

Wingham (1994). "A comparison of the performance of the ice and ocean tracking modes of the ERS-1 radar altimeter over non-ocean surfaces". In: *Geophysical Research Letters* 21.7, pp. 553–556.

Sedlmeier, Andreas, Thomas Gabor, Thomy Phan, Lenz Belzner, and Claudia Linnhoff-Popien (2020). "Uncertainty-based Out-of-Distribution Classification in Deep Reinforcement Learning". In: *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications.

Seehaus, T., A. J. Cook, A. B. Silva, and M. Braun (2018). "Changes in glacier dynamics in the northern Antarctic Peninsula since 1985". In: *The Cryosphere* 12.2, pp. 577–594.

Sellers, William D. (1969). "A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System". In: *Journal of Applied Meteorology and Climatology* 8.3, pp. 392–400.

SentiWiki (n.d.). *S3 Altimetry Instruments*. Accessed: 27 April 2025.

Seroussi, Hélène, Tyler Pelle, William H. Lipscomb, Ayako Abe-Ouchi, Torsten Albrecht, Jorge Alvarez-Solas, Xylar Asay-Davis, Jean-Baptiste Barre, Constantijn J. Berends, Jorge Bernales, Javier Blasco, Justine Caillet, David M. Chandler, Violaine Coulon, Richard Cullather, Christophe Dumas, Benjamin K. Galton-Fenzi, Julius Garbe, Fabien Gillet-Chaulet, Rupert Gladstone, Heiko Goelzer, Nicholas Golledge, Ralf Greve, G. Hilmar Gudmundsson, Holly Kyeore Han, Trevor R. Hillebrand, Matthew J. Hoffman, Philippe Huybrechts, Nicolas C. Jourdain, Ann Kristin Klose, Petra M. Langebroek, Gunter R. Leguy, Daniel P. Lowry, Pierre Mathiot, Marisa Montoya, Mathieu Morlighem, Sophie Nowicki, Frank Pattyn, Antony J. Payne, Aurélien Quiquet, Ronja Reese, Alexander Robinson, Leopekka Saraste, Erika G. Simon, Sainan Sun, Jake P. Twarog, Luke D. Trusel, Benoit Urruty, Jonas Van Breedam, Roderik S. W. van de Wal, Yu Wang, Chen Zhao, and Thomas Zwinger (2024). "Evolution of the Antarctic

Ice Sheet Over the Next Three Centuries From an ISMIP6 Model Ensemble". In: *Earth's Future* 12.9. e2024EF004561 2024EF004561, e2024EF004561.

Shallue, Christopher J., Jaehoon Lee, Joseph Antognini, Jascha Sohl-Dickstein, Roy Frostig, and George E. Dahl (2019). *Measuring the Effects of Data Parallelism on Neural Network Training.*

Shao, Changpeng (2018). *A Quantum Model for Multilayer Perceptron.*

Shepherd, Andrew, Lin Gilbert, Alan S. Muir, Hannes Konrad, Malcolm McMillan, Thomas Slater, Kate H. Briggs, Aud V. Sundal, Anna E. Hogg, and Marcus E. Engdahl (2019). "Trends in Antarctic Ice Sheet Elevation and Mass". In: *Geophysical Research Letters* 46.14, pp. 8174–8183.

Shepherd, Andrew, Erik Ivins, Geruo Aa, Valentina Barletta, Michael Bentley, Srinivas Bettadpur, Kate Briggs, David Bromwich, Rene Forsberg, Natalia Galin, Martin Horwath, Stan Jacobs, Ian Joughin, Matt King, Jan Lenaerts, Jilu Li, Stefan Ligtenberg, Adrian Luckman, Scott Luthcke, and H. Zwally (Nov. 2012). "A Reconciled Estimate of Ice-Sheet Mass Balance". In: *Science (New York, N.Y.)* 338, pp. 1183–9.

Shepherd, Andrew, Duncan Wingham, and Eric Rignot (2004). "Warm ocean is eroding West Antarctic Ice Sheet". In: *Geophysical Research Letters* 31.23.

Shridhar, Kumar, Felix Laumann, and Marcus Liwicki (2019). *Uncertainty Estimations by Softplus normalization in Bayesian Convolutional Neural Networks with Variational Inference.*

Siegfried, Matthew R. and Helen A. Fricker (2018). "Thirteen years of subglacial lake activity in Antarctica from multi-mission satellite altimetry". In: *Annals of Glaciology* 59.76pt1, pp. 42–55.

Simonsen, Sebastian B., Valentina R. Barletta, William T. Colgan, and Louise Sandberg Sørensen (2021). "Greenland Ice Sheet Mass Balance (1992–2020) From Calibrated Radar Altimetry". In: *Geophysical Research Letters* 48.3. e2020GL091216 2020GL091216, e2020GL091216.

Simonsen, Sebastian B. and Louise Sandberg Sørensen (2017). "Implications of changing scattering properties on Greenland ice sheet volume change from Cryosat-2 altimetry". In: *Remote Sensing of Environment* 190, pp. 207–216.

Simonyan, Karen and Andrew Zisserman (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*.

Slater, T., A. Shepherd, M. McMillan, A. Muir, L. Gilbert, A. E. Hogg, H. Konrad, and T. Parrinello (2018). "A new digital elevation model of Antarctica derived from CryoSat-2 altimetry". In: *The Cryosphere* 12.4, pp. 1551–1562.

Slater, Thomas, Anna E. Hogg, and Ruth Mottram (Oct. 2020). "Ice-sheet losses track high-end sea-level rise projections". In: *Nature Climate Change* 10.10, pp. 879–881.

Slater, Thomas, Andrew Shepherd, Malcolm Mcmillan, Thomas W. K. Armitage, Inès Otosaka, and Robert J. Arthern (2019). "Compensating Changes in the Penetration Depth of Pulse-Limited Radar Altimetry Over the Greenland Ice Sheet". In: *IEEE Transactions on Geoscience and Remote Sensing* 57.12, pp. 9633–9642.

Slater, Thomas, Andrew Shepherd, Malcolm McMillan, Amber Leeson, Lin Gilbert, Alan Muir, Peter Kuipers Munneke, Brice Noël, Xavier Fettweis, Michiel van den Broeke, and Kate Briggs (2021). "Increased variability in Greenland Ice Sheet runoff from satellite observations". In: *Nature Communications* 12.1, p. 6069.

Sluijterman, Laurens, Eric Cator, and Tom Heskes (May 2024). "How to evaluate uncertainty estimates in machine learning for regression?" In: *Neural Networks* 173, p. 106203.

Smith, B. E., B. Medley, X. Fettweis, T. Sutterley, P. Alexander, D. Porter, and M. Tedesco (2023). "Evaluating Greenland surface-mass-balance and firn-densification data using ICESat-2 altimetry". In: *The Cryosphere* 17.2, pp. 789–808.

Smith, Ben, Benjamin Jelley, Suzanne Dickinson, Tyler Sutterley, Thomas Neumann, and Kaitlin Harbeck (2021). *ATLAS/ICESat-2 L3B Gridded Antarctic and Arctic Land Ice Height, Version 1.*

Smith, Benjamin (Dec. 2021). *ICESat-2 Algorithm Theoretical Basis Document (ATBD) for Land-ice DEM (ATL14) and Land-ice Height Change (ATL15).* Tech. rep. Release 001. Prepared by Ben Smith (University of Washington Applied Physics Lab Polar Science Center), approved by Tom Neumann (NASA Goddard Space Flight Center). Available at `https://icesat-2.gsfc.nasa.gov/`. National Aeronautics and Space Administration (NASA), Goddard Space Flight Center.

Smith, Benjamin, Helen A. Fricker, Nicholas Holschuh, Alex S. Gardner, Susheel Adusumilli, Kelly M. Brunt, Beata Csatho, Kaitlin Harbeck, Alex Huth, Thomas Neumann, Johan Nilsson, and Matthew R. Siegfried (2019). "Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter". In: *Remote Sensing of Environment* 233, p. 111352.

Smith, Freddie Bickford, Jannik Kossen, Eleanor Trollope, Mark van der Wilk, Adam Foster, and Tom Rainforth (2025). *Rethinking Aleatoric and Epistemic Uncertainty.*

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958.

Stephens, Graeme L, Julia M Slingo, Eric Rignot, John T Reager, Maria Z Hakuba, Paul J Durack, John Worden, and Remy Rocca (2020). "Earth's water reservoirs in a changing climate". In: *Proceedings of the Royal Society A* 476.2236, p. 20190458.

Straneo, Fiammetta, Gordon S. Hamilton, David A. Sutherland, Leigh A. Stearns, Fraser Davidson, Mike O. Hammill, Garry B. Stenson, and Aqqalu Rosing-Asvid (Mar. 2010). "Rapid circulation of warm subtropical waters in a major glacial fjord in East Greenland". In: *Nature Geoscience* 3.3, pp. 182–186.

Straneo, Fiammetta and Patrick Heimbach (Dec. 2013). "North Atlantic warming and the retreat of Greenland's outlet glaciers". In: *Nature* 504.7478, pp. 36–43.

Sumbul, Gencer, Marcela Charfuelan, Begum Demir, and Volker Markl (July 2019). "Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding". In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE.

Suryawanshi, M. R., M. McMillan, J. Maddalena, F. Piras, J. Aublanc, J.-A. Daguzé, C. Grau, and Q. Huang (2025). "New radar altimetry datasets of Greenland and Antarctic surface elevation, 1991–2012". In: *The Cryosphere* 19.8, pp. 2855–2880.

Szandała, Tomasz (2021). *Bio-inspired Neurocomputing*. Springer Singapore.

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich (2015). "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9.

Tan, Aik Rui, Shingo Urata, Samuel Goldman, Johannes C. B. Dietschreit, and Rafael Gómez-Bombarelli (Dec. 2023). "Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles". In: *npj Computational Materials* 9.1, p. 225.

Tan, Mingxing and Quoc V. Le (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*.

Tang, Wensi, Guodong Long, Lu Liu, Tianyi Zhou, Michael Blumenstein, and Jing Jiang (2022). *Omni-Scale CNNs: a simple and effective kernel size configuration for time series classification*.

Tedesco, M. and X. Fettweis (2020). "Unprecedented atmospheric conditions (1948–2019) drive the 2019 exceptional melting season over the Greenland ice sheet". In: *The Cryosphere* 14.4, pp. 1209–1223.

Thakkar, Krushna (July 2024). *ResNets. Deep Learning Series*. URL: `https://medium.com/operations-research-bit/deep-learning-resnets-08e79e604fa0` (visited on 09/27/2025).

Thapa, Aakash, Teerayut Horanont, and Bipul Neupane (2022). "Parcel-Level Flood and Drought Detection for Insurance Using Sentinel-2A, Sentinel-1 SAR GRD and Mobile Images". In: *Remote Sensing* 14.23.

Thompson-Munson, M., J. E. Kay, and B. R. Markle (2024). "Greenland's firn responds more to warming than to cooling". In: *The Cryosphere* 18.7, pp. 3333–3350.

Thuy, Arthur and Dries F. Benoit (Dec. 2024). *Fast and reliable uncertainty quantification with neural network ensembles for industrial image classification.* arXiv preprint arXiv:2403.10182. Version 3.

Tilling, Rachel L, Andy Ridout, Andrew Shepherd, and Duncan J Wingham (2015). "Increased Arctic sea ice volume after anomalously low melting in 2013". In: *Nature Geoscience* 8.8, pp. 643–646.

Tilling, Rachel L., Andy Ridout, and Andrew Shepherd (2018). "Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data". In: *Advances in Space Research* 62.6. The CryoSat Satellite Altimetry Mission: Eight Years of Scientific Exploitation, pp. 1203–1225.

Trevisani, Sebastiano, Giordano Teza, and Peter L. Guth (2023). "Hacking the topographic ruggedness index". In: *Geomorphology* 439, p. 108838.

Van den Broeke, Michiel R., Jonathan Bamber, Jan Lenaerts, and Eric Rignot (Sept. 2011). "Ice Sheets and Sea Level: Thinking Outside the Box". In: *Surveys in Geophysics* 32.4, pp. 495–505.

Velicogna, I., T. C. Sutterley, and M. R. van den Broeke (2014). "Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data". In: *Geophysical Research Letters* 41.22, pp. 8130–8137.

Velicogna, I. and J. Wahr (2013). "Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the GRACE satellite data". In: *Geophysical Research Letters* 40.12, pp. 3055–3063.

Velicogna, Isabella, Yara Mohajerani, Geruo A, Felix Landerer, Jeremie Mouginot, Brice Noel, Eric Rignot, Tyler Sutterley, Michiel van den Broeke, Melchior

van Wessem, and David Wiese (2020). "Continuity of Ice Sheet Mass Loss in Greenland and Antarctica From the GRACE and GRACE Follow-On Missions". In: *Geophysical Research Letters* 47.8. e2020GL087291 10.1029/2020GL087291, e2020GL087291.

Villadsen, Heidi, Xiaoli Deng, Ole B Andersen, Lars Stenseng, Karina Nielsen, and Per Knudsen (2016). "Improved inland water levels from SAR altimetry using novel empirical and physical retrackers". In: *Journal of Hydrology* 537, pp. 234–247.

Wang, Fang, Jonathan L. Bamber, and Xiao Cheng (2015). "Accuracy and Performance of CryoSat-2 SARIn Mode Data Over Antarctica". In: *IEEE Geoscience and Remote Sensing Letters* 12.7, pp. 1516–1520.

Wang, Fang, Jonathan L. Bamber, Shuangbao Yang, Xiao Cheng, Shubin Liu, Youzhi Zhang, and Yunping Chi (2019). "Accuracy assessment and waveform analysis of CryoSat-2 SARIn mode data over Antarctica". In: *International Journal of Remote Sensing* 40.22, pp. 8418–8431.

Wang, Jiwen, Qiangqiang Yuan, Huanfeng Shen, Tingting Liu, Tongwen Li, Linwei Yue, Xiaogang Shi, and Liangpei Zhang (2020). "Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach". In: *Journal of Hydrology* 585, p. 124828.

Wang, Y., M. Ding, C. H. Reijmer, P. C. J. P. Smeets, S. Hou, and C. Xiao (2021). "The AntSMB dataset: a comprehensive compilation of surface mass balance field observations over the Antarctic Ice Sheet". In: *Earth System Science Data* 13.6, pp. 3057–3074.

Weertman, J. (1961). "Equilibrium Profile of Ice Caps". In: *Journal of Glaciology* 3.30, pp. 953–964.

Weyn, Jonathan A., Dale R. Durran, and Rich Caruana (2020). "Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere". In: *Journal of Advances in Modeling Earth Systems* 12.9. e2020MS002109 10.1029/2020MS002109, e2020MS002109.

Williamson, Andrew G., Ian C. Willis, Neil S. Arnold, and Alison F. Banwell (2018). "Controls on rapid supraglacial lake drainage in West Greenland: an Exploratory Data Analysis approach". In: *Journal of Glaciology* 64.244, pp. 208–226.

Wilson, Margaret F. J., Brian O'Connell, Colin Brown, Janine C. Guinan, and Anthony J. Grehan (2007). "Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope". In: *Marine Geodesy* 1-2, pp. 3–35.

Wingham, D. J., D. W. Wallis, and A. Shepherd (2009). "Spatial and temporal evolution of Pine Island Glacier thinning, 1995–2006". In: *Geophysical Research Letters* 36.17.

Wingham, D.J., C.R. Francis, S. Baker, C. Bouzinac, D. Brockley, R. Cullen, P. de Chateau-Thierry, S.W. Laxon, U. Mallow, C. Mavrocordatos, L. Phalippou, G. Ratier, L. Rey, F. Rostan, P. Viau, and D.W. Wallis (2006a). "CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields". In: *Advances in Space Research* 37.4. Natural Hazards and Oceanographic Processes from Satellite Data, pp. 841–871.

Wingham, Duncan, Chris Rapley, and Griffiths D (Sept. 1986). "New Techniques in Satellite Altimeter Tracking Systems". In: *IGARSS 86 Symposium*.

Wingham, Duncan J, Martin J Siegert, Andrew Shepherd, and Alan S Muir (Apr. 2006b). "Rapid discharge connects Antarctic subglacial lakes". en. In: *Nature* 440.7087, pp. 1033–1036.

Wingham, Duncan J., Andrew J. Ridout, Remko Scharroo, Robert J. Arthern, and C. K. Shum (1998). "Antarctic Elevation Change from 1992 to 1996". In: *Science* 282.5388, pp. 456–458.

Wouters, Bert and Ingo Sasgen (Dec. 2022). "Increasing Freshwater Fluxes from the Greenland Ice Sheet Observed from Space". In: *Oceanography* 35.4, pp. 22–23.

Wu, Yuxin and Kaiming He (2018). *Group Normalization*.

Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He (2017). "Aggregated Residual Transformations for Deep Neural Networks". In: *2017 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987–5995.

Xu, Bing, Naiyan Wang, Tianqi Chen, and Mu Li (2015). *Empirical Evaluation of Rectified Activations in Convolutional Network*.

Yan, Changqing, Guangpeng Yang, Zeyun Liang, Han Cheng, Genghong Wu, Amit Kumar Srivastava, Qiang Yu, and Gang Zhao (2025). "CO-ResNetRS50-SSL: Enhanced convolution and semi-supervised learning for accurate rice growth stage recognition in complex field conditions". In: *European Journal of Agronomy* 168, p. 127631.

Yi, Donghui, H. Jay Zwally, and Xiaoli Sun (2005). "ICESat measurement of Greenland ice sheet surface slope and roughness". In: *Annals of Glaciology* 42, pp. 83–89.

Yu, Xiao, Xie Hu, Guoquan Wang, Kaicun Wang, and Xuelong Chen (2022). "Machine-Learning Estimation of Snow Depth in 2021 Texas Statewide Winter Storm Using SAR Imagery". In: *Geophysical Research Letters* 49.17. e2022GL099119 2022GL099119, e2022GL099119.

Zagoruyko, Sergey and Nikos Komodakis (2017). *Wide Residual Networks*.

Zemp, M., E. Thibert, M. Huss, D. Stumm, C. Rolstad Denby, C. Nuth, S. U. Nussbaumer, G. Moholdt, A. Mercer, C. Mayer, P. C. Joerg, P. Jansson, B. Hynek, A. Fischer, H. Escher-Vetter, H. Elvehøy, and L. M. Andreassen (2013). "Reanalysing glacier mass balance measurement series". In: *The Cryosphere* 7.4, pp. 1227–1245.

Zevenbergen, Lyle W. and Colin R. Thorne (1987). "Quantitative analysis of land surface topography". In: *Earth Surface Processes and Landforms* 12.1, pp. 47–56.

Zhang, Aston (2020). *Dive into Deep Learning - Residual Networks (ResNet). Section 7.6*.

Zhang, Xin, Liangxiu Han, Lianghao Han, and Liang Zhu (2020). "How Well Do Deep Learning-Based Methods for Land Cover Classification and Object

Detection Perform on High Resolution Remote Sensing Imagery?" In: *Remote Sensing* 12.3.

Zheng, Yue, Xiaoran Zhang, Yongchao Zhou, Yiping Zhang, Tuqiao Zhang, and Raziyeh Farmani (Apr. 2025). "Deep representation learning enables cross-basin water quality prediction under data-scarce conditions". In: *npj Clean Water* 8.1, p. 33.

Zhi, Wei, Alison P. Appling, Heather E. Golden, Joel Podgorski, and Li Li (Mar. 2024). "Deep learning for water quality". In: *Nature Water* 2.3, pp. 228–241.

Zinck, Ann-Sofie Priergaard, Bert Wouters, Erwin Lambert, and Stef Lhermitte (Sept. 2023). "Unveiling spatial variability within the Dotson Melt Channel through high-resolution basal melt rates from the Reference Elevation Model of Antarctica". In: *The Cryosphere* 17.9, pp. 3785–3801.

Zwally, H. J., M. B. Giovinetto, M. A. Beckley, and J. L. Saba (2012). *Antarctic and Greenland drainage systems*. Dataset.

Zwally, H. Jay (1989). "Growth of Greenland Ice Sheet: Interpretation". In: *Science* 246.4937, pp. 1589–1591.

Zwally, H. Jay, Mario B. Giovinetto, Jun Li, Helen G. Cornejo, Matthew A. Beckley, Anita C. Brenner, Jack L. Saba, and Donghui Yi (2005). "Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992–2002". In: *Journal of Glaciology* 51.175, pp. 509–527.

Zwally, H. Jay, Jun Li, Anita C. Brenner, Matthew Beckley, Helen G. Cornejo, John DiMarzio, Mario B. Giovinetto, Thomas A. Neumann, John Robbins, Jack L. Saba, and et al. (2011). "Greenland ice sheet mass balance: distribution of increased mass loss with climate warming; 2003–07 versus 1992–2002". In: *Journal of Glaciology* 57.201, pp. 88–102.