Developing a radar-based machine learning snowfall retrieval algorithm

by

Fraser King

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Geography

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Dr. Norman B. Wood Research Scientist, Space Science and Engineering Center, University of Wisconsin-Madison
Supervisor(s):	Dr. Christopher G. Fletcher Associate Professor, Dept. of Geography and Environmental Management, University of Waterloo
Internal Member:	Dr. Richard Kelly Professor, Dept. Geography and Environmental Management, University of Waterloo
Internal-External Member:	Dr. Andrea Scott Associate Professor, Systems Design Engineering (SYDE), University of Waterloo
Other Member(s):	Dr. Lisa Milani ESSIC Assistant Research Scientist, NASA Goddard Space Flight Center, University of Maryland

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis is presented as a collection of peer-reviewed journal articles which assess the capabilities of current snowfall retrieval algorithms and offer new insights into algorithmic enhancements using remote sensing datasets and machine learning techniques. Each manuscript was produced in a collaborative manner with contributions from multiple colleagues at various international institutions, as described below for each respective article.

The first of these papers (presented in chapter 2) was published in *Earth and Space Science* and evaluates the capabilities and limitations of using spaceborne remote sensing snowfall estimates from the NASA CloudSat satellite for constraining current gridded snow water equivalent datasets. The data collection, experimental design, analysis and writing was done by myself (FK) with subsequent manuscript draft editing, suggestions and general project direction provided by Dr. Christopher Fletcher (CF).

King, F., & Fletcher, C. G. (2020). Using CloudSat-CPR Retrievals to Estimate Snow Accumulation in the Canadian Arctic. *Earth and Space Science*, 7(2), e2019EA000776. https://doi.org/10.1029/2019EA000776

The second paper (chapter 3) was published in the *Journal of Applied Meteorology and Climatology* and contains a novel, proof-of-concept machine learning-based snowfall retrieval algorithm developed from vertically pointing surface radar data at a single site in southern Ontario. Data collection, experimental design, algorithm development, analysis and writing were performed by FK; project idea was developed by George Duffy (GD) and FK; editing and project direction was provided by GD and CF.

King, F., Duffy, G., & Fletcher, C. G. (2022). A Centimeter-Wavelength Snowfall Retrieval

Algorithm Using Machine Learning. *Journal of Applied Meteorology and Climatology*, 1(aop). https://doi.org/10.1175/JAMC-D-22-0036.1

The third and final paper presented in this work (chapter 4) was published in *Atmospheric Measurement Techniques* and expands upon the retrieval algorithm introduced in chapter 3. Chapter 4 uses additional surface radar data from multiple sites across the northern hemisphere along with more sophisticated deep learning methods, to develop a low uncertainty, robust precipitation retrieval. The project concept was developed by FK and GD, data provided by FK, Claire Pettersen (CP), Kerstin Ebell (KE) and Lisa Milani (LM), methods developed by FK, GD and CF, model design by FK, data processing and experiments performed by FK, manuscript writing from FK, editing from FK, CF, GD, LM, CP and KE.

King, F., Duffy, G., Milani, L., Fletcher, C. G., Pettersen, C., & Ebell, K. (2022). DeepPrecip: A deep neural network for precipitation retrievals. *Atmospheric Measurement Techniques*, 15(20), 6035–6050. https://doi.org/10.5194/amt-15-6035-2022

Abstract

As global temperatures continue to rise, snowfall patterns are expected to respond in a complex, nonlinear manner. Changes in the quantity of snowmelt-derived freshwater will impact global water-energy budgets, flood frequencies and intensities, and regional water resource management practices. Remote sensing is an observational alternative to poorly constrained reanalysis-derived estimates of snow water equivalent (SWE) across Arctic regions. We present an outlier detection methodology which leverages remote sensing data from CloudSat to constrain reanalysis product estimates of SWE. This analysis highlights areas and periods of high uncertainty in the gridded reanalysis datasets, and identifies a systematic positive SWE bias (of 14.9%) in a blended reanalysis product as a consequence of these low-quality estimates.

The ability to use remotely sensed observations to characterize error in surface SWE estimates is incredibly powerful, however, remote sensing datasets are not without their own sources of uncertainty. This work also, therefore, examines the capabilities of a machine learning snowfall retrieval algorithm trained on vertically pointing surface radar data at a Global Precipitation Measurement (GPM) validation experiment site in southern Ontario. Random forest (RF) retrieval performance is compared to an ensemble of traditional $Z_e - S$ power law relationships, with the RF consistently displaying the lowest overall error. The RF also demonstrates a level of robustness not present in the power law relations, with low error when applied to unseen observations from a study site with a different regional climate.

We further extend this methodology to a general machine learning precipitation retrieval across the wider northern hemisphere (NH) using additional input covariates, data from multiple sites over a longer time period, and by adopting a more sophisticated deep learning paradigm.

The DeepPrecip retrieval algorithm displays a 187% improvement in snowfall retrieval accuracy when compared to traditional $Z_e - S/R$ power law relationships, and a 21% improvement over the aforementioned RF. The highly generalized nature of DeepPrecip facilitates its application to unseen data with only a small performance reduction. DeepPrecip also provides insight into the regions of the vertical column (below 1 km and between $1.5 - 2$ km) most effective in contributing to high retrieval accuracy, highlighting the important role of ML in current and future spaceborne remote sensing precipitation missions.

Keywords: machine learning, remote sensing, snowfall, precipitation, radar, CloudSat, snow water equivalent, snow accumulation, neural network

Acknowledgements

I would like to express my deepest appreciation and thanks to my advisor Dr. Christopher Fletcher for his continued support, guidance and supervision over the past three years. His unwavering dedication to the scientific method, combined with his enthusiasm for research, have been invaluable resources and a strong source of personal motivation during my PhD. I would also like to extend a special thank you to Dr. Richard Kelly for his continued mentorship and collaboration across multiple research projects, and for freely sharing his deep love and appreciation of snow. Additionally, I thank my other committee members for their insightful comments and adventurous spirit (Dr. Andrea Scott and Dr. Lisa Milani), along with all of my other incredible coauthors over the years. I would also like to recognize the Canadian Space Agency (CSA) and Natural Sciences and Engineering Research Council (NSERC) for funding the majority of this research. Finally, this endeavour would not have been possible without the ongoing support of my family, friends, Riley, and Bruce, for which I will always be grateful.

Dedication

For PD King.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1D 1-dimensional

3D-VAR . . . . . 3-dimensional variational

AMSR-E . . . . . Advanced Microwave Scanning Radiometer for Earth Observing System

B4 . . . . . Blended-4

C . . . . . Celsius

C3VP . . . . . Canadian CloudSat/CALIPSO Validation Project

CAA . . . . . Canadian Arctic archipelago

CALIPSO . . . . . Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation

CanSISE . . . . . Canadian Sea Ice and Snow Evolution

CARE . . . . . Centre for Atmospheric Research Experiments

CCI . . . . . Climate change initiative

CESM . . . . . Community Earth System Model

CHRS . . . . . University of California Irvine Center for Hydrometeorology and Remote Sensing

CMC . . . . . Canadian Meteorological Center

CMIP5 . . . . . Community Model Intercomparison Project 5

CNN . . . . . Convolutional neural network

CPR . . . . . Cloud profiling radar

CV . . . . . Cross validation

dBZ . . . . . Decibel relative to Z

DFIR . . . . . Double Fence Intercomparison Reference

DJF . . . . . December, January, February

DO-Op . . . . . Daylight only operation

DOV . . . . . Doppler velocity

DP . . . . . DeepPrecip

DPR . . . . . Dual-frequency precipitation radar

DT . . . . . Decision tree

ECCC . . . . . Environment and Climate Change Canada

ECMWF . . . . . European Centre for Medium-Range Weather Forecasts

ERA . . . . . European Centre for Medium-Range Weather Forecasts Re-Analysis

ESA . . . . . European Space Agency

ESM . . . . . Earth system model

FSS . . . . . Federal snow sampler

GCPEx . . . . . Global Precipitation Measurement Cold Season Precipitation Experiment

GHz . . . . . Gigahertz

GJ . . . . . Gigajoules

GLDAS . . . . . Global Land Data Assimilation System

GPM . . . . . Global Precipitation Measurement

GPU . . . . . Graphics processing unit

GRACE . . . . . Gravity Recovery and Climate Experiment

GZS . . . . . GCPEx reflectivity-snowfall

ICE-POP . . . . . International Collaborative Experiments for Pyeongchang 2018 Olympic and Paralympic Winter Games

IFS . . . . . Integrated forecasting system

IMProToo . . . . . Improved MRR processing tool

IPU . . . . . Intelligence processing unit

IPZS . . . . . ICE-POP reflectivity-snowfall

ISBA . . . . . Interactions between Soil, Biosphere, and Atmosphere

LiDAR . . . . . Light Detection and Ranging

LR . . . . . Linear regression

MERRA . . . . . Modern-Era Retrospective analysis for Research and Applications

ML . . . . . Machine learning

MLP . . . . . Multilayer perceptron

MLR . . . . . Multiple linear regression

MP . . . . . Marshall-Palmer

MRR . . . . . Micro rain radar

MSE . . . . . Mean square error

MXPol . . . . . Mobile X-band polarimetric weather radar

NA . . . . . North America

NASA . . . . . National Aeronautics and Space Administration

NCMR . . . . . National Centre for Meteorological Research

NH . . . . . Northern hemisphere

NN . . . . . Neural network

OE . . . . . Optimal estimation

OI . . . . . Optimal interpolation

OLS . . . . . Ordinary least squares

P . . . . . Precipitation

PDF . . . . . Probability density function

PERSIANN . . . Precipitation Estimation from Remotely Sensed Information using

PG . . . . . Profile group

PI . . . . . Prediction interval

PIP . . . . . Precipitation Imaging Package

PSD . . . . . Particle size distribution

R . . . . . Rainfall

RCP . . . . . Representative concentration pathway

RF . . . . . Random forest

RFL/Z_e . . . . . Reflectivity

RMSE . . . . . Root mean square error

S . . . . . Snowfall

SCD . . . . . Snow cover days

SCE . . . . . Snow cover extent

SD . . . . . Snow depth

SE . . . . . Standard error

SH . . . . . Southern hemisphere

SHAP . . . . . SHapley Additive exPlanations

SON . . . . . September, October, November

SPW . . . . . Spectral width

SWCN . . . . . Surface Weather and Climate monitoring Network

SWE . . . . . Snow water equivalent

SWIPA . . . . . Snow, Water, Ice and Permafrost in the Arctic

SWM . . . . . Snow water mass

SVM . . . . . Support vector machine

TMP . . . . . Temperature

TPU . . . . . Tensor processing unit

VertiX . . . . . Vertically pointing X-band radar

WMO-SPICE . World Meteorological Organization Solid Precipitation Intercomparison Experiment

WVL . . . . . Vertical wind velocity

Research Graphic

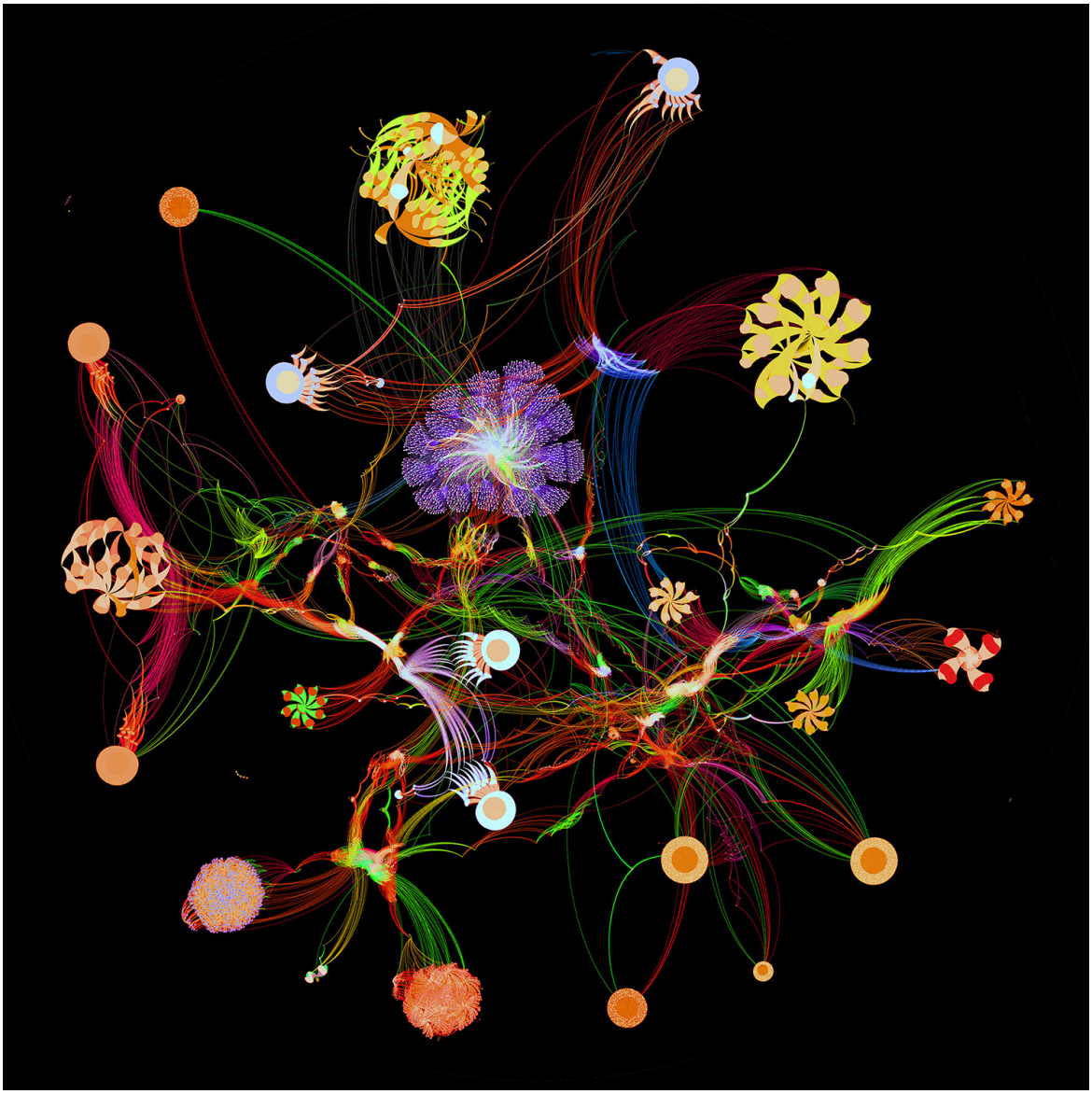

Figure 1: **DeepPrecip model compute graph**. This network diagram render is the culmination of the years of work put into completing my PhD. It shows the connections between 1.7 million nodes using 3.8 million edges in our convolutional deep learning precipitation retrieval.

Chapter 1

Introduction

1.1 Background

Anthropogenic climate change has led to irreversible impacts to the Arctic, as natural processes and human systems are bent beyond their ability to adapt and keep pace with warming global temperatures (Cassotta et al., 2022; Brown et al., 2017). Arctic air temperatures have increased by more than double that of the global mean, with winter temperatures now 6° C higher than 1981-2010 averages due to increases in atmospheric greenhouse gas concentrations (Nummelin et al., 2017). This extreme regional warming (also known as Arctic amplification), is a phenomenon which threatens to substantially impact the daily lives of high latitude communities and their surrounding ecosystems (Cohen et al., 2014; Tei et al., 2017). Increases in surface air temperature are tightly coupled to many other systems within the terrestrial cryosphere, including a higher frequency of wildfires, permafrost degradation, increased river discharge, higher surface runoff intensity, and changes to snow and ice quantities (Fig. 1.1) (Troy et al., 2012; Schuur et al., 2015; Kasischke and Turetsky, 2006). Changes in one or more components of this system can

trigger positive feedback loops (e.g. snow albedo feedback), where increased warming facilitates additional changes within the system (e.g. more losses to snow cover), which in turn contributes to further warming (Thackeray and Fletcher, 2016). Understanding how these systems are changing throughout the cryosphere is therefore of critical importance in the adaptation and mitigation of climate change.

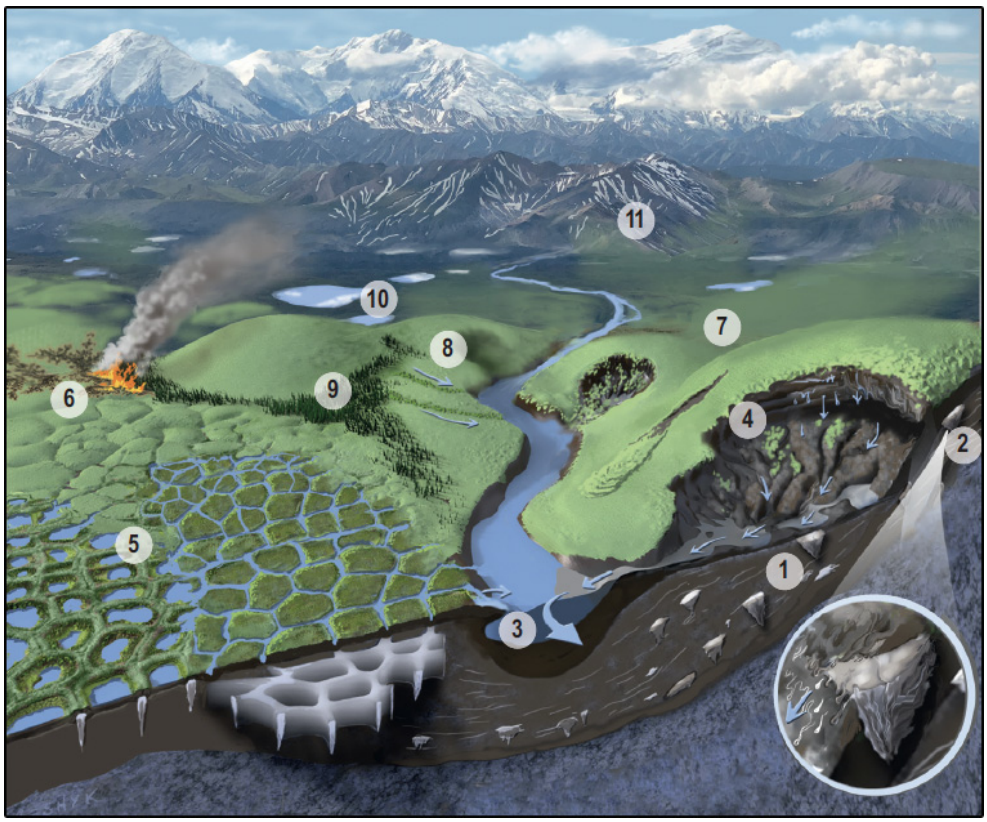

Figure 1.1: **Land surface components of the Arctic** including 1) permafrost; 2) surface ice; 3) river discharge; 4) surface thawing; 5) sitting water; 6) wildfire; 7) tundra; 8) low-lying shrubs; 9) boreal forest; 10) lake ice; and 11) seasonal snow (Cassotta et al., 2022).

Snowfall and surface snow water equivalent (SWE) quantities are tightly coupled climate variables of hydrologic significance in the cryosphere. Both variables are integral, linked components

of the global water and energy cycle with significant impacts to regional freshwater availability and ecosystem sustainability (Musselman et al., 2021; Etchevers et al., 2017; Gray and Landine, 2011; Pörtner et al., 2019). The majority of people living in the northern hemisphere (NH) receive a substantial quantity of annual precipitation in the form of snowfall, including, even, southern regions like California which are typically thought of as being snow-free (Fig. 1.2) (Sturm et al., 2017). Over a sixth of the world's population relies on freshwater produced from snowmelt for general consumption and horticultural purposes, with many more areas relying on snowmelt as a primary energy source (Sturm et al., 2017). The financial implications stemming from climate change reductions in snow cover extent (SCE) to the outdoor recreational industry and agricultural sector are astounding, as losses are anticipated to be in the *trillions*, not billions, of dollars over the next several decades (Euskirchen et al., 2013).

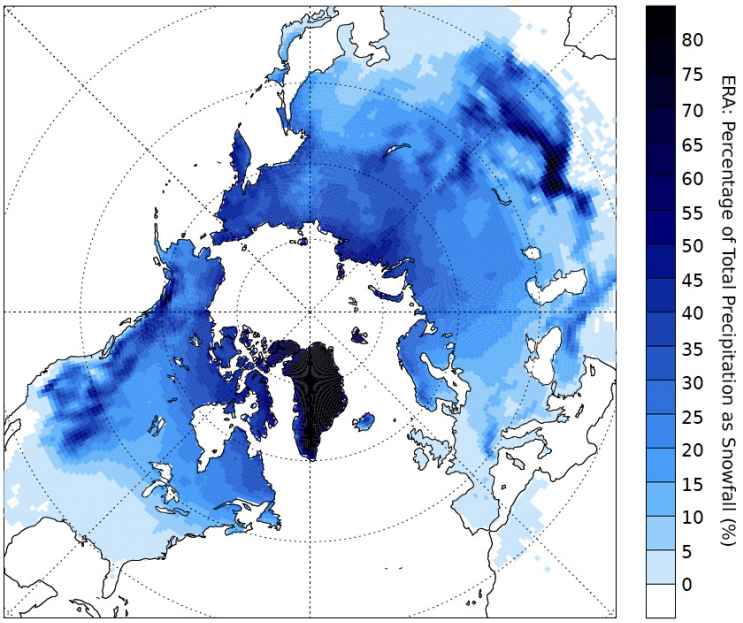

Figure 1.2: **Percentage of total NH precipitation falling as snow** calculated for the period 2000-2010 from ERA-Interim (Sturm et al., 2017).

However, increased Arctic air temperatures do not necessarily coincide with less snow throughout the region, as interactions with precipitation result in complex, nonlinear changes to snow quantity, duration and extent (Cassotta et al., 2022; Liston and Hiemstra, 2011; Trenberth, 2011). The 2017 Snow, Water, Ice and Permafrost in the Arctic (SWIPA) project demonstrate that Community Model Intercomparison Project 5 (CMIP5) projections (forced using an RCP8.5 emission scenario) display increases in maximum snow water equivalent (SWE_{max}) of $15-30\%$ in high Arctic regions over the next 50 years, driven by a moisture-rich atmosphere combined with continued cold temperatures to facilitate increased snowfall (Fig. 1.3.a) (Brown et al., 2017). Changes in snow depth (SD) over North America (NA) are not as clearly defined as SWE_{max} and do not appear statistically significant (Vincent et al., 2015). Under the same CMIP5 scenario, SWIPA also notes that the total number of snow cover days (SCD) are expected to decrease by $10-20\%$ across the Arctic by 2090 (Fig. 1.3.b). Pan-Arctic multi-dataset analyses of springtime SCE have also shown negative trends of $14-46\%$ over recent decades (with similar decreases over summer), further highlighting the extreme changes occurring throughout the region (Brown et al., 2010; Mudryk et al., 2017). The magnitude of these changes also depends greatly on the forcing scenario used, with substantial differences in SWE_{max}, SD, SCE and SCD projected changes between RCP2.6 and RCP8.5 scenarios (Brown et al., 2017; Cassotta et al., 2022).

New in situ campaigns, improvements to climate model resolution and representation of subgrid-scale processes, and large remote sensing data repositories from continuously orbiting Earth-observing satellites (now regularly measured in petabytes) have been invaluable resources for enhancing our current understanding of changing snow trends (Guo et al., 2017). However, due to the high variability in snowfall occurrence, sublimation effects, blowing snow, and the complex microphysical properties of frozen hydrometeors, large uncertainties still remain (Skofronick-

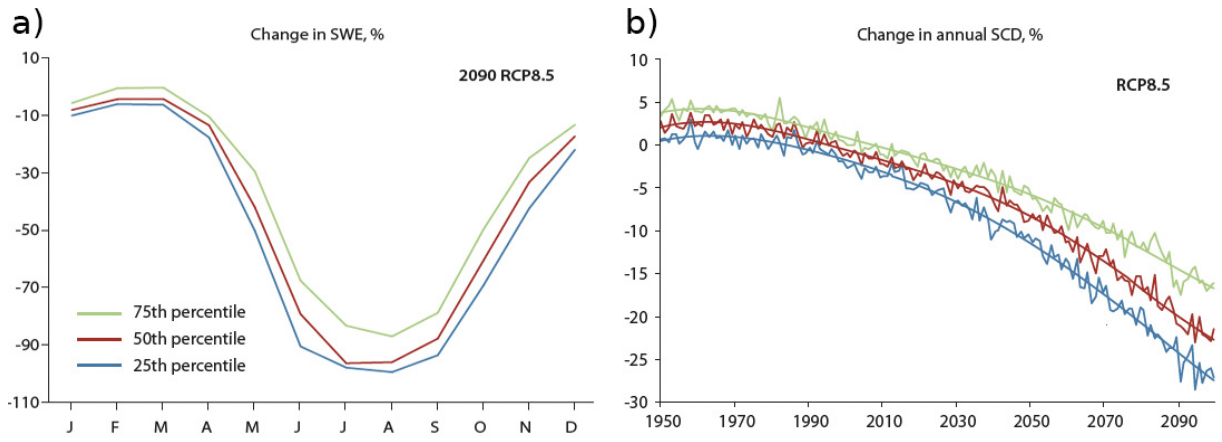

Figure 1.3: **Projected relative change in percent Arctic SWE and SCD to 1986-2005 averages**. **a)** CMIP5 (16 ensemble members under RCP8.5 forcing scenario) seasonal projected relative changes in SWE (%) by 2090; and **b)** the same in **a)** but for changes in total annual SCD. Each colored line represents a different percentile level from the ensemble (Brown et al., 2017).

Jackson et al., 2019; Sturm and Stuefer, 2013; Anderton et al., 2004). Shrinking the magnitude of the aforementioned snow trend ranges, and reducing uncertainties in future projections, are critically important steps towards preparing for and mitigating climate change driven disasters in the coming decades (Sturm et al., 2017). In the following sections, we will review the strengths and limitations of commonly used techniques for measuring snow accumulation over vast, remote regions.

1.1.1 Snow accumulation estimation systems

In situ measurements

Measurements taken in place (in situ) are typically considered as some of the highest quality observations for monitoring changes in snow accumulation (Doesken and Robinson, 2009).

Accumulation is measured as the change in depth of fresh snow falling to the surface over a given region in terms of its snow water equivalent (SWE) (Groisman and Davies, 2001). The SWE content of a snowpack at a point i is defined as $SWE_i = \rho_i \times SD_i$, where ρ is the snowpack density in kg m^3 and SD is the snow depth in mm. In situ SWE monitoring techniques typically involve either snow gauge instrumentation at a single point (e.g. Pluvio/Geonor gauges) or from performing multiple manual snow survey measurements from a sampling device like the Federal Snow Sampler (FSS) or the Environment Canada ESC30 (Egli et al., 2009; Kinar and Pomeroy, 2015; Brown and Mote, 2009). King et al. (2022c) have also demonstrated the effectiveness of using the built-in Light Detection and Ranging (LiDAR) modules on consumer smartphones to map the spatial variability of snow depth, however these methods have not yet been validated on larger, km-scale measurement campaigns.

The high temporal sampling frequency of in situ snowfall gauges allows for detailed monitoring of accumulation at minute-scale resolution (King and Fletcher, 2019; Saha et al., 2021). This high sampling frequency facilitates analyses of short-duration snowfall events and provides detailed insights into the evolution of synoptic storms. However, the cost of installing and maintaining such equipment often results in a sparsely distributed measurement network (Doesken and Robinson, 2009). This observational sparsity is especially notable in Canada, with less than 1% of all Canadian gauges currently operating in Arctic regions above 70° North (Mekis et al., 2018). Undercatch is another issue which contributes to additional uncertainty in gauge estimates (Fassnacht, 2004). During periods of strong winds, a substantial percentage of frozen hydrometeors can be carried over and away from the gauge top orifice, leading to an underestimation of the true accumulation value by upwards of 100% depending on the wind shielding techniques being employed (Skofronick-Jackson et al., 2017; Kochendorfer et al., 2022). Transfer functions

can be applied to account for this underestimation, however these functions are closely coupled to the atmospheric conditions in which they are derived and therefore do not generalize well to new regional climates (Colli et al., 2020).

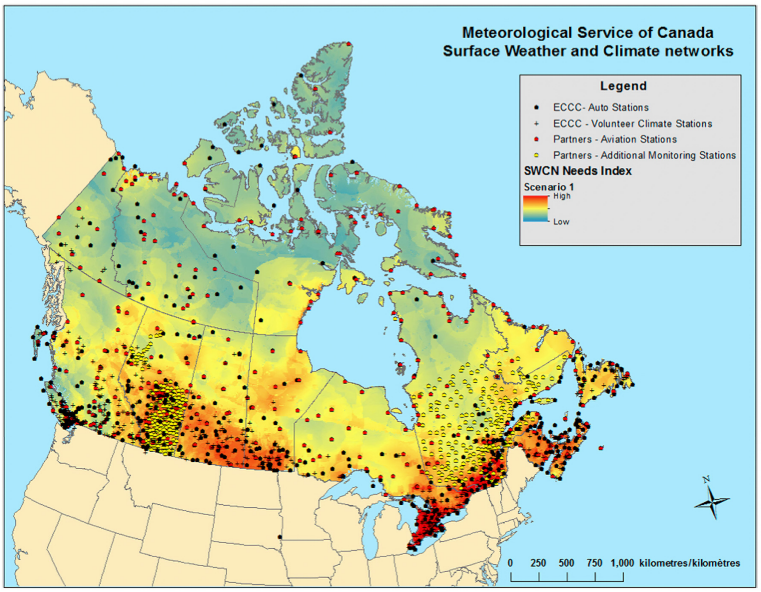

Figure 1.4: **Weather station locations across Canada**. All 1735 Environment and Climate Change Canada (ECCC) and partner surface weather station locations (as of 2016) within Canada's Surface Weather and Climate monitoring Network (SWCN) (Mekis et al., 2018).

Manual measurements do not suffer from the same sets of problems as in situ gauges, as these observations are performed by hand using trained field technicians (Doesken and Robinson, 2009). This is a labour-intensive process typically performed at various points across an observation site (ranging from a few hundred meters to a km in scale) over multiple days, which provides detailed snapshots in time of accumulated snowfall (Dyer and Mote, 2006). As a consequence of using field technicians, the temporal frequency of observations gathered from these measurement campaigns is much lower than that of weather gauges (Kinar and Pomeroy, 2015). Further, observations

are also prone to human observer bias and uncertainty from performing the measurement by eye (Brown and Mote, 2009). The limited length of measurement campaigns (ranging from a few days to weeks at a time), facilitates large temporal gaps in manual observation networks. Accumulated snowfall must therefore be estimated using ancillary methods during these gap periods.

Reanalyses and climate models

Reanalysis systems are popular methods for filling spatiotemporal gaps in snow observations recorded using in situ techniques. These systems use historical assimilated observations to constrain numerical models which can then simulate land surface, ocean and atmospheric interactions which, in turn, drive the global climate (Hurrell et al., 2013; Dee et al., 2011; Sun et al., 2004). The data assimilation portion of this process can be performed in a variety of different ways of varying complexity (e.g. sequential continuous, non-sequential intermittent), with the goal of bringing model estimates closer to reality, and producing a more reliable estimator (Hersbach et al., 2020; Gelaro et al., 2017; Navon, 2009). With complete spatial coverage, these systems provide a wealth of information which can be used to advance our understanding of changing snow patterns (Derksen and Brown, 2012; Rienecker et al., 2011; Dee et al., 2011; Duffy et al., 2021). However, due to the requisite historical data for constraining the numerical model, reanalysis systems are unable to estimate projected future changes in snow accumulation.

When unconstrained by historical data, climate models forced with potential future climate scenarios (e.g. possible future pathways describing global emissions of greenhouse gases) are able to forecast anticipated changes in snow accumulation under different emerging climates (Cassotta et al., 2022; Brown et al., 2017). The predictive capacity of climate models allows for

planning to begin earlier (or to initiate social programs in time) to mitigate issues arising from warming global temperatures (Forster et al., 2020; Bonan, 2008). Physically-based fully-coupled climate models with multiple component modules (Fig. 1.5) are able to simulate many of the natural processes occurring between land, ocean and atmosphere in order to capture emergent climate responses which are more challenging to model using conceptual models (at the cost of being significantly more computationally expensive) (Danabasoglu et al., 2020; Hurrell et al., 2013). However, climate models suffer from biases and uncertainties in their assumptions and internal parameterizations which can propagate through to their estimates of snow accumulation (Meehl et al., 2000; Mullan et al., 2017; Cassotta et al., 2022).

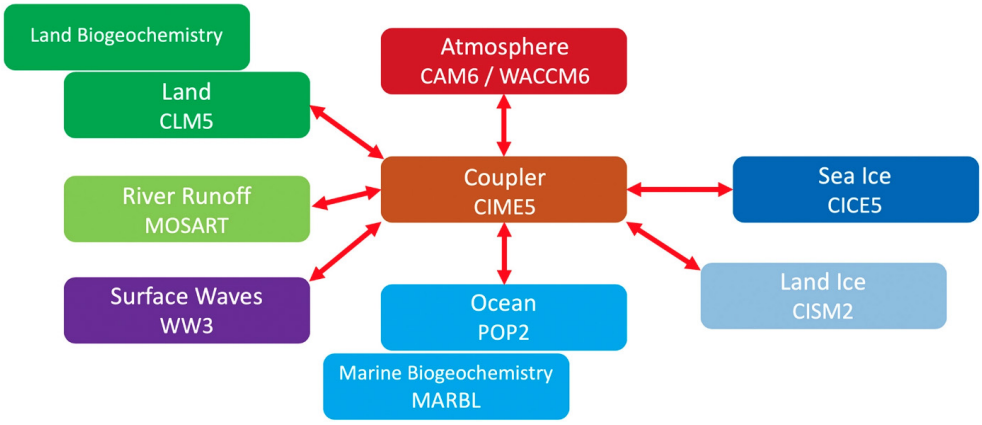

Figure 1.5: **Diagram of the Community Earth System Model version 2 (CESM-2) component modules and coupling mechanism** (Danabasoglu et al., 2020).

Reanalysis systems are beholden to their own set of internal assumptions, biases in assimilated data, and uncertainties stemming from simplified empirical relationships which represent important sub grid-scale processes like convection and precipitation (Lange et al., 2015; Franzke et al., 2015; Stone and Risbey, 1990). While the vast majority of reanalysis systems display increases in Arctic precipitation in recent decades, there remains high uncertainty regarding the magnitude

of this trend, as these products often demonstrate an inability to close the Arctic freshwater budget (Lindsay et al., 2014). A similar trend is also noted in reanalysis snow water mass (SWM) estimates, with individual Arctic reanalysis systems displaying large inter-product spread in their climatological means and anomalies, as shown in Fig. 1.6 (Mudryk et al., 2015). Comparisons of long-term (1891-2014) global precipitation trends from CMIP5 and CMIP6 show that the multi-model mean closely matches observed precipitation trends, however this is with a large spread in individual models within the ensemble (Vicente-Serrano et al., 2022). Additionally, large uncertainties in snowfall trends remain a problem at the regional scale due to the strong interannual variability of precipitation (Cassotta et al., 2022).

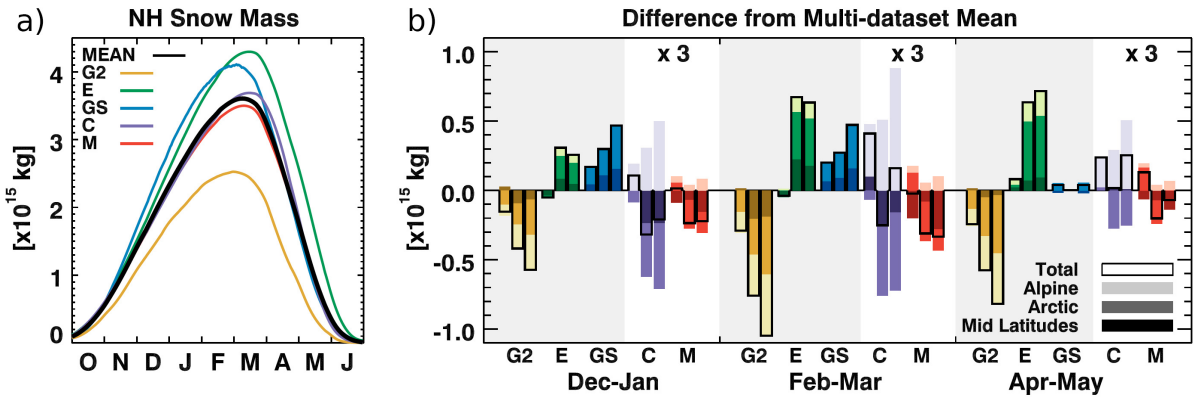

Figure 1.6: **NH SWM differences between five gridded snow model and reanalysis products (G2: GLDAS-2, E: ERA-Interim/Land, M: MERRA, GS: GlobSnow, C: Crocus). a)** Climatological SWM estimates from each gridded product (along with the multidataset mean) from 1981-2010; and **b)** Differences from the multi-dataset mean in each product's estimate of SWM for various seasons and regions (Mudryk et al., 2015).

Remote sensing retrievals

Satellite-based remote sensing instruments are an alternative method for filling in situ observational gaps using radar retrieval algorithms with near global coverage (Dietz et al., 2012; Skofronick-Jackson et al., 2019; Stephens et al., 2018; King and Fletcher, 2021). Instead of directly observing surface SWE quantities, radar instrument estimates of snowfall are derived, which are then linked to snow accumulation. Retrievals from radar systems like CloudSat's cloud profiling radar (CPR) or the Global Precipitation Measurement (GPM) core satellite's Dual-frequency precipitation radar (DPR) measure the intensity of backscattered energy from Rayleigh scattering interactions with falling snow particles in the atmosphere (Im et al., 2001; Skofronick-Jackson et al., 2019). In order to relate this backscattered energy (x) to a specific climatic variable like snowfall ($\hat{y}$), additional information about the atmospheric state at the time of measurement ($\hat{b}$) like the particle size, shape, density or fallspeed is required (Wood and L'Ecuyer, 2021). More formally, this type of Bayesian retrieval is called Optimal Estimation (OE). OE allows for a priori information to be explicitly included in the retrieval (Eq. 1.1) where the forward model F transforms a set of inputs x and additional state variables $\hat{b}$ into a snowfall estimate $\hat{y}$ with ϵ representing the total uncertainty, error and noise from all sources (Maahn et al., 2020). The resulting snowfall estimates can then be used to produce spatially-complete, gridded snowfall datasets over vast, remote Arctic regions as shown in Figure 1.7 produced using CloudSat's 2C-SNOW-PROFILE (Kulie and Milani, 2018).

$$\hat{y} = F(x, \hat{b}) + \epsilon \tag{1.1}$$

A consequence of physically-based retrievals is that the accuracy of the snowfall estimate

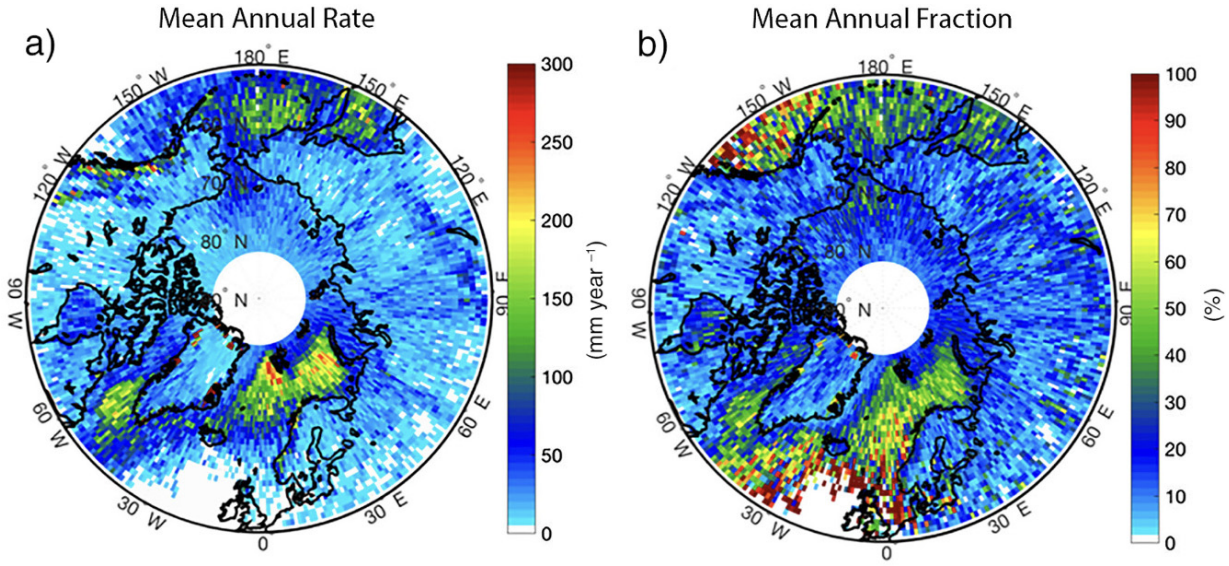

Figure 1.7: **Gridded NH CloudSat snowfall estimates from 2006-2010**. **a)** Mean annual cumuliform snowfall rate; and **b)** mean annual fraction of total snowfall attributed as cumuliform snowfall (Kulie and Milani, 2018).

is closely tied to the ability in which the OE algorithm can minimize a loss function and find a maximal probability density function (PDF) based on an assumed atmospheric state (Maahn et al., 2020). Knowledge gaps in a priori databases of observed particle microphysics for said assumed state will directly propagate through the forward model inversion of the retrieval process to produce an estimate of snowfall with high uncertainty (Wood et al., 2015; Chase et al., 2020; Souverijns et al., 2017; Sun et al., 2011). Recent work by Wellmann et al. (2020) suggests that two thirds of the precipitation variance for graupel and hail comes from uncertainties in microphysical properties like particle fallspeed. Gilmore et al. (2004) present similar sensitivities to microphysical assumptions, with surface accumulation estimates of solid precipitation varying by a factor of 3 to 4 based on changes to assumed particle density parameterizations. New validation campaigns (e.g. the Canadian CloudSat/CALIPSO Validation Programme (C3VP), GPM Cold

Season Precipitation Experiment (GCPEx), and the International Collaborative Experiments for Pyeongchang 2018 Olympic and Paralympic Winter Games (ICE-POP)) are increasing the size and quality of current a priori knowledge databases of hydrometeor particle microphysics under different regional climates, however large knowledge gaps still remain (Petersen et al., 2007; Skofronick-Jackson et al., 2015; Kim et al., 2021).

Recently, researchers have begun to turn their attention towards statistical models to fill these knowledge gaps and reduce the reliance of current retrievals on tightly coupled microphysical assumptions (Wellmann et al., 2020; Besic et al., 2016). In particular, machine learning (ML) algorithms have become popular tools in precipitation research as they greatly benefit from the large databases of available remote sensing observations to extract complex, emergent relationships between multiple climate variable inputs (Reichstein et al., 2019; Chen et al., 2020a; Chase et al., 2021). Further, new satellite launches combined with continued monitoring from current spaceborne and surface radar missions has led to an explosion in the growth of available remote sensing data (on the order of hundreds of terabytes per day), which can be used as extensive training datasets for these models (Agapiou, 2017; Guo et al., 2017).

1.1.2 Machine learning retrievals

Machine learning attempts to simulate human intelligence by learning from a set of input contexts without an explicit set of instructions (El Naqa and Murphy, 2015). The resulting nonlinear models created through this process are capable of learning complex relationships without an understanding of the underlying physical properties of the system (Burke et al., 2020; Li et al., 2020; Mao and Sorteberg, 2020). In general, ML is defined as a collection of empirical techniques

which fit a series of internal model parameters using a training set of observations to optimize some predefined loss (or cost) function (Chase et al., 2022). The size and complexity of these models have evolved over time, facilitating the ability of increasingly sophisticated model architectures to discover previously unknown, intricate structures in large databases, and to uncover novel data representations through multiple levels of abstraction (LeCun et al., 2015).

Machine learning in the Geosciences has become increasingly popular over the last decade (Fig. 1.8), yet the origins of this relationship traces back to nearly 60 years ago (Davenport and Diffenbaugh, 2021). Decision trees, Markov chains and k-means clustering have been widely used in the Geosciences since the mid-1960s for a variety of environmental safety and energy-related applications (Krumbein and Dacey, 1969; Newendorp, 1976). Only recently has the incredible quantity of remote sensing data combined with technological advancements in cloud computing resources (e.g. Google Cloud, Amazon Web Systems, Microsoft Azure, Graphcore), and powerful machine learning libraries (e.g. scikit-learn, Keras, PyTorch, Tensorflow) allowed for ML-based algorithms to truly flourish in this field (Zhang et al., 2016; Shen, 2018; Karpatne et al., 2019).

Unlike the OE snowfall retrieval methods introduced in Section 1.1.1, as empirical models, ML-based retrievals are not constrained by a priori assumptions of underlying hydrometeor microphysics. This does not imply that ML-based models do not have their own set of statistical assumptions about the training data (conditions of independence, normality, multicollinearity, autocorrelation and equal variance must still be considered), however these assumptions are generally less restrictive than those required for OE retrievals (Karniadakis et al., 2021). While less sophisticated statistical models like linear regression (LR) are often favoured over more complex ML methods due to their simplicity and interoperability, the nonlinear nature of ML allows for

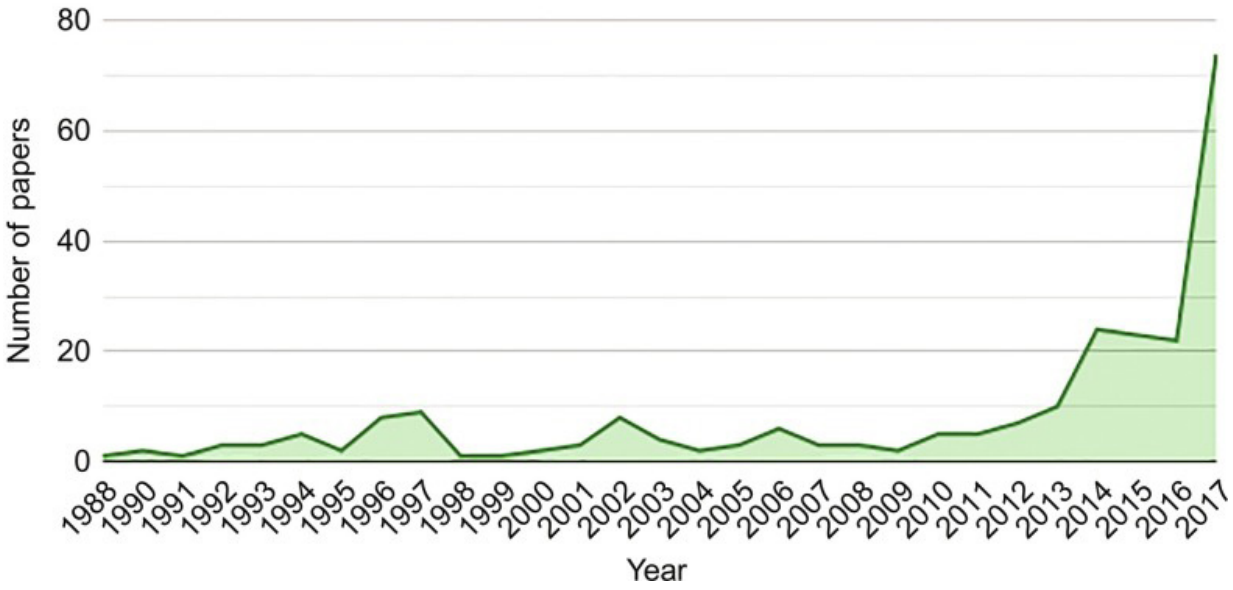

Figure 1.8: **Bibliometric analysis of 242 ML papers published in the Geosciences over 30 years (sorted by publication year)** (Dramsch, 2020).

models to be developed which can capture increasingly complex, and desirable, behaviours in precipitation (i.e. seasonality and spatial variability); response signals which are typically lost when relying solely on linear methods (Holte, 1993; Maahn et al., 2020; King et al., 2020). This process of combining remote sensing data and ML has led to many new advancements in our understanding of global precipitation occurrence and intensity (Chen et al., 2020a; Shi et al., 2015; Ehsani et al., 2021; Adhikari et al., 2020b; Chase et al., 2021).

The catalogue of widely-available ML methods is vast, and selecting the optimal method for a project strongly depends on the problem context (Chase et al., 2022). The timeline in Figure 1.9 depicts a subset of popular ML algorithm milestones since the inception of the field in the 1950s. As the work described in this thesis is focused on predictions of precipitation using radar data and collocated in situ measurements, a supervised ML paradigm is applied wherein the training

data is partitioned into a set of predictors (i.e. input covariates) and a response (i.e. what the model is learning to predict). For continuous variables like precipitation, thinking of this as a classification problem would be inappropriate, as there are an infinite number of possible classes to be predicted. Therefore, regression-based ML is performed herein to obtain a numerical output representing a surface precipitation rate (Breiman, 2017).

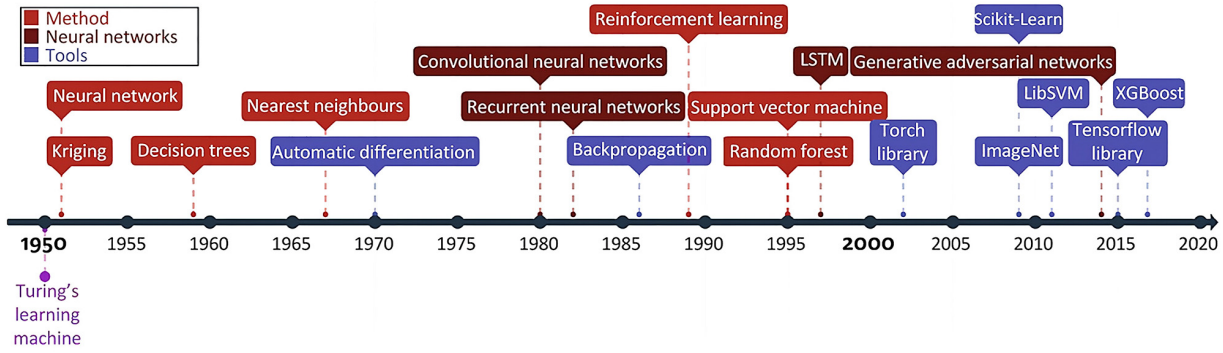

Figure 1.9: **Timeline of notable milestone in machine learning (1960-2020)** (Dramsch, 2019).

There exist many supervised, regression-based ML models of varying complexity which could be appropriate for this type of problem. Support vector machines (SVM), for instance, are a mathematically similar method to LR which defines a linear boundary in the form of Eq. 1.2

$$\hat{y} = w^T * x + b \tag{1.2}$$

to perform predictions with w as the weight vector, x containing the input covariates, b as the bias term and $\hat{y}$ as the predicted output (Vapnik, 1963). Note the similarity between Eq. 1.2 and LR (i.e. $y = m * x + b$), along with the fact that this general structure of weights and biases is also a key component of more sophisticated ML model like neural networks (NN) (Huang et al., 2006). However, unlike LR, the SVM is able to model much more complex, nonlinear

relationships in the data. Decision tree (DT) methods are another common ML technique which operate by making a cascade of choices through a flow chart-like data structure (Fig. 1.10) (Ali et al., 2012). Combining multiple DTs together forms another tree-based ML model known as a random forest (RF). Constructed from many individual trees, the RF is a much more robust model than any single DT. At the core of the RF algorithm is the concept of bagging (or bootstrapped aggregation). Bagging is an ensemble learning process which randomly samples the training data with replacement (i.e. bootstrapping) to generate multiple independently trained DTs which are then averaged together (i.e. aggregation) to yield a less variable and more robust estimate of the response variable being predicted (Breiman, 2001; Prasad et al., 2006).

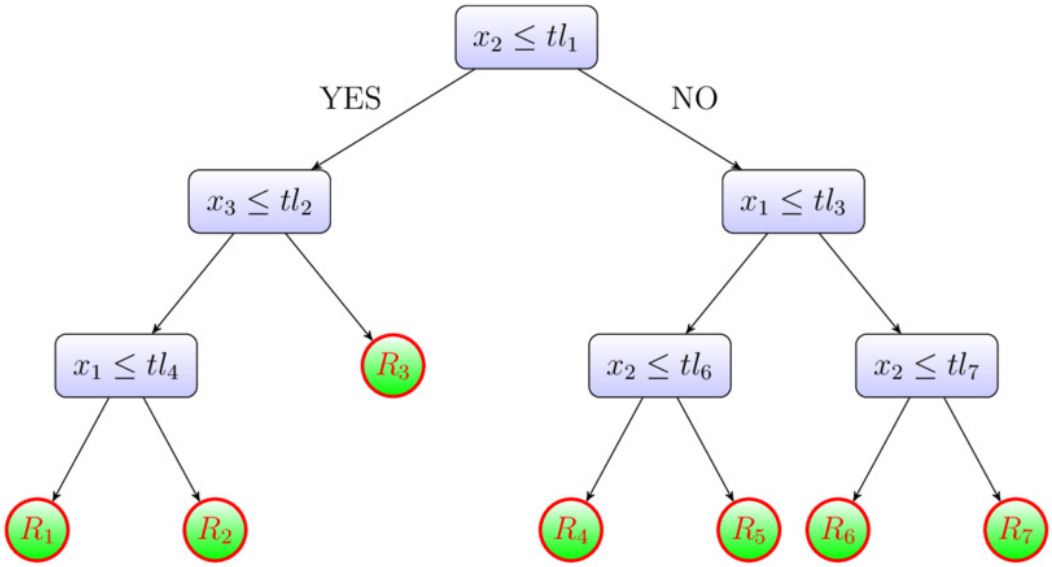

Figure 1.10: **Schematic of an example decision tree with three predictors (x_1,x_2 and x_3) and seven leaves ($R_1 \cdots R_7$) with a maximum depth of 3** (Akbari et al., 2021).

On the other end of the complexity spectrum to LR methods are neural networks. These models are composed of a network of interconnected processing nodes (i.e. neurons), each of

which receives an input signal, multiplies that signal by some weight w, adds a bias b term and applies an activation function (e.g. linear, sigmoid, tanh) to produce a signal sent to the next connected neuron in the network (Schmidhuber, 2015). Network learning is performed by incrementally updating internal model weight and bias parameters through a stochastic gradient descent optimization scheme to minimize a predefined loss function and in turn, teach the NN to behave in a desirable manner (e.g. driving a car, identifying a cat, or predicting snowfall) (Amari, 1993). The overall complexity of these networks is highly dependent on whether shallow learning (one neuron layer) or deep learning (many neuron layers) is being performed (Uzair and Jamil, 2020).

ML models like the RF have been used in King et al. (2020) and Zahmatkesh et al. (2019) for bias correcting precipitation data in NA (improvements in RMSE of 86%), in Wolfe and Snider (2012) for predicting precipitation extremes (with up to 90% accuracy), and in Das et al. (2017) for general precipitation retrievals. Neural network architectures and deep learning methods have also shown promise in precipitation retrieval algorithms by taking advantage of the vast database of available remote sensing observations. For instance, the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) is a daily, gridded precipitation product (at 0.25 degrees) developed by the University of California Irvine Center for Hydrometeorology and Remote Sensing (CHRS) which uses GridSat-B1 infrared satellite data as an input to an NN in order to produce high resolution long-term estimates of precipitation (Ashouri et al., 2015). Recently, the use of convolutional neural networks (CNNs) have also demonstrated effectiveness in precipitation estimation by identifying and extracting relevant features from the remote sensing input data to more accurately identify precipitating storm layers (Xue et al., 2021; Pan et al., 2019).

While ML-based models have contributed to advancements in the prediction of rain over the past decade, current literature related to snowfall prediction using radar remains limited due to the challenging variable density and non-spherical shape of frozen hydrometeors (Liu, 2020). Further, the majority of ML-based precipitation models assimilate passive remote sensing data (i.e. horizontally-structured data) and do not consider the vertical profile to capture different precipitating cloud layers. Recent advancements using surface radar data as training inputs to machine learning models during periods of snowfall have further demonstrated the important role of ML in precipitation retrievals (Chen et al., 2020a). However, accurate precipitation phase discrimination remains a substantial challenge (Casellas Masana, 2022). The strengths and limitations of vertically-structured snowfall retrievals will be examined in detail in chapters 3 and 4 using a variety of ML-based approaches and remote sensing datasets.

1.2 Research Motivation

As Arctic air temperatures continue to rise, investigations of new methods for snow estimation are becoming critically important topics with far-reaching impacts to the wider Geosciences. As previously discussed, the micro and macro-scale impacts of rapid nonlinear changes in SWE quantities across the Arctic are anticipated to produce knock-on effects to global water-energy budgets which will impact billions of people in the coming decades. The ability to mitigate the resulting damages from these changes begins with high quality, observationally-derived estimates of snow accumulation. While southern hemisphere (SH) snowfall studies are also important, improved satellite data sampling over the NH combined with better spatiotemporal coverage from terrestrial weather stations and in situ measurement networks concentrates our research in this

work on the NH. Therefore, the focus of this thesis is to evaluate the accuracy of current NH SWE products using remote sensing techniques, and to identify new methods for low uncertainty retrievals of snowfall using machine learning.

The motivation behind chapter 2 focuses on the former of these two issues. The large spread in SWE from state-of-the-art gridded snow products over much of the Arctic suggests high uncertainty in any individual product's general accuracy (Mudryk et al., 2015). Reanalysis and model predictions throughout this region heavily rely on in situ observational networks for constraining their estimates, however the sparsely distributed Arctic in situ network contains large, unobserved gaps (Mekis et al., 2018). Spaceborne, remote sensing observations from the NASA CloudSat-CPR's 2C-SNOW-PROFILE provide an observation-based set of snowfall estimates that can be used for independent comparative analyses. CloudSat's polar orbit results in high-frequency sampling (up to daily sampling at high latitudes) and has demonstrated good skill in estimating surface accumulation from snowfall measurements at monthly timescales (King and Fletcher, 2020). This chapter (King and Fletcher, 2021) therefore seeks to develop a methodology which can be used to objectively flag periods and regions within gridded SWE products which display poor data quality in order to enhance our understanding of snow accumulation across Arctic regions.

Following our evaluation of gridded SWE products using remote sensing, chapter 3 introduces a nascent ML-based retrieval to compare against traditional reflectivity (Z_e) snowfall (S) relationships. $Z_e - S$ power law relations (i.e. $Z_e = a \times S^b$) have been used for decades to estimate surface snowfall from atmospheric backscatter (Hiley et al., 2011). Deriving appropriate values for a and b in this formula is challenging, as these variables are related to various hydrometer microphysical properties (i.e. shape, size, distribution, fallspeed and density) and therefore many

unique relationships exist (e.g. Fig. 1.11). The specific combination of microphysical properties used in the derivation of the $Z_e - S$ relationship results in a power law that is non-robust to new regional climates (i.e. where these properties may vary). In chapter 3 (King et al., 2022a), we examine the role of machine learning methods (which are not beholden to the same sets of microphysical assumptions) in retrieving surface snowfall estimates when trained using remote sensing data.

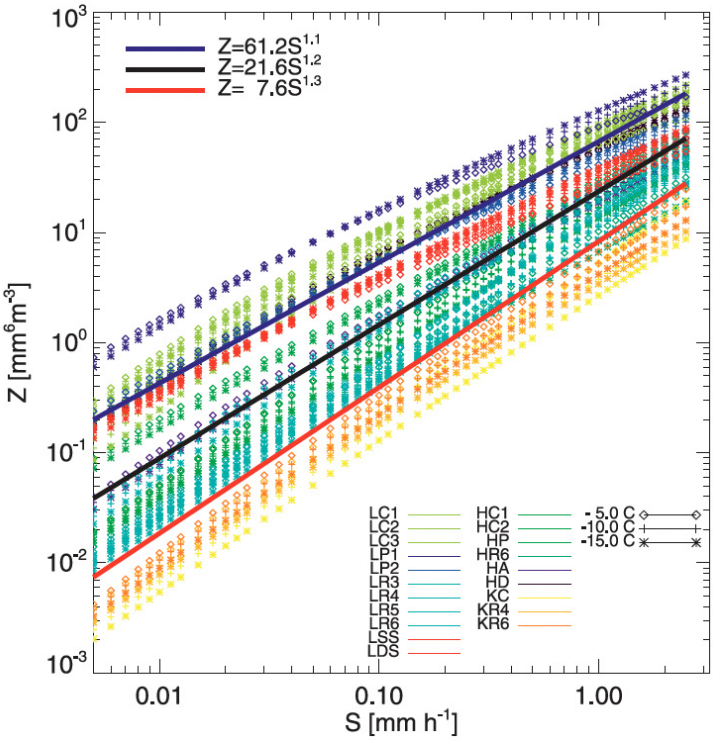

Figure 1.11: **An ensemble of $Z_e - S$ relationships from different studies**. Each color represents a unique $Z_e - S$ relation for non-spherical ice particles from Hong (2007), Kim et al. (2007) and Liu (2008b) at multiple temperature thresholds (Hiley et al., 2011).

Chapter 4 (King et al., 2022b) builds on the ML-based retrieval from chapter 3, expanding the scope of the method to use additional data from multiple sites. In order to combat the

non-robust nature of radar-based retrievals, it is important to develop the model in a manner which can handle different precipitation regimes at various locations and times of the year. Deep learning methods have demonstrated skill in extracting relevant features from inputs that are not immediately recognizable by humans (Jogin et al., 2018). Further, deep learning excels when working on problems with large quantities of data (Chen and Lin, 2014). This final chapter therefore attempts to produce a phase-agnostic precipitation retrieval using surface radar data from multiple locations, and compares the accuracy of said model against traditional $Z_e - S/R$ relationships as well as the RF model from chapter 3.

1.3 Research Objectives

The goal of this work is to improve our understanding of NH snow accumulation by evaluating the robustness and uncertainty of SWE products and snowfall retrieval methods using remote sensing observations and machine learning. This work also highlights the important role of machine learning in remote sensing retrievals of snowfall for future spaceborne precipitation measurement missions. We therefore aim to answer the following research questions over each subsequent chapter:

1. How can remote sensing observations be used to quantify and reduce uncertainties in current snow water equivalent reanalysis products?

2. What surface snowfall retrieval accuracy can be achieved using a supervised, machine learning algorithm when trained on surface radar observations?

3. Can we generalize the ML model from the previous question to new regional climates using additional inputs and deep learning techniques? What additional information can this model supply with respect to vertical column feature importances?

1.4 Thesis Structure

This thesis is separated into 5 primary chapters. Chapter 1 (the introduction) provides a literature review of the current state of the snowfall research in Arctic regions. Descriptions of current observational methods, climate model estimates and remote sensing techniques are also provided, followed by a review of the increasingly relevant role of machine learning in this field in recent years. The body of this thesis is contained within chapters 2-4, each comprised of a published manuscript produced over the course of my PhD. Chapter 2 (King and Fletcher, 2021) defines a novel approach for constraining current gridded SWE products using CloudSat-CPR snowfall observations. Chapter 3 (King et al., 2022a) introduces a nascent method of snowfall retrieval using a random forest model trained on surface radar data (X-band) from a site in southern Ontario, Canada. Finally, in chapter 4 (King et al., 2022b) we develop a deep learning, phase-agnostic precipitation retrieval from surface radar (K-band) at multiple sites across the northern hemisphere. Thesis findings are summarized in chapter 5 (conclusions), followed by a discussion regarding the main limitations encountered in this work and suggestions for relevant future studies.

Chapter 2

Using CloudSat-derived snow accumulation estimates to constrain gridded snow water equivalent products

2.1 Overview

Changes in the quantity of terrestrial Arctic snow have far-reaching implications for the global water-energy budget, ecosystem development and cold region flooding. Snow water equivalent (SWE) is a useful metric for monitoring these changes, however only a sparse observing network is available in the Arctic. Space-based remote sensing offers the potential to fill these measurement gaps, and the CloudSat cloud profiling radar (CPR) has been shown to provide high-quality estimates of surface snowfall rates across the Arctic. We propose a novel method to leverage

monthly information from CloudSat to identify data quality issues in the Blended-4 gridded SWE product. A regression of estimated monthly mean snow accumulation between CloudSat and B4 provides a confidence interval to objectively flag individual months in B4 with data quality issues. Applying this method to a case study in the Canadian high Arctic identifies an unphysical January melt event in B4, which is traced back to a measurement error in the assimilated station observations. We generalize this technique to a one-degree grid and find a total of 4885 cases in B4 with low data quality. We find that the low-quality SWE product values are not random errors, and that they introduce a systematic bias of -3.1 mm on B4's estimates of mean Arctic SWE. CloudSat is uniquely positioned as one of the few observational datasets for observing snowfall at high latitudes, and when combined with the methodology developed here, its estimates can be used to further enhance the accuracy of current gridded SWE products.

2.2 Introduction

Long term changes in Arctic snow mass has major impacts to the world's water and energy budget, cold region flooding, and ecosystem development (Berghuijs et al., 2019; Déry and Brown, 2007; Peng et al., 2010). As global temperatures continue to warm, the quantity of Arctic snow is expected to change in a significant, non-linear manner (Lemke et al., 2007; Vaughan et al., 2013). Understanding these changes is of critical importance, as accurate estimates of snow accumulation help us better prepare for and mitigate against the consequences of a warming climate (Hosaka et al., 2005). Snow water equivalent (SWE) is the amount of water stored in a snowpack when it is completely and instantaneously melted, and is a useful metric for monitoring changes in accumulating snowfall (Brown, 2000).

Traditional techniques for measuring SWE include automated in situ observation gauges, and manual measurements which are performed by a human observer (Doesken and Robinson, 2009; Authorities, 1985). These methods provide high accuracy point measurements of SWE, however, due to the time and cost of implementing and maintaining these systems, or performing these measurements by hand, there often exist large gaps between measurement sites (Tait, 1998; Derksen and Brown, 2012; Liston, 2004). Issues in maintaining active measurement sites are further exasperated by the vast size, remote nature and cold climate that is typical of Arctic regions. For instance, only approximately 1% of all active weather stations in Canada are located poleward of 70° N (Mekis et al., 2018). Additionally, the number of automated snowfall gauges in Canada has been declining for the past 30 years, with approximately 75% fewer measurement locations today than were available in the early 1990s (Mekis and Vincent, 2011). Filling these measurement gaps is therefore of critical importance in creating a comprehensive understanding of changes to Arctic snow accumulation (Pulliainen et al., 2020).

Numerical weather and hydrologic forecasts, along with data assimilation techniques, are powerful tools which can be used to estimate SWE in data sparse regions (Rienecker et al., 2011; Takala et al., 2011; Balsamo et al., 2015). Gridded SWE products have complete spatio-temporal coverage and have been shown to demonstrate good agreement with in situ measurements at global scales (Kulie et al., 2020a; Zahmatkesh et al., 2019; Lindsay et al., 2014). Observation based reanalyses like the European Center for Medium-Range Weather Forecasts (ECMWF) Re-Analysis version 5 (ERA5) and the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) from NASA Gelaro et al. (2017), are two common examples of gridded products that output a suite of surface climate parameters like snow depth (SD) and SWE (Dee et al., 2011; Rienecker et al., 2011). However, all of these gridded SWE products are

derived from numerical models which contain inaccuracies stemming from internal numerical model estimates, uncertainties in prescribed forcing data, or due to issues in the assimilated observational datasets (Teutschbein and Seibert, 2012; Broxton et al., 2016).

Using an ensemble of gridded datasets is one technique for addressing biases in individual component estimates and improving overall product accuracy (Mudryk et al., 2015; Guan et al., 2013). The process of blending together multiple datasets can be performed in a variety of ways, however in this work we consider it as taking the unweighted multidataset mean of all component products to create a new *blended* dataset. Blending has been used in previous literature by Mudryk et al. (2015) and demonstrated effectiveness at reducing the multidataset spread of SWE. Monthly normal SWE values from the blended dataset compare well with observed SWE distributions across the Northern Hemisphere (NH) from 60° N to 82° N (Brown and Mote, 2009). Additionally, combining information from multiple datasets through the blending process allows us to identify where one product may be negatively influencing the final product through comparisons between individual component datasets (Foster et al., 2011; Orsolini et al., 2013; Mudryk et al., 2015).

Another source of observational snowfall data comes from remote sensing missions. With their global coverage, satellites are able to overcome many of the traditional issues experienced in observing remote Arctic regions. SWE products like GlobSnow (Pulliainen et al., 2020) and other climate change initiative (CCI) products focusing on temperature, wind speed and soil moisture, all take advantage of remote sensing instruments which provide additional data to refine product estimates (Jiménez-Muñoz and Sobrino, 2003; Witschas et al., 2020; Kerr et al., 2010). The Cloud Profiling Radar (CPR) instrument installed on the NASA CloudSat satellite has been shown to provide high quality snowfall estimates across Arctic regions (Hiley et al., 2010; King

and Fletcher, 2020; Palerme et al., 2014; Milani et al., 2018; Hudak et al., 2008; Bennartz et al., 2019). As a consequence of this high accuracy, we therefore expect good agreement during cold seasons (with temperatures below 0° Celsius) between the average CloudSat snow accumulation estimate over two consecutive months and the monthly change in snow on ground in the gridded SWE products over that same period. However, a complete CloudSat gridded snowfall product has yet to be used to evaluate snow accumulation in other gridded SWE products at NH high latitudes.

The primary objectives of this work are to:

1. Characterize the relationship between CloudSat snow accumulation estimates and a gridded, blended SWE product across the Arctic;
2. Use CloudSat accumulation estimates to identify periods of low data quality in the gridded SWE products; and
3. Quantify the bias in climatological SWE induced by low quality observations on mean SWE in the gridded products

2.3 Data and Methods

2.3.1 Gridded SWE Products

The Blended-4 (B4) product is a daily, 0.5° SWE dataset spanning 2006 to 2015. B4 is an updated version (extending beyond 2010) of the Blended-5 SWE product, which was developed

by the Canadian Sea Ice and Snow Evolution (CanSISE) network (Mudryk et al., 2015). The B4 dataset is produced by taking the multi-dataset mean of four independent SWE products across the NH. These component datasets include 1) the Canadian Meteorological Center (CMC) gridded snow product; 2) the CROCUS snow model; 3) the GlobSnow gridded SWE product; and 4) The Modern-Era Retrospective analysis for Research and Applications (version 2) (MERRA-2) gridded reanalysis product. Each product's estimates of SWE are derived independently using model representations of snow accumulation and melt of variable sophistication, ranging from simple temperature index models based on overlying air temperatures (CMC) to more complex multi-layer, physically-based representations of snowpack evolution (CROCUS). The details of each SWE product are summarized in Table 2.1, and briefly described below.

The CMC product is a daily, $1/3°$ NH gridded SWE dataset produced in Brown et al. (2003) and updated to include data until the end of 2015. The CMC data combines a simple temperature index snow scheme driven by 6-hourly precipitation and temperature data from the European Center for Medium-Range Weather Forecasts (ECMWF), along with snow density estimates derived by Sturm et al. (2010), and assimilated daily snow depth observations from approximately 8000 United States and Canadian climate stations to provide estimates of surface SWE. Optimal interpolation (OI) is used to combine in situ observations with background model estimates of SWE.

CROCUS is an energy-balance snow model developed and maintained by the National Centre for Meteorological Research (NCMR), which was applied to the terrestrial NH at 1 km resolution (Brun et al., 1989, 2013). Daily estimates of SWE on ground are driven by 2 meter air temperature and humidity, surface wind velocity, shortwave and longwave radiation and precipitation estimates from ERA-Interim. CROCUS divides the snow-pack into multiple layers and represents internal

Table 2.1: Gridded SWE product summary table for the datasets used in this study, along with area-weighted summary statistics of accumulated snow between CloudSat and each of the B4 component products (units of mm SWE).

Product	Snow Scheme	Resolution	Mean Bias	RMSE	Source
CMC	Simple	$1/3° \times 1/3°$	8.82	29.61	(Brown et al., 2003)
$CROCUS$	Complex	$1° \times 1°$	9.14	37.75	(Brun et al., 1989)
$GlobSnow$	Satellite & In Situ	25 km	9.23	31.81	(Takala et al., 2011)
$MERRA-2$	Intermediate	$1/2° \times 2/3°$	9.33	39.57	(Gelaro et al., 2017)

snowpack processes like settling, water transfer and heat exchange to simulate snowpack evolution in a physically-based manner (Brun et al., 2013). The CROCUS snow model is run as a land surface component in the Interactions between Soil, Biosphere, and Atmosphere (ISBA) land model to generate estimates of SWE (Mudryk et al., 2015).

GlobSnow is a 25 km gridded SWE product developed by European Space Agency (ESA) which provides daily estimates of SWE on ground through a combination of both in situ and satellite brightness temperature observations (Takala et al., 2011). A single layer snow emission model is forced with in situ snow depth measurements to simulate passive microwave brightness temperature estimates across the NH (Pulliainen et al., 1999). Estimates of snow grain size are then generated by minimizing the disagreement between simulated and observed brightness temperatures to derive a contiguous SWE product across the NH.

MERRA-2 is a reanalysis product developed by NASA's Global Modeling and Assimilation Office that provides daily SWE estimates at $1/2° \times 2/3°$ resolution (Gelaro et al., 2017). MERRA-2 uses a finite-volume dynamical core and cube-sphere horizontal discretization of the grid in its forecast model to distribute estimates of its surface variable suite (Putman and Lin, 2007). 3-Dimensional variational (3D-VAR) data assimilation is used to minimize error between

the forward model and observational data, to incorporate in situ climate information along with space-based aerosol data. SWE estimates are then generated using this assimilated data within the Catchment land surface model, which includes a complex multi-layer snow scheme to represent changes to snowpack accumulation and ablation (Mudryk et al., 2015; Koster et al., 2000).

Each of the above products is first regridded to a common 0.5° NH grid using bilinear interpolation. A land fraction mask provided by MERRA is also applied to our data in order to select only terrestrial SWE. The B4 product is then generated by taking the ensemble average of all component datasets. More formally, this is written as the unweighted arithmetic mean of the four products across all grid cells:

$$\overline{B} = \frac{1}{n} \sum_{i=0}^{4} g_i * f \tag{2.1}$$

where $\overline{B}$ is the derived blended dataset, f is the land fraction mask, and each g_i represents a component gridded product.

2.3.2 CloudSat-CPR surface snow estimates

The NASA satellite CloudSat was launched in 2006 to advance our understanding of cloud macro and microphysical properties at global scales (Stephens et al., 2002; Kulie et al., 2020a). CloudSat is equipped with a 94-GHz nadir-looking Cloud Profiling Radar (CPR) instrument with 500-m vertical range gate resolution that observes clouds from the Earth's surface up to approximately 30 km into the lower stratosphere (Liu, 2008b; Stephens et al., 2008). The CPR allows CloudSat to view the interior structure of clouds by measuring the power backscatter reflected from particles within the cloud as a function of distance from the radar instrument in

orbit (Stephens, 2017). These reflectivity profiles provide information on cloud type, shape and the presence of hydrometeors within the cloud, along with corresponding internal precipitation rate estimates (Kulie et al., 2020a). CloudSat also demonstrates high accuracy in determining precipitation phase to differentiate between rain, mixed-phase and snowfall events (Kodamana and Fletcher, 2021). CloudSat has a 16-day repeat cycle with granule tracks that extend to an 82° N/S orbital maxima which converge towards the poles. As a remote sensing option, CloudSat does not suffer from traditional challenges that ground-based measurements encounter in measuring snowfall across such a vast, cold and remote locale, and therefore provides a unique perspective towards estimating accumulated SWE at global scales.

Reflectivity backscatter is derived using a combination of radar range information, hydrometeor positions along the path of the beam and retrieval response time for a set of assumed hydrometeor properties including particle size, shape and fallspeed (Matrosov, 2007; Liu, 2008a). As a result of the high frequency 94 GHz (3.19 mm) radar wavelength being similar to the length of the hydrometeor particles being measured, single scattering properties of snowflakes do not follow traditional Rayleigh scattering approximations and instead align more closely with Mie scattering (Battaglia et al., 2007; Haynes et al., 2009; Stephens et al., 2002). This W-band frequency is therefore well-suited for light intensity snowfall (which is typical in Arctic regions), but begins to attenuate during high intensity precipitation events (Hudak et al., 2008). Optimal estimations of snowfall rates are obtained by minimizing a cost function between estimated and observed a priori assumptions about particle size, shape, fallspeed and other snow microphysical properties (Rodgers, 2000). A forward model: $y = F(x, \tilde{b}) + \epsilon$ is constructed which approximates the physical relationship between an observed reflectivity profile y to a set of unknown snow microphysical properties x, additional potentially unknown supporting parameters $\tilde{b}$, and measurement/forward

model uncertainty ϵ (Wood and L'Ecuyer, 2021; Wood et al., 2015, 2014; Stephens, 2017). CPR snowfall rate estimates within the lowest precipitating bin are then extrapolated to the surface below and packaged into the 2C-SNOW-PROFILE (version 5) product, which is what is used in calculating snow accumulation for the remainder of this study.

Individual radar profiles from CloudSat have also shown in work by Duffy and Bennartz (2018) to exhibit uncertainties in snowfall rates of upwards of 200%. However, aggregation of many retrievals has demonstrated skill in reducing error in snowfall rate estimates (Hiley et al., 2010; Milani et al., 2018; Palerme et al., 2014). In our previous study King and Fletcher (2020) on the validation of CloudSat snowfall estimates over the Canadian Arctic Archipelago (CAA), we aggregated individual CloudSat retrievals into overpasses within a grid box. By averaging many of these overpasses together over the course of a month, we found that CloudSat typically displays good agreement (RMSE below 10 mm SWE and correlations above 0.5) with in situ measurements of accumulated SWE when CPR retrievals are gridded at 1° resolution at high latitude locations. Generalizing this technique to a hemispheric scale allows us to generate a monthly, terrestrial gridded SWE accumulation product (Fig. 2.1.b) which can then be used in comparisons with other gridded SWE datasets.

Duffy et al. (2021) found that when gridded at sufficiently large scales, CloudSat exhibits high correlations and uncertainties below 50% for much of the NH when compared with B4. Generalizing this aggregation technique to a wider NH grid from 60° to 82° N produces a spatial map of monthly snowfall accumulation and uncertainty from January 2007 to December 2015 (Fig. 2.1.b and Fig. 2.1.c). Figure 2.1.d also shows a low average coefficient of variation ($\bar{CV} = 1.27$) across the study region, with the lowest dataset variability with respect to the mean over the CAA, central Eurasia and Eastern Siberia.

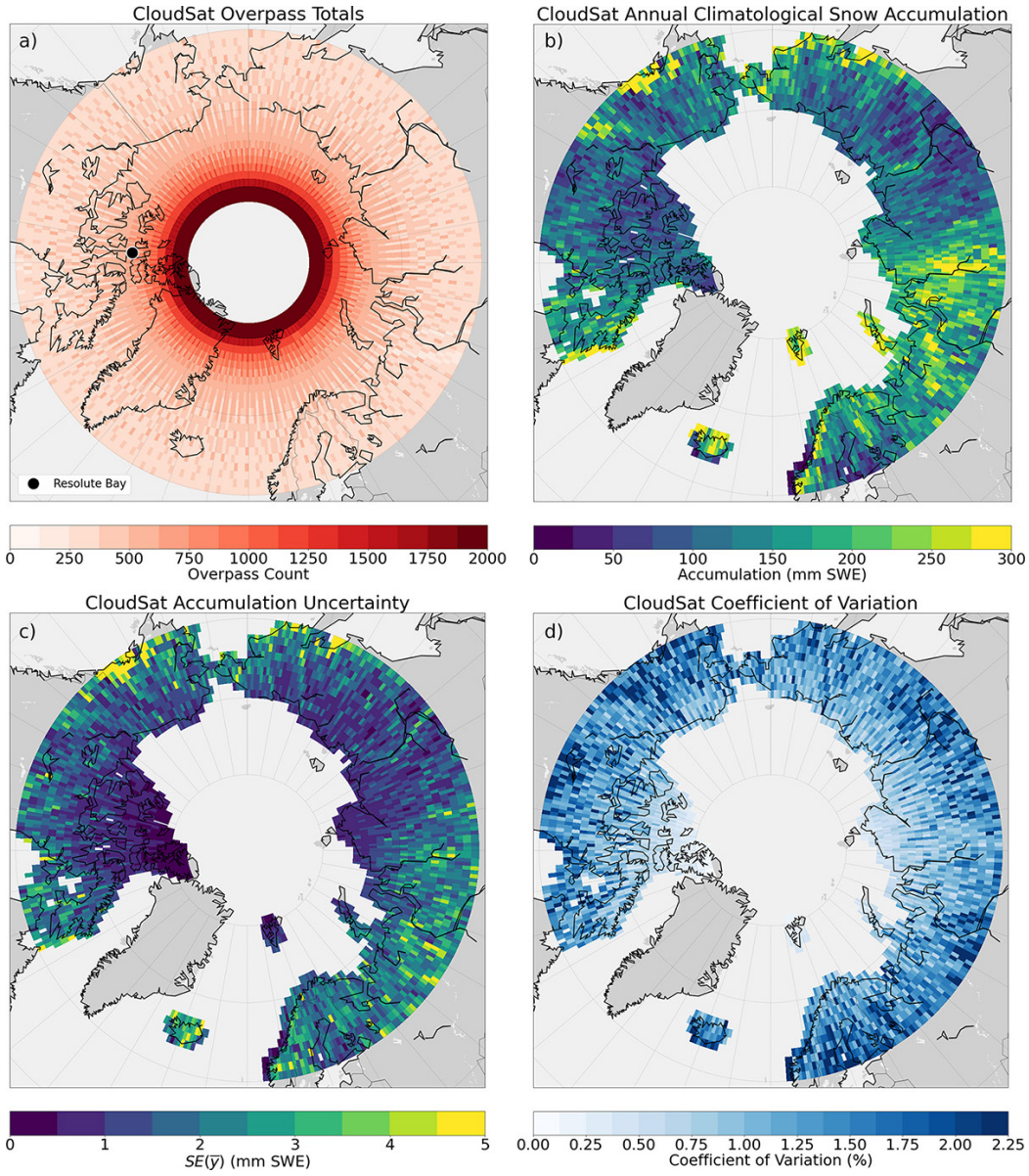

Figure 2.1: **Gridded CloudSat sampling frequency and snow accumulation estimates and uncertainties**. **a)** Distribution of CloudSat overpasses from 60° N to 82° N, gridded at 1° resolution from January 2007 to December 2015; **b)** gridded annual climatological snow accumulation derived from CloudSat; **c)** CloudSat annual snowfall sampling uncertainty; and **d)** CloudSat gridded coefficient of variation ($CV = \frac{\sigma}{\mu}$).

Repeated CloudSat battery failures are an additional source of uncertainty which may impact the capabilities of our outlier detection methodology due to variability in the CPR's observational frequency over a region. A critical battery failure occurred between 2011-2012, which disabled the CPR instrument for nearly a year and required the satellite to henceforth operate in daylight only operational (DO-Op) mode. As implied by its name, DO-Op mode requires CloudSat to have a direct line of sight to the sun for the CPR to have sufficient power to operate. The new DO-Op energy requirements have been shown to substantially reduce the CloudSat sample in the southern hemisphere, along with smaller reductions in its sample at high latitudes in the NH (Milani and Wood, 2021; Kulie et al., 2020a). An examination of the impact of repeated CloudSat battery failures on the CPR sample size across the Arctic was performed in our previous work King and Fletcher (2020), and the impacts of these battery failures were found to be negligible with respect to CloudSat's ability to accurately quantify snow accumulation in the region (additional details are discussed in Section 2.4.2).

2.3.3 Quantitative Methodology for Assessing Data Quality

Since the number of component products being blended together is typically small ($n = 4$ in the case of B4), data quality issues in the SWE estimates from one or more component datasets can significantly impact the accuracy of the resulting blended product. Understanding when, where and how often these data quality issues occur can therefore provide us with improved estimates of Arctic SWE. To identify and flag consecutive months (i.e. a month-pair) as a statistically inconsistent outlier in a gridded product, we examine the relationship between monthly changes in SWE estimated from two independent quantities: ΔSWE_B and $\overline{SWE}_C$. ΔSWE_B is derived

from the B4 product as the difference in monthly mean SWE on ground between two consecutive months, providing an estimate of effective changes in SWE due to accumulation, melt, and other ablative processes. $\overline{SWE}_C$ is calculated as the average snowfall rate from CloudSat over the two months (derived using techniques described in King and Fletcher (2020)) and, assuming a constant snowfall rate (in mm/hr), multiplied by the total number of hours over the two months to derive a corresponding estimate (in mm) of accumulated SWE on ground.

Since CloudSat provides instantaneous estimates of snowfall rates, it is therefore unable to provide information related to ablative processes like snow melt or sublimation. To minimize discrepancies in the quantification of SWE during periods of melt, we restrict our analysis to only months when the maximum daily 2 meter air temperature from ERA5 remains below 0° C for the entire month. We also examined the sensitivity to this temperature threshold by restricting the maximum monthly temperature to three additional values (-1° C, -3° C, and -5° C), under the assumption that a colder temperature threshold would result in a lower likelihood of including month-pairs with surface melt occurring. Our conclusions from this study were not significantly altered by using a different threshold (not shown), and so the 0° C threshold was selected as it provides the largest CloudSat sample size.

Further data preprocessing constraints are also applied to the set of ΔSWE_B and $\overline{SWE}_C$ month-pairs in order to mask cases with an insufficient sample of CloudSat observations. This step limits our comparisons to cases where we have a high level of confidence in CloudSat's accumulation estimates. Restrictions based on the number of available overpasses in a month (N_{month}) along with the total number of month-pairs in a grid cell (M_{cell}) were examined using a simple grid-search procedure. This technique revealed that limiting the analysis to cases where $N_{month} \geq 12$ and $N_{cell} \geq 3$ improves the accuracy of the method and produces a sufficiently large

sample over the majority of North America (NA) and Siberia.

Using estimates of accumulated SWE from $\overline{SWE}_C$ as a predictor and ΔSWE_B as the response variable for all months when the maximum temperature is below 0° C in an ordinary least squares (OLS) regression, we derive a relationship between snowfall and accumulation for each grid cell. The general regression model is described in Eq. 2.2 as:

$$\Delta \hat{SWE}_B = b_0 + b_1 \cdot \overline{SWE}_C \tag{2.2}$$

where the b coefficients are the respective model y-intercept and regression slope. Statistical assumptions required for OLS regression include linearity, normality and homoscedasticity for a set of independent grid cells, and all were found to meet acceptable criteria in our analysis. We also examine the leverage (h_i) and studentized residuals (t_i) of each month-pair (i) per grid cell to remove influential cases suffering from issues in the CloudSat sample. These cases can arise when CloudSat incorrectly estimates the average accumulation in a grid cell as a result of missing snowfall events that occur when CloudSat is not overhead. Typical thresholds of $h_i > 0.15$ and $t_i > 2$ standard deviations are used to identify and remove such extreme cases which would significantly negatively influence the accuracy of our regression model fit.

Applying a 95% prediction interval (PI) to the regression model allows us to identify specific month-pairs within each grid cell that lie outside the expected range of variation for individual month-pairs. We apply leave-one-out cross-validation for each month-pair in a grid cell to determine if it falls outside the PI when excluded from the construction of the regression model. Month-pairs which fall outside the PI are automatically flagged as statistically inconsistent outliers for further investigation.

2.4 Results

2.4.1 Resolute Bay Case Study

We begin by presenting a case study of our method at Resolute Bay; a high latitude (75° N) community on Cornwallis Island in the CAA (location shown in Figure 2.1.a). The selection of this site follows from the results of King and Fletcher (2020) who found strong agreement between CloudSat and in situ estimates of snow accumulation at this location. Resolute Bay is equipped with Environment and Climate Change Canada (ECCC) weather station instruments to record in situ snow data including SWE, which provides an additional independent observation of accumulation for comparison purposes. This case study provides a local-scale test for identifying statistically inconsistent periods in B4 and its component products before generalizing and applying this technique to the wider NH in Section 2.4.2.

The relationship between $\overline{SWE_C}$ and ΔSWE_B at Resolute Bay (Fig. 2.2) reveals a weak positive correlation ($r = 0.24$) between CloudSat and B4, with plenty of scatter introduced by uncertainties in B4 and CloudSat as discussed in more detail in Section 2.5. The vast majority of month-pairs (94%) reside within the 95% prediction interval, and three outlier month-pairs fall outside the interval. One of these outliers is highlighted in red and stands out as a case of interest as it is the only month-pair where B4 shows a loss in SWE on ground between two months (excluding this case, the correlation improves to $r = 0.384$). This outlier represents a wintertime melt event occurring between December 2007 and January 2008 when daily observed maximum temperatures at the weather station were below -20° C and wind speeds were negligible. A reduction in SWE of this magnitude, therefore, indicates a potential problem with the estimate of

ΔSWE_B for these months.

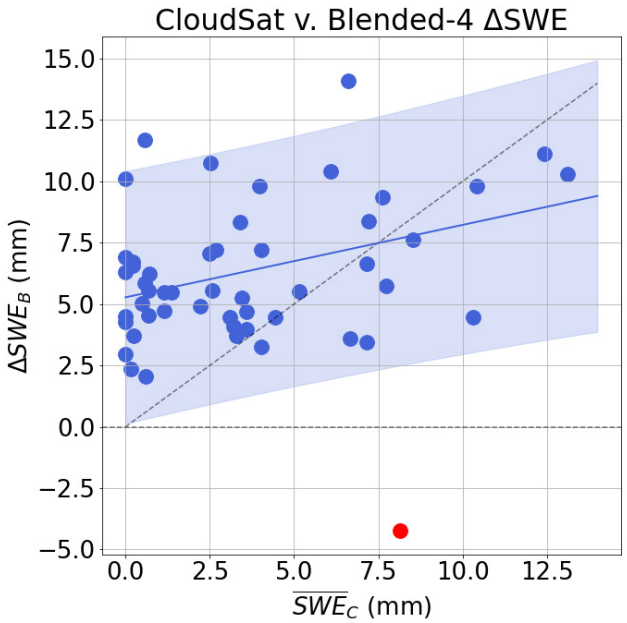

Figure 2.2: **CloudSat and Blended-4 monthly accumulated SWE scatter**. $\overline{SWE_C}$ snow accumulation vs. ΔSWE_B for all frozen months at Resolute Bay (case study highlighted in red).

Next, we examine the daily evolution of the B4 gridded product and its component datasets for this outlier event, compared to other years at the same time and location. SWE on ground generally increases as temperatures remain too cold for any significant ablation to occur and losses due to blowing snow are negligible (Fig. 2.3.a). However, for the 2007-2008 period, the blended product shows two large ablation events in early December, and late December/early Jan (Fig.2.3.a, black line). It is this second, larger melt event that is identified here by the outlier detection method in Fig.2.2.

Examining the B4 component datasets reveals GlobSnow as the product that is responsible for this unphysical loss in SWE during Dec-Jan 2007/08 (Fig. 2.3.b). GlobSnow is unique among

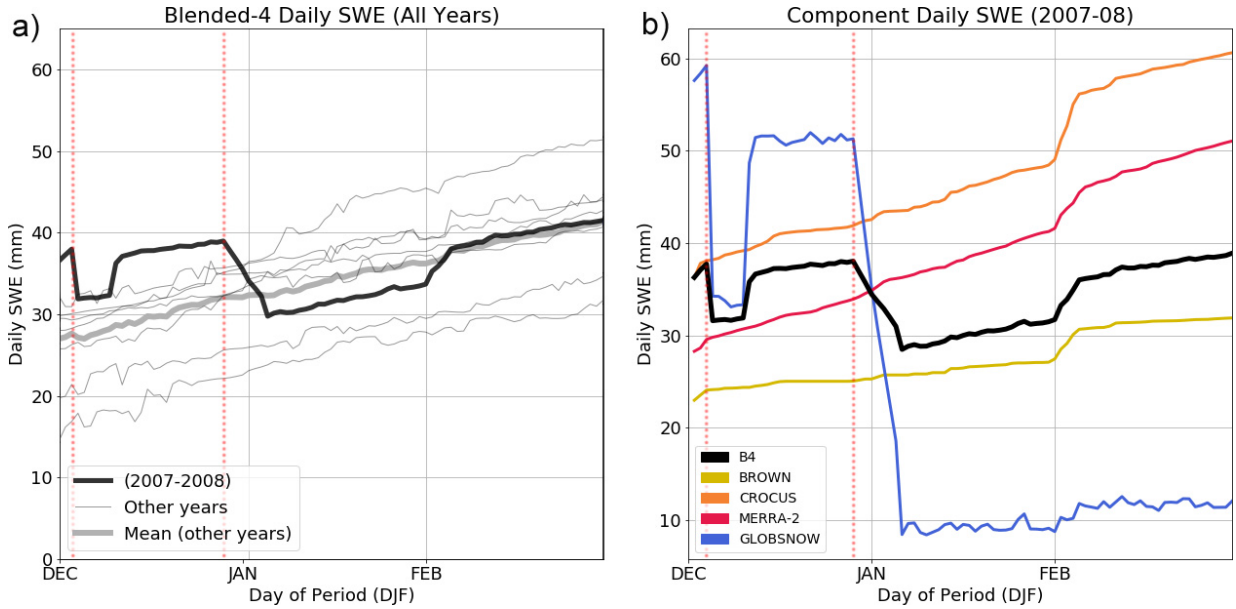

Figure 2.3: **Blended-4 melt event annual comparisons and product breakdown**. **a)** B4 daily SWE on ground for December through February for all years, with the 2007-2008 period of interest in black, and the average of all other periods in bold grey; and **b)** B4 gridded component breakdown of daily SWE on ground for December 2007 to February 2008, highlighting the impact of the GlobSnow product on B4. Dashed red lines indicate the start of major ablative events.

the products in B4 because it is satellite-based and assimilates surface snow depth data. An examination of the in situ snow depth measurements provided by ECCC for this station reveals similar instantaneous drops in snow depth observations during the last week of December 2007 (from 70 cm to 31 cm) and again on January 3 and January 4 (from 32 cm to 16 cm) which were then assimilated by GlobSnow, likely resulting in this unphysical cold season melt event. This single statistically inconsistent case was automatically flagged by our detection algorithm, which allowed us to trace back to the component product causing the issue, and finally the assimilated in situ surface observation.

2.4.2 Northern Hemisphere Evaluation

Applying this technique across the 60° N to 82° N NH grid allows us to examine where and when outliers occur even in locations without in situ measurements, and to quantify the frequency of the outliers in each product. Spatio-temporal relationships between ΔSWE_B and $\overline{SWE}_C$ provide important insight into when this technique has the highest accuracy in detecting month-pair outliers. As shown in Figure 2.2, positive agreement between the two products results in a better regression fit and associated PI with lower uncertainty, which consequently improves the ability of our method in detecting outlier cases. Correlations between ΔSWE_B and $\overline{SWE}_C$ for the entire NH domain are shown in Figure 2.4.a with an average correlation of $r = 0.29$ across all 2971 grid cells. It is encouraging that the general tendency is for these quantities to be positively correlated, indicating that at most locations, CloudSat-derived snowfall provides at least some constraint on B4 snow accumulation. However, there exists considerable spatial variation with the distribution of these correlations as shown in Figure 2.4.b, where 344 grid cells with correlations below zero

and 1612 grid cells above the average correlation of 0.29.

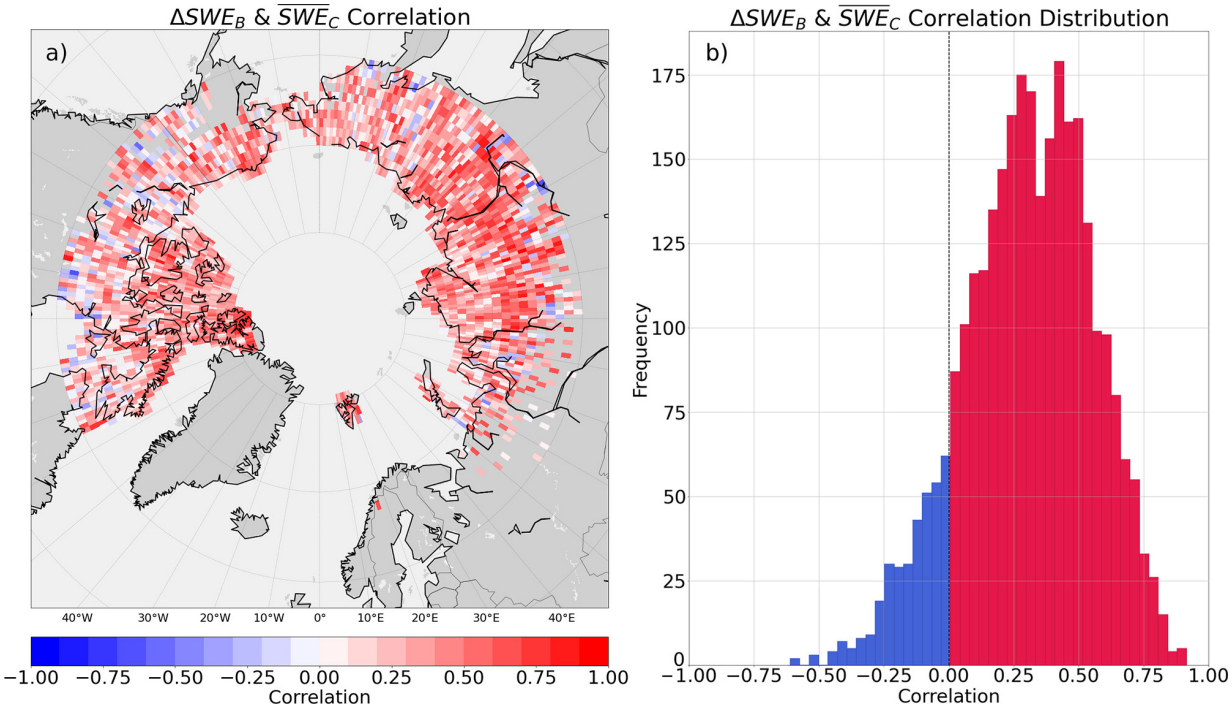

Figure 2.4: **Gridded correlations of CloudSat and B4 SWE accumulation**. **a)** ΔSWE_B and $\overline{SWE_C}$ frozen month-pair snow accumulation estimates correlated over 2007-2015 for each grid cell with data preprocessing applied using the methodology described in Section 2.3.3; and **b)** the distribution of correlations from **a)**.

While the majority of grid cells produce correlations above 0 and most of the correlations are strongly positive overall, the dataset agreement generally decreases as we move to lower latitudes as a result of CloudSat's polar orbit and reduced coverage at low latitudes (Fig. 2.4.a). A typical example from a grid cell at lower latitudes with a negative correlation is shown in Figure 2.5.a. The sign and magnitude of the relationship results primarily from a single erroneous $\overline{SWE_C}$ accumulation estimate (indicated by the red dot) which is the result of a small CloudSat sample of only $n = 3$ overpasses at this grid cell during this period. A small sample size can result in large

biases in the month-pair snow accumulation estimate if CloudSat is unable to accurately capture the presence and magnitude of localized storm systems at the grid cell level. This can lead to a less robust relationship with B4 accumulation at these latitudes. Furthermore, the dashed line in Figure 2.5.a shows what the OLS relationship would look like without the influence of the outlier. Examining the relationship between CloudSat's sampling uncertainty and the average correlation as a function of latitude (Fig. 2.5.b) shows that as CloudSat's sample size increases and standard error (SE) decreases, the correlation improves. This latitudinal component of the CloudSat estimate error is directly related to CloudSat's orbit and the reduction in its observational sample size as we move towards the equator (Fig. 2.5.c).

This connection between CloudSat's estimate uncertainty and sample size with respect to latitude has also been examined in previous work by King and Fletcher (2020) and Duffy et al. (2021), which came to similar conclusions that high latitude regions generally exhibit improved accuracy and lower uncertainty from CloudSat as a result of the larger sample size. These findings, along with the spatial distribution of grid cells in Figure 2.4, provide valuable information into identifying where we have a sufficient sample to use CloudSat as an observational constraint for comparisons at 1° resolution. Outlier totals were also calculated for pre and post 2011 for B4 to quantify the impacts of the primary battery failure, and we found outlier totals were similar in magnitude (n_{pre} = 2540 and n_{post} = 2231), with only a slight decrease after the battery failure. The spatial differences in outlier counts between these two periods appear randomly distributed across our study grid, and we argue that the CPR sample size remains sufficient when aggregated at 1° across high latitude regions to apply this data quality assessment and identify statistically inconsistent accumulation periods.

Another way to demonstrate the impact of the reduced sample size of CloudSat observations

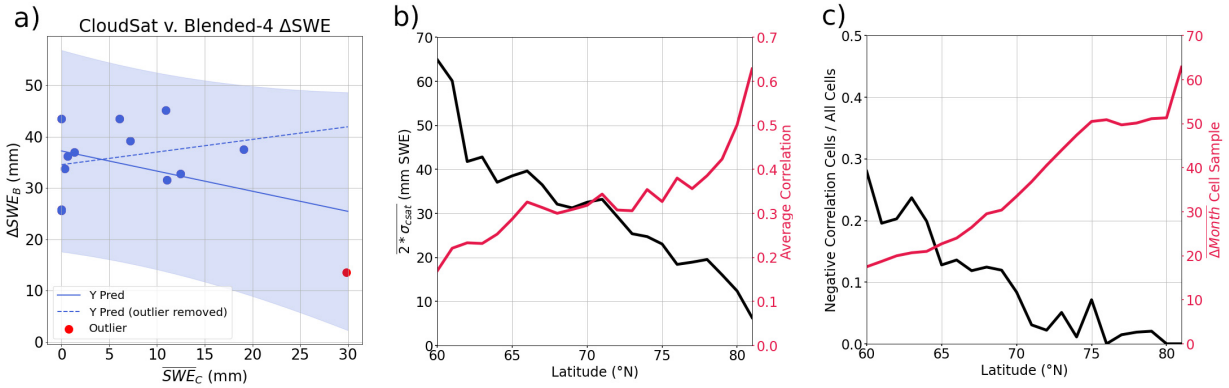

Figure 2.5: **Negative correlation case study and the relationships between latitude and CloudSat uncertainty**. **a)** Example of a negative correlation between ΔSWE_B and $\overline{SWE_C}$ at (63.5° N, 56.5° E) due to the influence of CloudSat sampling error and its impact on the regression slope; **b)** the latitudinal influence of CloudSat's sample size on the CloudSat standard error estimate and CloudSat's correlation with B4; and **c)** the relationship between the number of negatively correlated cells and the average number of frozen month-pairs per grid cell at that latitude band.

is to resample snowfall rates from a randomly selected high latitude terrestrial grid cell (in this example it is located at 81.5° N), by randomly choosing smaller sample sizes to simulate CloudSat's observational sample at lower latitudes. We first calculate the average snowfall rate from all available overpasses in the grid cell ($n = 3439$), which is considered as the true population mean snowfall rate ($\mu_s = 7.25$ mm). We then randomly resample the snowfall rate 1000 times using the average overpass sample size found at 60° N ($n = 225$) to generate a sampling distribution of estimated mean snowfall rates ($\overline{s}$). $\overline{s}$ and μ_s are then compared to yield a distribution of errors in $\overline{s}$ at that grid cell (Fig. 2.6.a).

Monthly mean accumulation estimates were found to converge to the population value when the CloudSat sample includes upwards of approximately 250 overpasses (Fig. 2.6.b), which is slightly larger than the smallest number of overpasses we would expect to obtain (on average) at

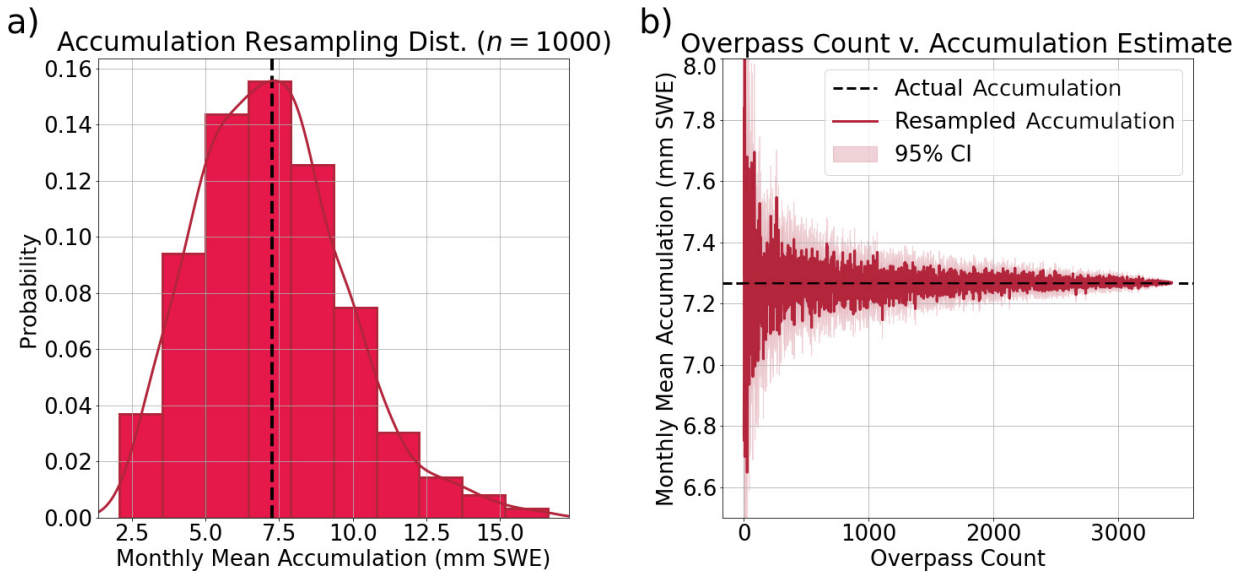

Figure 2.6: **CloudSat high latitude resampling uncertainty**. **a)** Monthly mean accumulation sampling distribution for a high latitude grid cell at 81.5° N, generated by resampling accumulation estimates 1000 times using the average overpass sample size at 60° N; and **b)** line plot of monthly mean accumulation estimates as we increase the number of overpasses in the sample from 1 to the full sample size ($n = 3439$) at the same grid cell in **a)**. The black dashed line indicates the true snow accumulation (μ_s).

our lowest latitude grid cells around 60° N (Fig. 2.1.a). Performing this resampling procedure at monthly timescales produces similar results to the climatological analysis, with overpasses also converging to the population snowfall rate at a latitude of approximately 60° N (not shown). Building on the work of King and Fletcher (2020) and Hiley et al. (2010) which found that higher latitude stations exhibit stronger correlations and lower RMSE between CloudSat and in situ snow observations, the results of this resampling experiment further demonstrate that while uncertainties from a low CloudSat sample at 60° N can influence the accuracy of the accumulation estimate in some cases, on average this latitude is an acceptable lower boundary for comparison at 1° resolution.

2.4.3 Quantifying Biases in NH SWE due to Low Data Quality

Applying the outlier detection technique across our NH grid allows us to produce a spatial map of outlier count totals for each of the B4 component datasets (Fig. 2.7). Higher outlier counts are found in the CAA and central Eurasia, along with a general trend of decreasing outlier totals further South. GlobSnow is the component with the most outliers, particularly in the CAA and Eurasia (Fig. 2.7.c). This can likely be attributed in part to the poorly constrained nature of these regions where few in situ observations are available coupled with limitations in GlobSnow estimates over areas of complex topography Takala et al. (2011), and day-to-day variability in its estimates of SWE as GlobSnow does not include data from previous daily estimates in its calculation of SWE for the current day (Mudryk et al., 2015). Examining the spatial distribution of outliers at monthly time scales (not shown) reveals that grid cells with data quality issues often appear as spatial clusters but seem to be randomly located geographically.

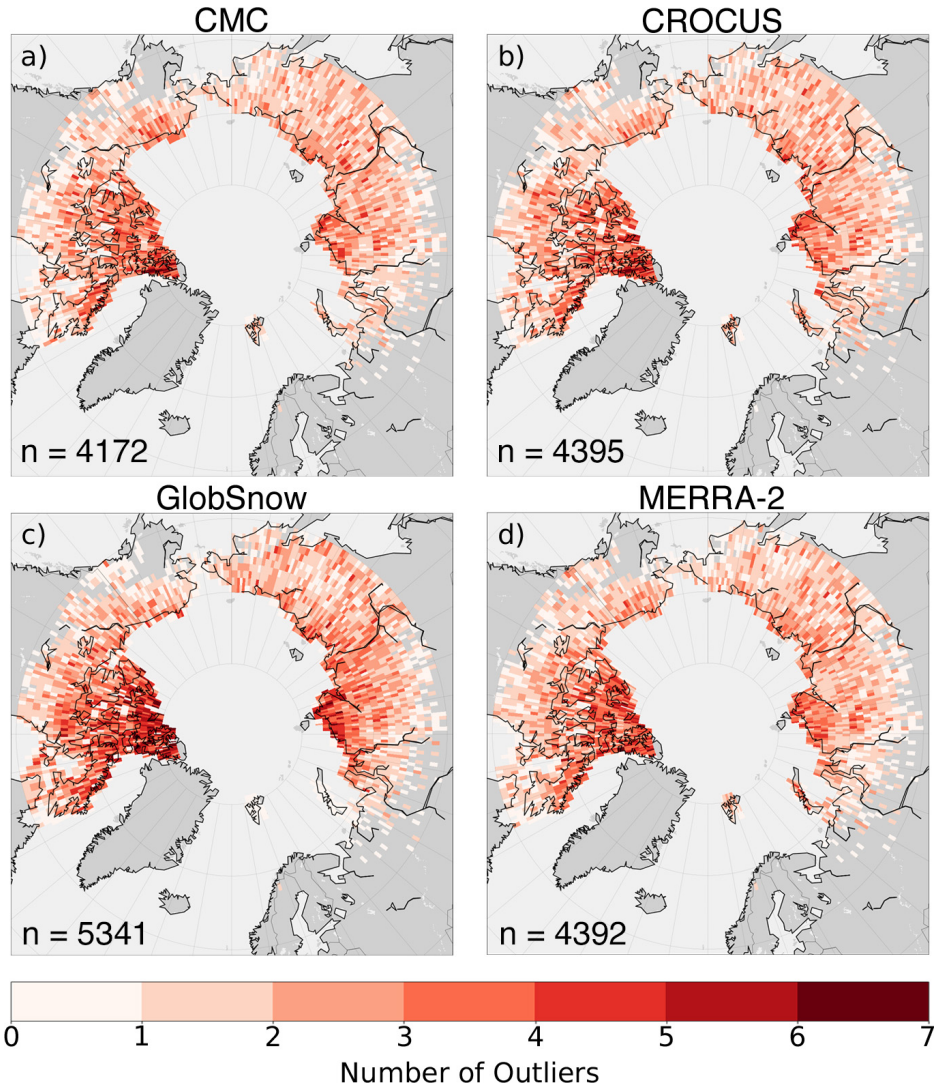

Figure 2.7: **B4 gridded product component spatial outlier maps**. Showing outlier count spatial distributions and grid totals (n) for **a)** CMC, **b)** CROCUS, **c)** GlobSnow and **d)** MERRA-2.

Monthly outlier count totals are highest during periods of heavy accumulation (October through December) and cold temperatures (Fig. 2.8.a). We note a steady decline in outliers during spring and summer months, when there is less frozen precipitation. Seasonal comparisons of outliers over September, October, November (SON) and December, January, February (DJF) exhibit a spatial structure consistent with monthly temporal distributions, as 79% of the total number of outliers occur over the CAA and central Eurasia during SON, contrasted by smaller outlier clusters in Eastern Siberia and lower Canada in DJF. Yearly outlier totals appear closely related to CloudSat data availability with similar outlier counts across all years excluding those where CloudSat experienced prolonged battery failures in 2009, 2011 and 2012 (Fig. 2.8.b).

Having flagged a total of 4885 month pairs in B4 with suspected data quality issues, our final contribution is to quantify the impact of data quality on estimated mean SWE in B4. First, an adjusted B4 SWE ($SWE_{B_{Adj}}$) is constructed as the mean of the four component products after masking the monthly SWE values in grid cells identified as statistically inconsistent using the method from Section 2.3.3. Next, the SWE bias attributed to low data quality is calculated as $SWE_{BIAS} = SWE_{B_{Adj}} - SWE_B$ for each grid cell (Fig. 2.9). This bias is generally negative across the NH, with a mean value of -3.1 mm in SON, which represents an approximate 14.9% reduction in total SWE on ground during this period. The distribution of this bias for all months in SON suggests that while in the majority of months the bias is less than 5 mm, there are some cells with extreme negative biases of more than 20 mm. These differences are visible in Fig. 2.9, where specific local regions in central and eastern Eurasia, the CAA and parts of southern Alaska, exhibit very large negative biases in mean SWE.

This absolute difference in total NH SWE over all SON month-pairs ($bias_{abs} \approx -310,000$ kg m^{-3}) represents a substantial amount of available water sitting on the surface in the form

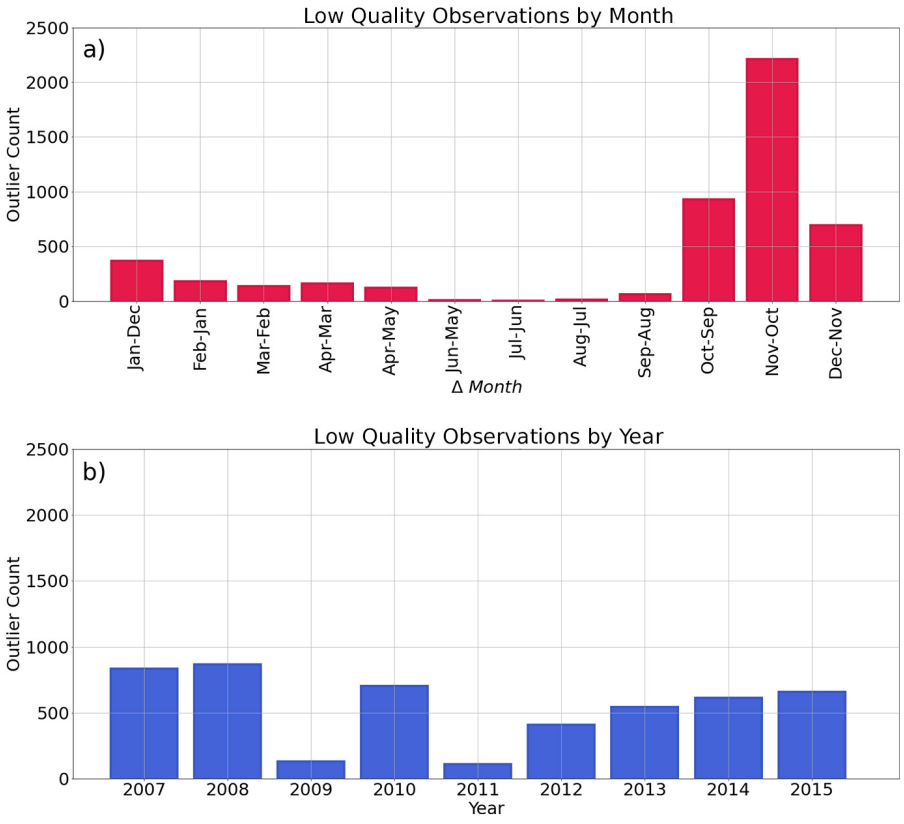

Figure 2.8: **Temporal distribution of outlier counts**. Outliers organized in B4 [2007-2015] for **a)** monthly and **b)** yearly totals.

of a frozen snowpack (the approximate water volume of a lake with surface water area 100 m^2 and 30 meter depth). Absolute SWE differences during DJF also exhibit a slightly negative average climatological bias, however DJF exhibits substantial spatial scatter with respect to the sign of the bias across the region, likely as a result of a reduced CloudSat sample from lower accumulation amounts during this time (see Fig. 2.8.a). The tendency for the SWE bias to be negative in Fig. 2.9 suggests that on average there exists excess snow accumulation in the gridded SWE products at these high latitudes in SON, likely in part as a result of the sparsely distributed in situ observational network which is a limiting factor in the accuracy of assimilation-based SWE products. Applying this masking technique at the Resolute Bay grid cell (Section 2.4.1), also resulted in an approximate 40% decrease in RMSE between in situ station observations and $SWE_{B_{Adj}}$ compared to using the default SWE_B. Our data quality methodology was able to automatically identify statistically inconsistent cases (Fig. 2.2) at the site and exclude problematic component products to improve the accuracy of the final adjusted blended product at this location.

2.5 Discussion and Conclusions

Through the process of aggregating CloudSat overpasses using techniques described in King and Fletcher (2020) and Hiley et al. (2010), we have produced a gridded snow accumulation product independent of current gridded SWE products that can be used as an additional observational constraint to enhance the accuracy and robustness of operational SWE datasets. Comparisons of NH gridded ΔSWE_B and $\overline{SWE}_C$ for all frozen months between January 2007 and December 2015 exhibit strong correlations over heavy accumulation periods (SON) with improved agreement at higher latitudes due to an increase in the available CloudSat sample across these regions. Using

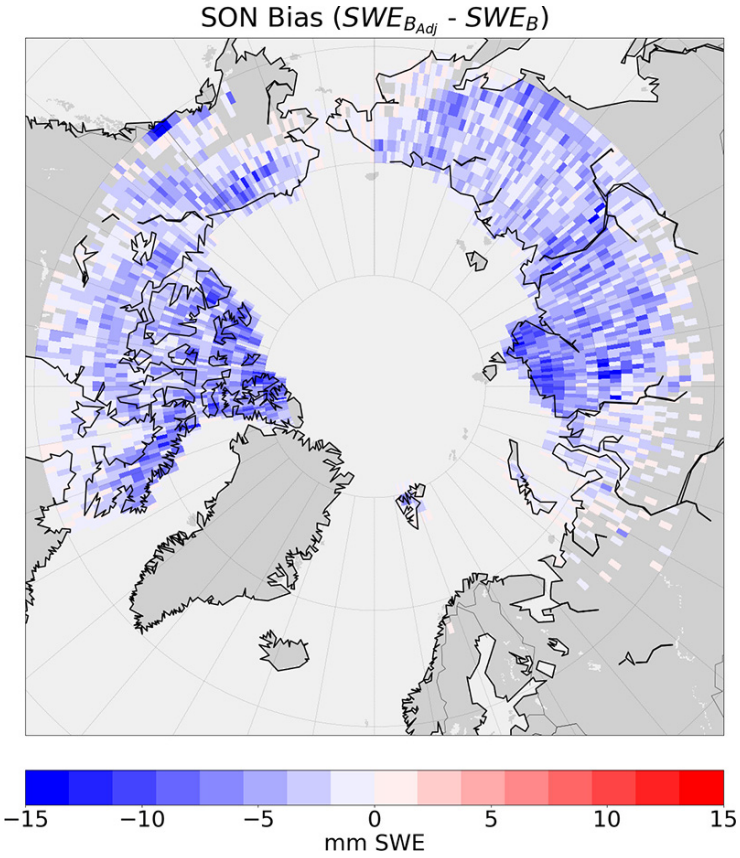

Figure 2.9: **Spatial bias with masked B4 outliers over SON**. Spatial differences between $SWE_{B_{Adj}}$ and SWE_B over SON where $SWE_{B_{Adj}}$ is a recalculated blended SWE product from component datasets, with periods highlighted as outliers masked in the blending process.

the 95% PI generated from a regression model for the Resolute Bay case study, this method was able to provide important insight into identifying the component product responsible for the data quality issues in B4. Applying this outlier detection technique to the wider NH grid produces a spatial distribution of outliers clustered primarily across poorly constrained high Arctic regions of the CAA and Siberia. Temporal outlier analyses also indicate that the majority (approximately 79%) of the total number of outlier events occur during periods of heavy accumulation in SON. Masking outlier month-pairs and re-blending the B4 product results in an average 14.9% reduction in SWE across the NH over SON, which represents a substantial difference in sitting surface water in the form of a snowpack (approximately $-310,000$ kg m^{-3}) and suggests that on average, portions of the B4 component products are accumulating too much SWE during this period.

A major assumption in the comparison of ΔSWE_B and $\overline{SWE}_C$ is that we consider $\overline{SWE}_C$ accumulation estimates as the truth when flagging outliers in the gridded products. In reality, uncertainties in the CloudSat retrieval of snowfall also contribute to uncertainty in the outlier detection via both sampling and instrumental uncertainty (Stephens, 2017; Wood et al., 2014; Duffy et al., 2021). One advantage of the outlier detection method proposed in this study is that it includes a measure of its own uncertainty via the width of the 95% PI (Fig. 2.2). Our analysis has found that this uncertainty tends to be largest in periods with a weaker relationship between CloudSat and B4, and that random variability in the location of cells which display low correlations is likely related to the effects of both a low CloudSat sample size coupled with biased CPR observations. Since CloudSat flies over some of the most southern grid cells only 3 to 4 times a month, if a heavy snowstorm occurs during one of those overpasses and is observed by CloudSat, it will bias the resulting accumulation estimate during that period which will then

lead to disagreement with estimates provided by B4. However, CloudSat has shown to generally demonstrate good accuracy at this spatio-temporal resolution at high latitudes during periods of strong accumulation, suggesting some confidence in the overall number of outliers detected during this time (Hiley et al., 2010; King and Fletcher, 2020; Bennartz et al., 2019; Palerme et al., 2014).

This methodology works most effectively if no systematic error obscures the relationship between ΔSWE_B and $\overline{SWE_C}$ (an assumption supported by the overall positive agreement shown in Fig 2.4). These systematic errors include overarching issues related to CloudSat's available sample, instrument retrieval accuracy and numerical model errors in the gridded SWE product forecasts of SWE on ground (Kulie and Bennartz, 2009; Broxton et al., 2016; Lindsay et al., 2014). These issues increase uncertainty in the relationship between ΔSWE_B and $\overline{SWE_C}$ and limit our ability to detect statistically inconsistent periods in the gridded products. The steps we take in restricting the CloudSat sample in Section 2.3.3.b, along with the results of complementary studies by Duffy et al. (2021) and King and Fletcher (2020), help to minimize the effect of such biases and suggest that the existence of large systematic errors between CloudSat and B4 across the NH are negligible.

Radar signal attenuation from complex terrain and ground clutter interference are two additional critical sources of uncertainty when considering the accuracy of CloudSat snowfall estimates (Bennartz et al., 2019; Palerme et al., 2019; L'Ecuyer et al., 2008; Hiley et al., 2010). Since snowfall is extrapolated down to the surface from the lowest precipitating bin, areas of high elevation and complex terrain can influence retrieval accuracy by attenuating the CPR signal directly above CloudSat's radar blind zone at 1.2 km (Marchand et al., 2008; Matrosov and Battaglia, 2009). The impacts of these phenomena influencing snowfall rate accuracy has been

noted in previous work by Kulie and Bennartz (2009) and the CPR radar blind zone can be seen in the yellow line at the bottom of Fig. 2 in (Hudak et al., 2008). Attenuation over alpine regions is critically important to consider, as large swaths of our study area include grid cells primary composed of high elevation topography and contain large quantities of SWE (Wrzesien et al., 2018). Furthermore, outputs of the regression model described in Section 2.3.3, such as the standard error or the y-intercept, exhibit larger biases in alpine regions across all B4 component products as a result of CloudSat-CPR retrieval deficiencies in these regions (Fig. 2.1.c).

Gridded products are capable of providing nearly complete spatio-temporal coverage of SWE across the NH. This improved coverage is an extremely valuable resource in cryospheric research and as inputs to hydrologic models. However, these products are not without uncertainty and error, and we should take whatever steps are possible to enhance the accuracy of their estimates. The outlier detection technique described here is a novel method which allows for satellite observations to constrain gridded products and, in turn, provide researchers with an automated check for flagging areas of statistically inconsistent SWE estimates. While this study uses CloudSat snowfall to constrain B4 SWE and its component datasets, we argue that this technique could be applied to any gridded SWE dataset to identify potential inconsistencies. This technique can support future blended SWE products by providing information regarding where and when to include component datasets in the ensemble averaging process. Additionally, this outlier detection methodology could also be used to identify and fix issues with surface instrumentation (e.g. ECCC acoustic snow depth measurements) which are assimilated by other products. CloudSat is uniquely positioned as one of the few available observational snowfall constraints in remote regions and, when retrievals are aggregated at appropriate scales, CPR snowfall estimates can provide valuable new insights into the accuracy of current and future

gridded SWE products.

Chapter 3

A Centimeter-Wavelength Snowfall Retrieval Algorithm Using Machine Learning

3.1 Overview

Remote sensing snowfall retrievals are powerful tools for advancing our understanding of global snow accumulation patterns. However, current satellite-based snowfall retrievals rely on assumptions about snowfall particle shape, size and distribution, which contribute to uncertainty and biases in their estimates. Vertical radar reflectivity profiles provided by the VertiX X-band radar instrument in Egbert, Ontario are compared with in situ surface snow accumulation measurements from January-March 2012 as a part of the Global Precipitation Measurement (GPM) Cold

Season Precipitation Experiment (GCPEx). In this work, we train a random forest (RF) machine learning model on VertiX radar profiles and European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA-5) atmospheric temperature estimates, to derive a surface snow accumulation regression model. Using event-based training-testing sets, the RF model demonstrates high predictive skill in estimating surface snow accumulation at 5-minute intervals with a low mean square error (MSE) of approximately 1.8×10^{-3} mm^2 when compared to collocated in situ measurements. The machine learning model outperformed other common radar-based snowfall retrievals ($Z_e - S$ relationships) which were unable to accurately capture the magnitudes of peaks and troughs in observed snow accumulation. The RF model also displayed consistent skill when applied to unseen data at a separate experimental site in South Korea. An estimate of predictor importance from the RF model reveals that combinations of multiple reflectivity measurement bins in the boundary layer below 2 km were the most significant features in predicting snow accumulation. Nonlinear machine learning-based retrievals like those explored in this work can offer new, important insights into global snow accumulation patterns and overcome traditional challenges resulting from sparse in situ observational networks.

3.2 Introduction

There are many reasons why it is important to measure falling snow. Accumulated snow has important linkages to regional flooding, water resource management practices and ecosystem development and sustainability (Buttle et al., 2016; Berghuijs et al., 2019; Déry and Brown, 2007; Peng et al., 2010). As global average temperatures continue to rise, the distribution and magnitude of synoptic snowfall events are expected to also change in a substantial, non-linear

manner (Vaughan et al., 2013). In many cases, the most effective way to measure snow is through radar. Satellite radar measurements, for example, can provide valuable snowfall information in remote regions such as the high Arctic, tundra or over sea ice (King and Fletcher, 2019; Cabaj et al., 2020; Duffy et al., 2021). They can also be used to retrieve measurements of melting snow in the ocean that may represent up to 10% of the seasonal ocean surface heat flux (Duffy and Bennartz, 2018). Satellite-based snowfall radar retrievals have also been used as independent observational datasets to evaluate and constrain current gridded snow products across Arctic regions, which are typically poorly observed due to sparsely distributed in situ measurement networks (King and Fletcher, 2021).

Radars are valuable resources as they can provide precipitation rates over a wider area and at a finer temporal scale compared to traditional in situ precipitation gauges (Lemonnier et al., 2019). Radars can only provide estimates of precipitation, however, and the accuracy of retrieved precipitation rates is largely dependent on the accuracy of the retrieval algorithm (Kulie and Bennartz, 2009; Hiley et al., 2010). There are several common methods typically used to convert radar measurements into precipitation rates. So-called $Z_e - S$ relationships (i.e. $Z_e = a \times S^b$) use algebraic power laws, with the prefactor a and exponent b derived from experimental observations or simulated models to convert near-surface reflectivity (Z_e) into precipitation rates (S) (Wolfe and Snider, 2012; Boucher and Wieler, 1985; Szyrmer and Zawadzki, 2010). Bayesian methods, such as the techniques used in the derivation of the 2C-SNOW-PROFILE snowfall retrieval product from CloudSat (Stephens, 2017), are also used for snowfall retrieval by incorporating prior assumptions of snowfall particle size, shape, distribution and fall speed (Wood and L'Ecuyer, 2021).

These empirically-derived relationships may only provide accurate estimates under specific

climate regimes as a consequence of the necessary a priori assumptions made regarding particle microphysics. Substantial error can therefore propagate through to the final snow accumulation estimates if the wrong $Z_e - S$ relationship is applied (Milani et al., 2018; von Lerber et al., 2017; Schoger et al., 2021). The variability in the structure and vertical extent of winter storms can also lead to additional uncertainty and error when deriving a surface snow accumulation estimate from only near-surface reflectivity bins when applying such a relationship. While the significance of each radar bin may vary, an ideal retrieval algorithm would incorporate all the relevant reflectivity information that a radar beam can provide to maximize the accuracy of the estimated surface accumulation. Such an algorithm would also be able to assimilate cloud features and environmental conditions that are known to further influence $Z_e - S$ relationships and Bayesian retrievals (Pettersen et al., 2020b; McIlhattan et al., 2020).

Traditional algebraic methods for precipitation retrievals can become impractical if we consider the full set of variables present in a typical radar retrieval. Machine learning (ML) algorithms, however, are unrestricted by algebra and can incorporate a virtually unlimited number of variables to derive complex and non-intuitive relationships between the provided predictors and response (Karpatne et al., 2019). Previous studies have used ML to construct snowfall retrieval algorithms which can detect and estimate surface snowfall rates under specified climate contexts using spaceborne passive microwave radar measurements (Adhikari et al., 2020b; Ehsani et al., 2021). NNs have also recently shown promise in retrieving particle size distribution (PSD) parameters from synthetic model data which, when applied to spaceborne radar observations, demonstrates improved estimates of snowfall compared to their default algorithms (Chase et al., 2021). These models are powerful tools for deriving snowfall estimates from radar data, and in further advancing our understanding of how accumulated snow quantities are changing under a

warming global climate.

Herein, we develop and evaluate a novel random forest (RF) snow accumulation retrieval algorithm that incorporates reflectivity and temperature from all applicable radar range gates. The RF model is then compared with traditional $Z_e - S$ relationships (derived from X-band radar) to evaluate model skill. Finally, both the RF and $Z_e - S$ relationships are applied to data from a completely unseen site in South Korea to evaluate model robustness. We use the RF algorithm to identify which radar gates appear as the most important, along with how many gates are necessary to maximize the accuracy and robustness of the retrieval. For the purposes of this methodological proof-of-concept study in the development of a retrieval strategy that could be applied to terrestrial and spaceborne radar, we are focusing on X-band radar measurements of snowfall to minimize potential complications that may arise from attenuation, non-Rayleigh reflectivity, or multiple scattering.

The main objectives of this paper are to:

1. Quantify the accuracy with which we can model surface snow accumulation from X-band radar data using an ensemble of statistical techniques of varying complexity

2. Determine whether machine learning models offer improved skill compared to traditional $Z_e - S$ relationships

3. Identify the most significant range gate estimates of temperature and reflectivity to enhance current and future retrievals of surface snow accumulation

3.3 Data and Methods

3.3.1 In Situ Snow Observations

In situ precipitation measurements were obtained from the Global Precipitation Measurement (GPM) Cold Season Precipitation Experiment (GCPEx) field campaign (Skofronick-Jackson et al., 2015). GCPEx was a joint venture between the National Air and Space Administration (NASA) and Environment and Climate Change Canada (ECCC). A primary goal of GCPEx was to provide multiple observations of snowstorms in order to evaluate GPM-style retrievals of precipitating clouds. GCPEx took place across several sites in Ontario, though for this study we focus on observations taken at the Centre for Atmospheric Research Experiments (CARE) in Egbert, Ontario, Canada from January to March 2012 (Fig. 3.1.a).

Surface snow accumulation was monitored using an OTT Pluvio2 automated precipitation weighing gauge at the GCPEx study site. The gauge was installed at a representative location in an open field, free from the influence of human activity and the presence of nearby vegetation. Snow accumulation estimates were recorded at one-minute temporal resolution to provide $n = 25$ days of non-zero precipitation observations. The Pluvio2 minimum observational intensity threshold is 0.2 mm per minute of snow accumulation for measurements to qualify as active precipitation and be registered by the system (i.e. trace amounts of accumulation). Accumulation estimates that fell below this minimum threshold were not considered as precipitation and therefore left unrecorded. This measurement gauge was equipped with a 200 cm^2 heated top to help prevent snow accumulation from covering the gauge top orifice and mitigate against underestimation. The Pluvio2 was also equipped with 8-ft diameter Alter shield and was situated in the centre of a

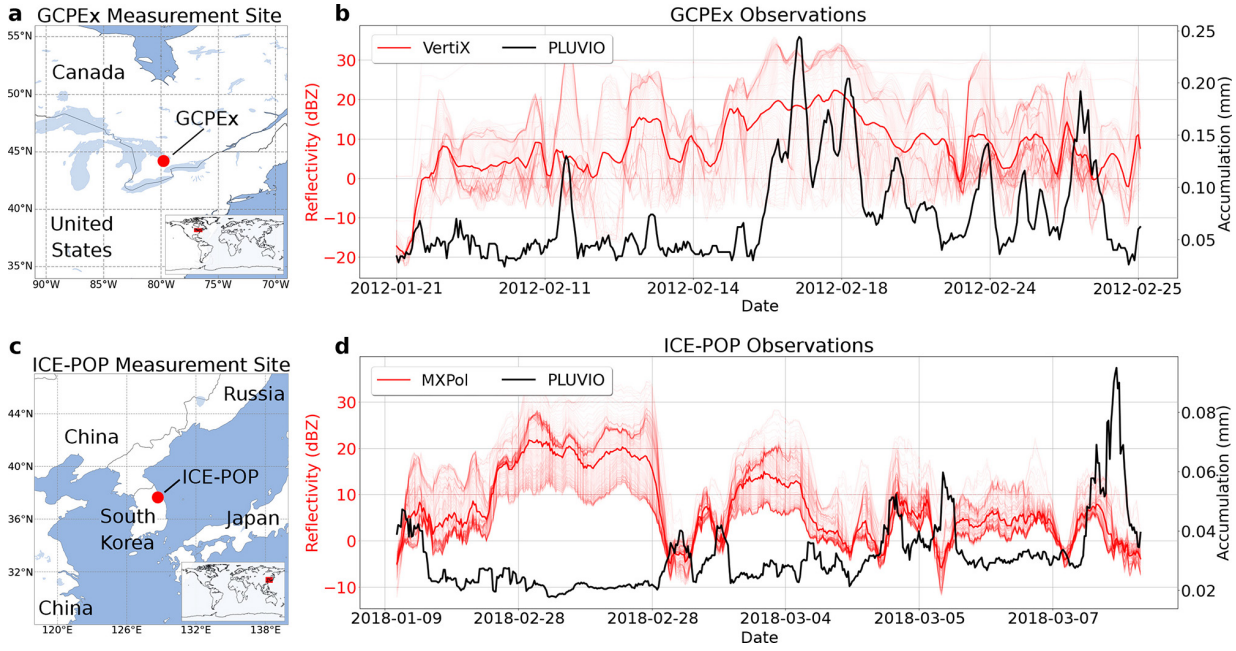

Figure 3.1: **Experiment locations and vertical radar profile summaries**. **a)** GCPEx measurement site location (Egbert, Ontario, Canada); **b)** 5-Minute VertiX reflectivity values (mean in bold red, individual vertical bins in light red) and Pluvio2 snow accumulation estimates (black) from the site in **a)**; **c)** ICE-POP measurement site location in South Korea; and **d)** the same as b) for ICE-POP using MXPol X-band radar observations.

surrounding Tretyakov fence (Smith, 2009; Metcalfe et al., 1997). This double-fence setup helps to mitigate the effects of turbulent winds, which can deflect incoming hydrometeors away from the gauge-top opening and lead to an underestimation of accumulation (i.e. undercatch effects) (Sevruk et al., 1991; Kochendorfer et al., 2017; Rasmussen and Rasmussen, 2014; Colli et al., 2020). To further mitigate against issues with strong winds and gauge undercatch, we constrain our analysis to periods where surface wind speeds recorded during GCPEx were less than 5 m/s (i.e. only low-to-medium intensity wind speed periods) for the duration of the measurement period (Kochendorfer et al., 2022; Rasmussen et al., 2002).

Additional surface snow accumulation observations and collocated X-band radar data were collected from the International Collaborative Experiment for PyeongChang Olympic and Paralympics (ICE-POP) in South Korea (Fig. 3.1.c). Similar to GCPEx, ICE-POP was another GPM Ground Validation campaign that collected a suite of atmospheric observations for researching frozen and mixed-phase precipitation (Gatlin, 2020). The experiment included a Pluvio2 gauge with a similar double-fenced shield to that at GCPEx, as well as a collocated MXPol X-band radar system (Gehring et al., 2020). As the MXPol data extends beyond 10 km, the lowest subset of the vertical column is extracted and aligned to the vertical resolution (44.5 m) of bins used at GCPEx (Fig. 3.1.d). Preprocessing of the MXPol and Pluvio2 data from ICE-POP follow the same steps as those applied to data at GCPEx (further details in Section 3.3.4).

To address uncertainties arising from trace amounts of accumulation being missed at the surface Pluvio2 gauge, we examined varying levels of temporal data averaging and also performed sensitivity analyses on minimum precipitation thresholds to determine whether certain low-intensity observations should be included in the model training data. We examined minimum accumulation thresholds from 0.01 mm to 0.16 mm and found that each model demonstrated similar strong performance (low MSE) for minimum thresholds of 0.01 and 0.02 mm, with reductions in performance as we move to larger minimum thresholds. For example, we note a 64% increase in MSE for RF_{full} trained using GCPEx data and applied at ICE-POP when the training threshold is increased from 0.01 to 0.04 mm (and a 175% increase in MSE when the training threshold is increased from 0.01 to 0.16 mm). To maintain high accuracy and a large sample size, we therefore employ a minimum 5-minute average accumulation threshold of 0.01 mm.

3.3.2 VertiX Radar

The vertically pointing X-band ($\lambda \approx 3$ cm wavelength) radar (VertiX) provides reflectivity profiles with a sampling frequency of 3 seconds, a range of 10 km with vertical bin resolution of approximately 44.5 meters, and a sensitivity below -20 dBZ. The VertiX instrument emits pulses of energy which scatter off of the particles within the cloud to provide precipitation echo measurements, which can then be used to derive information about precipitating hydrometeors contained within the vertical column above the instrument (Skofronick-Jackson et al., 2015). This type of X-band radar has been used to great effect in previous work by Gehring et al. (2020) during experiments focused on analyzing precipitation type as well as accumulation quantification. This instrument was positioned in the same study location as the Pluvio2 weighing gauge described in Section 3.3.1 to provide time-height estimates of reflectivity (dBZ) and Doppler velocity (m/s) for precipitating clouds directly overhead both instruments. Reflectivity values and Pluvio2 accumulation estimates are shown for the full study site in Figure 3.1.b. While X-band radars have traditionally been used for rainfall retrieval algorithms, recent studies suggest that this radar band can provide valuable information related to snowfall rates in cold environments, and the longer wavelength compared to Ku/Ka radar helps to mitigate against the effects of non-Rayleigh scattering of large snow particles (Matrosov et al., 2009, 2007).

3.3.3 ERA-5 Reanalysis

Atmospheric temperature data is provided by the European Center for Medium-Range Weather Forecast's (ECMWF) Reanalysis Version 5 (ERA-5) dataset. ERA-5 is a global reanalysis product that provides hourly estimates of a suite of climatic variables at $0.25° \times 0.25°$ resolution from

1979 to present, with vertical column estimates up to 80 km into the upper atmosphere (Hersbach et al., 2020). As a fifth generation product, ERA-5 is well documented and its estimates of temperature have been extensively analyzed and validated in previous literature (Tarek et al., 2020; Betts et al., 2019; Simmons et al., 2020; Wang et al., 2019a). Hourly ERA-5 atmospheric temperature data was extracted for each bin in the lowest 10 km of its vertical profile for the grid cell directly overlapping with the Egbert study site over our study period. Vertical bins in ERA-5 were then regridded to match the approximate 44.5 m vertical bins defined by the VertiX radar system. The resulting binned temperature data is then used as an additional predictor set in the snowfall retrieval techniques described in Section 3.3.4.

3.3.4 Snow Accumulation Retrieval Methodology

There were twelve days during GCPEx where snowfall was measured by both the VertiX and Pluvio2 instruments. The VertiX reflectivity profiles for these twelve days of collocated measurements are shown in Figure 3.2 along with the date of occurrence. We note the unique structure of precipitation in each day of Figure 3.2, along with a relatively uniform low-intensity band of reflectivity values (approximately equal to 0 dBZ on average), between the 0 to 1.5 km portion of the vertical profile. The impact of this near-surface reflectivity band on model skill is discussed in more detail in Section 3.5.

To begin the comparison process, each vertical bin of reflectivities in the VertiX radar profile (in physical units $mm^6\ m^{-3}$), along with each 1-minute in situ accumulation estimate is temporally averaged to common 5-minute intervals. This interval size was selected to help account for small jumps in trace amounts of precipitation from the in situ gauge, and form a more continuous

reference dataset while still maintaining a sufficient sample size for model training and testing ($n = 391$ observations). Furthermore, as only periods of non-zero precipitation were reported by the in situ gauge, our comparisons are restricted to periods of active snowfall.

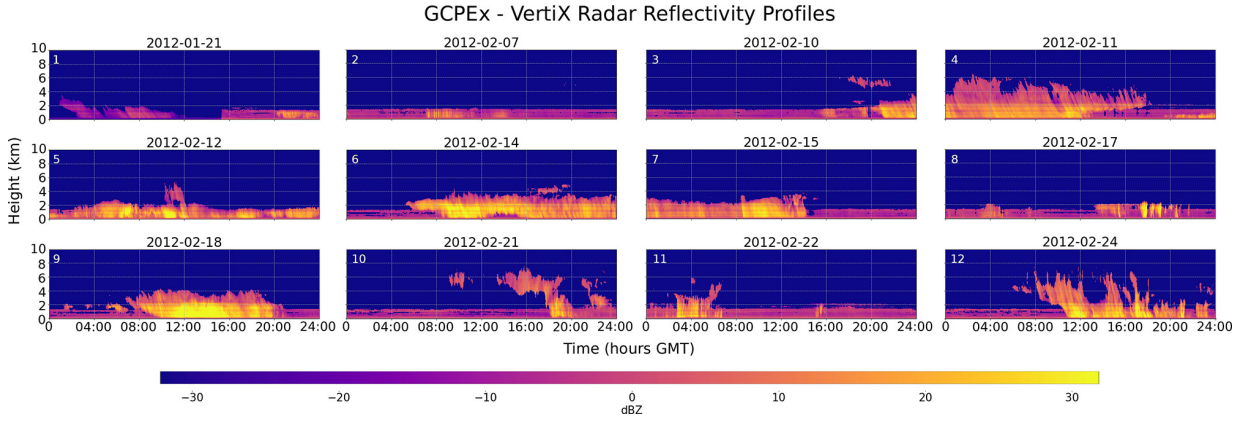

Figure 3.2: **Vertical radar reflectivity profiles (X-band)** generated from the VertiX radar over $n = 12$ collocated precipitation events up to 10 km above the surface.

Model performance is evaluated using mean squared error (MSE) calculated from a non-shuffled 10-fold cross validation (CV) in order to use the full set of available data and to mitigate against potential model overfitting. MSE is a common metric used in ML precipitation retrieval studies to quantify the average error between observed and predicted values, and assess model skill (Chen et al., 2020b; Kim and Bae, 2017; Shi et al., 2017). The non-shuffled nature of each 90/10 train/test split allows for the model's predictive skill to be examined when tested on unseen data, as retrievals taken close in time during training may be autocorrelated and lead to overfitting (i.e. a shuffled-split approach).

Each RF retrieval model examined here is provided with the same set of training data containing predictors and response variables. The response variable is the in situ snow accumulation observations provided by the Pluvio2 instrument. The default model predictors include the 221

Table 3.1: Model list and summary descriptions of the predictors used within each model.

Model	Variables	Inputs	Description
RF_{full}	Reflectivity, Temperature	442	All reflectivity and temperature bins
RF_{near}	Reflectivity, Temperature	90	Near-surface reflectivity and temperature bins
RF_{far}	Reflectivity, Temperature	352	Upper reflectivity and temperature bins
$Z_e - S$	Reflectivity	1	$Z_e - S$ relations described in more detail in Table 3.2

vertical reflectivity bins from the VertiX radar and the same 221 atmospheric temperature bins from ERA-5. Model predictors can also take the form of aggregations and subsets of these bins, depending on the model and experiment being performed. More information on the individual structures of these experiments can be found in Table 3.1, along with their general performance in Section 3.3.

Table 3.2: K-Band precipitation power law relationships derived from previous literature.

Name	Power Law	Source
DDU	$Z_e = 76 \times S^{0.91}$	(Grazioli et al., 2017)
B90A	$Z_e = 67 \times S^{1.28}$	(Matrosov et al., 2009)
W08A	$Z_e = 28 \times S^{1.44}$	(Matrosov et al., 2009)
W08B	$Z_e = 36 \times S^{1.56}$	(Matrosov et al., 2009)
W08C	$Z_e = 48 \times S^{1.45}$	(Matrosov et al., 2009)

In order to assess RF model accuracy, a set of five $Z_e - S$ relationships derived from X-band radar experiments were selected (Table 3.2) to act as baseline comparisons. These relationships were applied to the 11^{th} temporally averaged radar bin from the surface, as this region of the profile demonstrated strong predictive skill in our later analysis of vertical profile feature importance (Section 3.4.2) and lay far enough above the surface to avoid concerns of ground

clutter contamination. Each relationship's surface snow accumulation estimate was then recorded for comparison with the RF over a common 5-minute timestep.

A set of bespoke $Z_e - S$ relationships were also empirically derived using the 5-minute average reflectivity observations (in $mm^6\ m^{-3}$) and temporally aligned in situ snow accumulation measurements from the Pluvio2 at GCPEx and ICE-POP. These fits were performed under the same 10-fold CV strategy as the RF, using the Python SciPy package's $curve\_fit$ optimization algorithm to derive custom GCPEx and ICE-POP power law relations denoted as GZS and IPZS, respectively. Derivations were performed using a non-linear least-squares approach for finding optimal a and b coefficients to fit a $Z_e - S$ power law relationship (Virtanen et al., 2020). GZS and IPZS were then applied to unseen reflectivity observations in the testing set (in the same manner as the other $Z_e - S$ relationships) to derive corresponding surface snow accumulation estimates. GZS and IPZS were included in this work to act as additional baseline comparisons and to further demonstrate the robustness of a more sophisticated ML-based algorithm over traditional power law relationships when derived from the same sets of site-specific observations.

The RF regression model is implemented using Python's scikit-learn package from Pedregosa et al. (2011a), which provides access to a wide library of model setup, prediction and evaluation functions. RF model hyperparameterization is performed using a random search 10-fold cross-validation, where subsets of model hyperparameters are randomly tested across the full parameter space to identify a theoretical optimal set of values for our model and data. Using this technique, we examined the parameter space for the hyperparameters listed in Table 3.3 to identify optimal values.

Table 3.3: Random Forest hyperparameters and final tuned values.

Hyperparameter	Optimal Value	Parameter Space
Forest Size	400	[100, 200, 400, 800, 1600]
Bootstrap	True	[True, False]
Max Features	Sqrt	['auto', 'sqrt', 'log2']
Min. Samples Split	2	[1, 2, 5, 10, 20]
Min. Samples Leaf	10	[1, 2, 5, 10, 20]
Max Depth	10	[5, 10, 15, 25]

3.4 Results

3.4.1 Model Intercomparisons

Cumulative accumulation timeseries are shown for all models (colored lines) and Pluvio2 snow measurements (black dashed line) in Figure 3.3.a for the full study period. While all models display similar accumulation trends over the full study period, the RF appears to be the most capable in capturing peaks and troughs in the data over time, to more closely model the correct magnitude of the total recorded surface snow accumulation. W08A also closely models the total accumulation over the full period, but it (along with the other power law relationships) displays a strong negative bias during the first half of the study. W08A and DDU then display a positive total accumulation bias during the February 18 snowfall event, however all $Z_e - S$ relationships (excluding GZS) underestimate the precipitation rate following this event (as the slope of the accumulation curves appear to flatten out compared to in situ). While GZS closely follows in situ accumulation during the trace precipitation events at the beginning of GCPEx, on February 10 it begins to overestimate the amount of surface accumulation and remains positively biased by two

orders of magnitude more than RF_{full} for the remainder of the period when compared to in situ. We note that the RF displays the lowest overall error in its predictions, with MSE values smaller than any of the other $Z-S$ relationships.

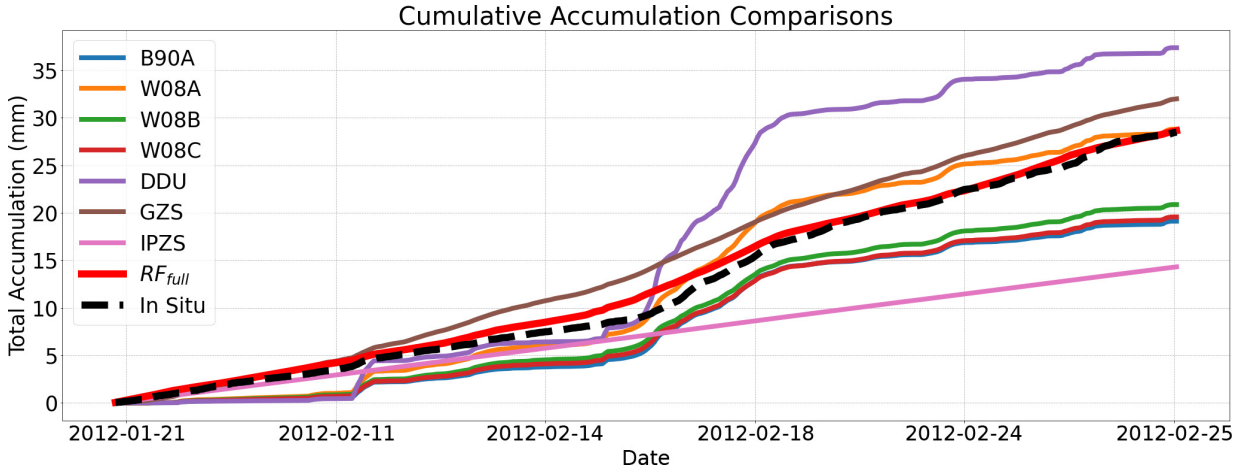

Figure 3.3: **Cumulative snow accumulation estimates**. Total snow accumulation intercomparisons between in situ measurements, RF_{full} and each respective $Z_e - S$ relationship over the course of all accumulation events at GCPEx.

As mentioned in Section 3.2, the referenced $Z_e - S$ power law coefficients are empirically derived from a set of observed atmospheric particle microphysics. These assumptions about particle microphysics are a large source of uncertainty in retrieval accuracy if the snowfall event being observed falls outside measured values recorded at the experiment from which the power law was derived. These uncertainties contribute to additional retrieval error when applied to data outside the training dataset, which is likely a large source of the error between each $Z_e - S$ relation and the Pluvio2 estimates.

To further examine model performance on unseen data and assess the robustness of the RF retrieval accuracy, we performed a secondary test against a similar experimental setup at the

ICE-POP site in South Korea (Fig. 3.4.a). Testing RF_{full} on ICE-POP data (while trained only on data from GCPEx) allows us to examine how well the model can generalize to completely unseen data and further assess the impacts of potential near-surface radar contamination from the VertiX radar.

The results of this test are shown in Figure 3.4.b, for each model at both sites. In general, the $Z_e - S$ relationships show a large change in MSE (e.g. DDU MSE improves by approximately 50% at ICE-POP), while the RF MSE remains fairly consistent between sites. Furthermore, while the $Z_e - S$ performance improves when applied at ICE-POP, the RF continues to exhibit the highest skill in predicting surface snow accumulation. While MSE values are similar for GZS and RF_{full} at GCPEx (a 5% increase in MSE for GZS), the power law performance degrades substantially at ICE-POP with an MSE increase of 40%. We also note a similar result from IPZS where, when applied at ICE-POP (i.e. the site it was trained), IPZS demonstrates very low MSE, however performance degrades by 150% when IPZS is applied at GCPEx. This reduction in GZS and IPZS performance when applied to new experiment sites further highlights the bespoke nature of empirically derived $Z_e - S$ relationships and their inability to generalize well to new regional climates. While the RF performance degrades slightly at ICE-POP, the overall similarity in performance suggests that the RF is capable of providing accurate estimates of accumulation, even when applied to completely unseen retrievals from a different climatic regime and time period.

While RF_{full} is able to capture the peaks in high intensity precipitation early and midway through the experiment, it struggles to accurately predict the high intensity precipitation event on February 24 – 25. These high cloud storms represent a small subset of the weather events observed at GCPEx and also include much more variability in the vertical reflectivity profile observations

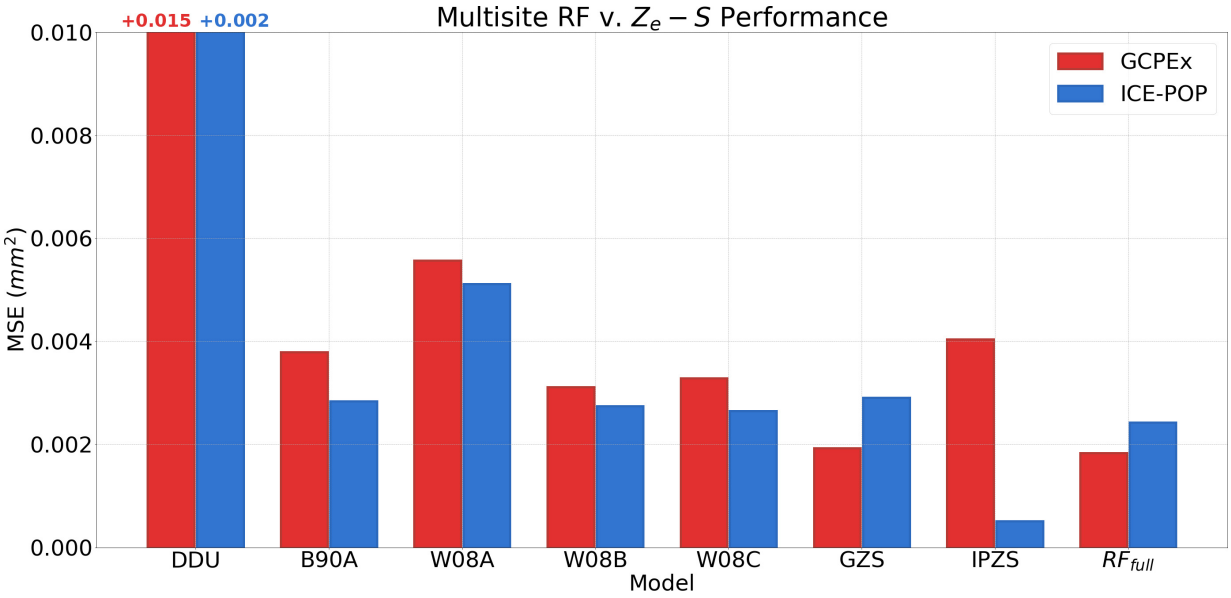

Figure 3.4: **MSE values from RF_{full} and the $Z_e - S$ relationships when tested against GCPEx observations (red) and ICE-POP (blue)**. Note that the ICE-POP tests were trained only using GCPEx observations (excluding IPZS which was derived from ICE-POP data), and that DDU MSE values extend beyond the chart top (extent labels are included).

(lower right panel of Figure 3.3). These uncertainties in the RF estimates are further visualized in Figure 3.5, which shows the distribution of model errors for the RF and each subset RF model (more details in Section 3.4.2) against the errors of the ensemble of $Z_e - S$ relationships examined in this work. We note a slight positive bias to the RF with an anomaly curve centred closer to zero compared to the negatively biased and long-tailed estimates from the $Z_e - S$ relations in Figure 3.5.a (excluding GZS, which is the only positively biased $Z_e - S$ relationship). Further, the RF model error during the high intensity February 24 – 25 event is clearly visible in the error timeseries (Fig. 3.5.b), where all $Z_e - S$ relationships and the RF display high levels of error in their estimates of surface accumulation. MSE values from RF_{full} are 100% higher on average when tested on the February 24 – 25 storm event compared to the other storm events at GCPEx.

We suggest that a larger, more robust reference dataset is required to allow the ML models to more clearly learn about differing storm types and varying cloud layer precipitation structures (which would need to be classified in advance using an additional, preprocessing supervised learning layer).

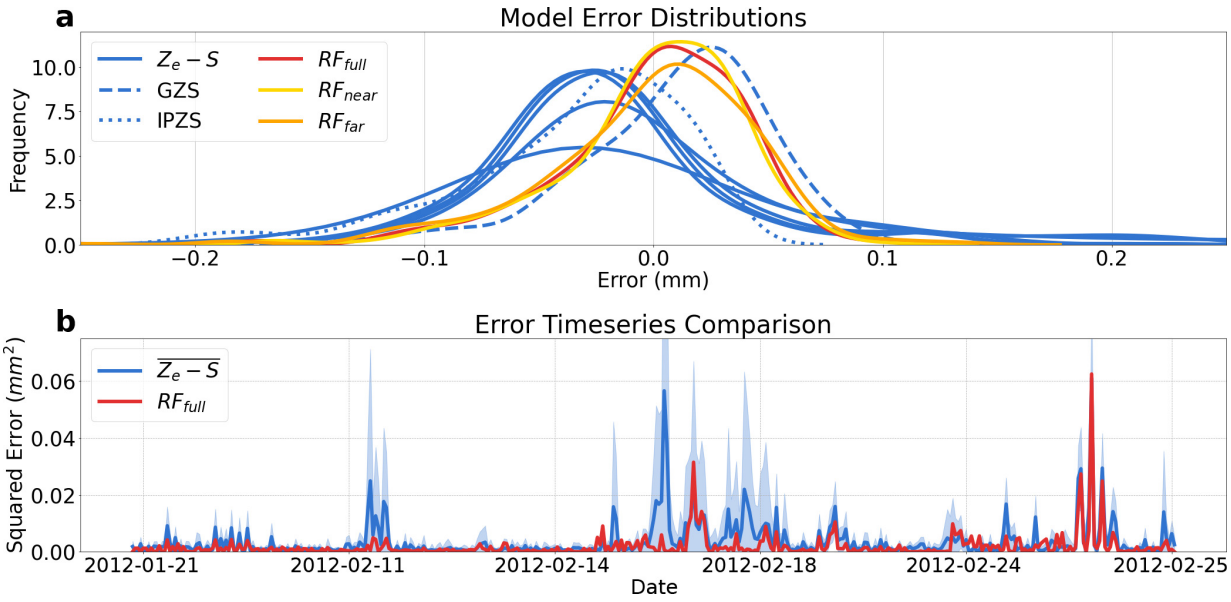

Figure 3.5: **Snowfall error distributions and uncertainty timeseries**. **a)** Distribution of model errors for RF_{full} (and subset models), along with the ensemble of $Z_e - S$ relationships for predictions at GCPEx; and **b)** error timeseries for the RF model estimates and the mean of all tested $Z_e - S$ relationships for each 5-minute timestep of observed reflectivities at GCPEx (shaded region shows a 95% confidence interval).

3.4.2 Vertical Profile Features

As a consequence of the splitting process used by individual decision tree nodes, we gain insight into the relative importance of each predictor contributing to the decisions made within the model. These predictor importance scores are useful metrics which provide additional

information regarding model behaviour and performance as a function of the atmospheric vertical column. Predictors in this study are based on the individual 221 VertiX reflectivity bins, the 221 ERA-5 atmospheric temperature bins, or as combinations/aggregations of these variables. A visualization of the importance of each individual vertical bin for reflectivity and temperature is shown in Figure 3.6, along with the respective frequency distributions for each bin in the vertical profile for RF_{full}. The frequency distribution of reflectivity values converges as we move further up the vertical profile, as missing regions are filled with the mean of the column. As noted in Figure 3.6.a, near surface reflectivity values comprise the most significant predictors by far, followed by a handful of cloud-top level ERA-5 temperature bins near 3.5 km above the surface. Reflectivity values are on average 10 times more influential in the model's decision-making process than the atmospheric temperature data.

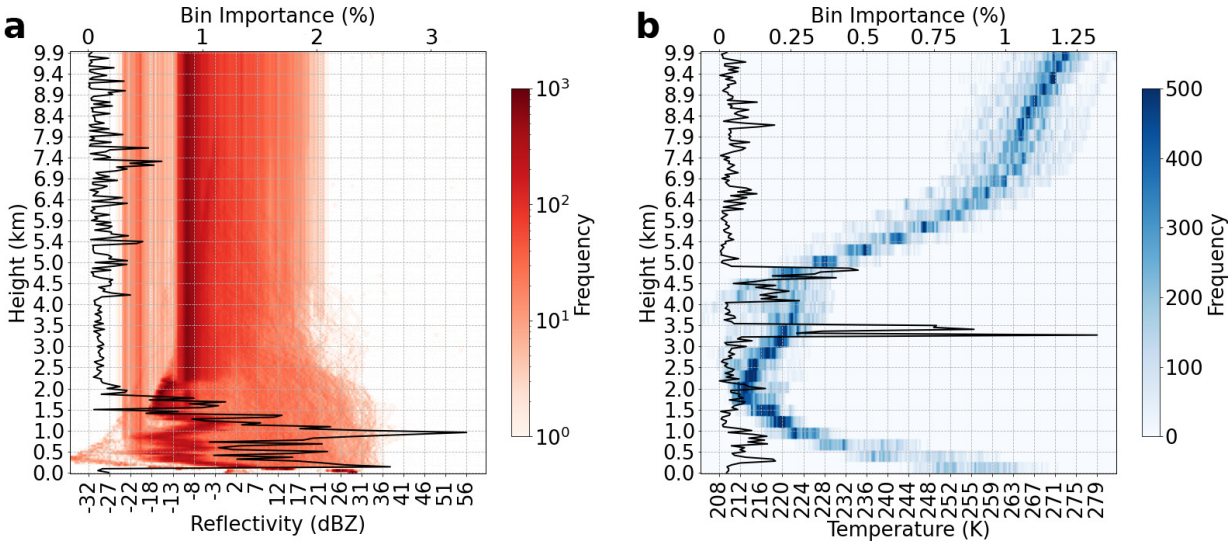

Figure 3.6: **RF feature importance scores and observation frequency histograms**. **a)** Histogram of reflectivity values from all snowfall events up to 10 km; and **b)** the same as in **a)** for ERA-5 temperature profiles. RF_{full} bin importance scores are shown in solid black for both datasets.

We attempt to simplify the model architecture by refining our set of predictors from the full set of hundreds of predictors (RF_{full}) to two subsets (Table 3.4). Aggregating multiple contiguous high-importance bins from Figure 3.6 results in two new predictor sets: RF_{near} (reflectivity/temperature bins 1-45; combined importance 65%) and RF_{far} (reflectivity/temperature bins 46-221; combined importance 35%). Rerunning the model with RF_{near} as the predictor set shows very similar skill to that of RF_{full} ($MSE_{near} = 1.77 \times 10^{-3}$ mm^2 vs. $MSE_{full} = 1.83 \times 10^{-3}$ mm^2), while using a much smaller total set of model predictors (Table 3.4). The RF_{far} model displays the poorest overall performance, with an approximate 19% drop in accuracy compared to using the full set of predictors.

These predictor importance scores also provide new insights into the regions of the precipitating atmospheric column which are most significant for predicting surface snow accumulation. We note in Fig 3.6.a, along with the results of the importance scores from RF_{near}, that the lowest 45 reflectivity bins (lowest 2 km of the vertical column) have a significant impact on model skill and are used in nearly two thirds of all decisions. Meanwhile, reflectivity bins above 2 km are used much more rarely within the model, with a combined total importance of less than 22%, spanning 176 bins (approximately 8 km) of the vertical column. These importance scores also closely correspond with the vertical profile reflectivity distributions of Figure 3.6.a, where there exists little activity in reflectivity values above this range and the most complex reflectivity distributions appear close to the surface. Similar importance information can be also extracted from the atmospheric temperature data, however these feature importances are typically dwarfed by the contributions from near-surface reflectivities.

Examining distinct combinations of adjacent bins along with their relative positions within the vertical profile provides us with a better understanding of how model performance varies when

trained on different subsets of the atmospheric column. Sensitivity tests varying the number of reflectivity predictors in the RF model, along with their relative position within the vertical profile, are shown in Figure 3.7 for a) MSE, b) Pearson correlation, and c) model bias. Each color represents a different number of adjacent predictors used in the model (i.e. red represents models trained using a sole predictor that is derived from a single bin, while dark purple represents models trained on 64 individual predictors composed of adjacent reflectivity bins in the vertical column). We note a general trend of improved performance for larger groups of adjacent predictors closer to the surface. These larger predictor-set near-surface models have similar performance to the RF_{near} model, which was generated by combining information from the most significant predictor bins.

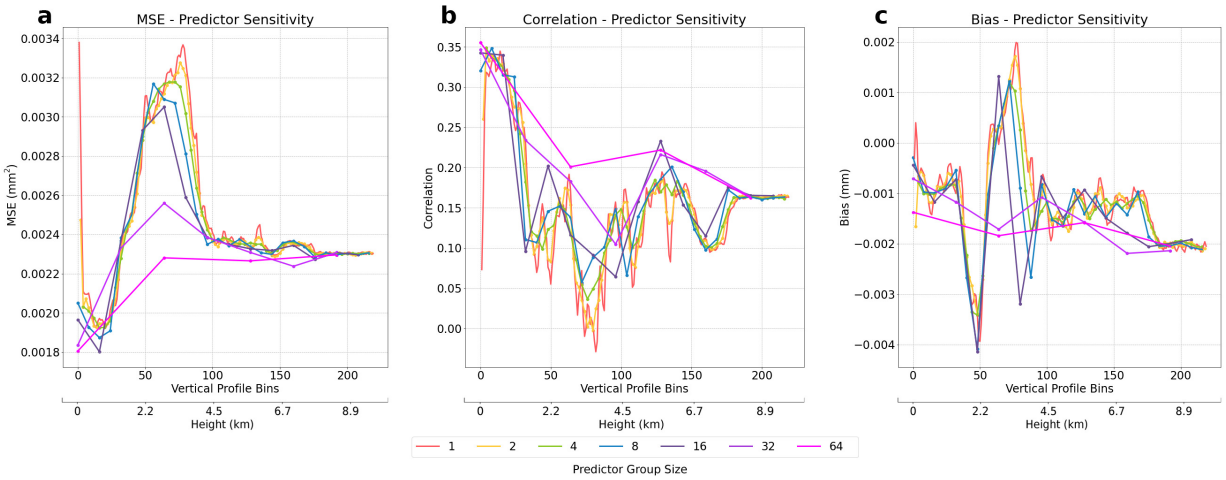

Figure 3.7: **Sensitivity analysis produced by varying RF predictor group sizes of adjacent bins, along with the locations of said predictor groups within the vertical profile** for: **a)** MSE; **b)** Pearson correlation; and **c)** mean bias ($model - observed$). Each color represents a different number of predictors used to train the model at different heights in the vertical column.

Model performance typically degrades and eventually converges to an MSE of 2.3×10^{-3} after the 100 vertical bin range (4.5 km above the surface). Another significant feature of the

Table 3.4: Summary of RF models and their respective performance as a function of vertical extent.

Name	Vertical Extent	MSE (mm^2)
RF_{near}	$1-45$	1.77×10^{-3}
RF_{far}	$46-221$	2.22×10^{-3}
RF_{full}	Full Profile	1.83×10^{-3}

sensitivity analysis is the parabolic dip in performance for bin group sizes below 32 predictors between the 40 and 90 bin ranges ($2-4$ km range), where we see substantial losses in model accuracy below that of even the models trained only on the uppermost bins in the atmospheric column near 10 km. While it is not entirely clear as to the cause of this performance dip in this region of the vertical profile, we believe it may be related to the fact that this region is typically between areas of shallow-cumuliform precipitation and high cloud precipitation in our training data (as seen in Figure 3.2) and therefore training only on these bins provides little information as to the actual precipitation occurring within the cloud. Spaceborne radar retrievals from satellites like CloudSat and GPM typically are blind to the lowest 1.5 kilometres of the atmosphere due to ground clutter interference, however Figure 3.2 shows that models trained on multiple bins within this near-surface range (bins $1-45$) appear as the most significant contributors to improved predictions of surface snow accumulation. The relative performance of each of the RF models examined in this sensitivity analysis can be used to help inform current and future retrieval methods that relate a reflectivity to snowfall rate. The information encoded in multiple, adjacent near-surface bins provides a more robust model when compared with single-bin estimates or full column averages.

3.5 Discussion and Conclusions

The VertiX radar and collocated in situ snowfall measurements from GCPEx have allowed us to derive a novel relationship between atmospheric reflectivity profiles and surface snow accumulation using ML-based regression. The accuracy of traditional $Z_e - S$ power law relationships was examined when applied to the VertiX data at GCPEx, and the RF continually demonstrated the highest skill, with the $Z_e - S$ relationships displaying clear biases in accumulation. RF predictor importance scores indicated that near surface bins in the lowest 2 km of the atmospheric column are by far the most influential regions of the vertical profile in quantifying surface snow accumulation. Sensitivity tests comparing different combinations of RF predictors suggest that the information found within groups of multiple, adjacent near-surface bins produce a model that better understands the relationship between reflectivities and accumulation. This information is highly relevant for improving current surface snowfall retrieval algorithms from spaceborne instruments, which also rely on a vertical column of reflectivity values to derive snowfall estimates for each atmospheric bin. These retrievals then extrapolate the snowfall rate from a single low-lying precipitating bin down to the surface, however this region of the vertical column is partially masked in spaceborne radar retrievals due to ground clutter attenuation in the radar signal, contributing additional uncertainty to the true location of the lowest precipitating cloud layer.

The small sample available from GCPEx is a major limiting factor in model performance and robustness. While we have demonstrated improved performance in the ML regression models, and tested the RF model against completely unseen data at ICE-POP, the highly variable nature of precipitation coupled with the effects of wind interference, the potential for mixed-phase pre-

cipitation occurrence, along with varying cloud layer precipitation structures, all contribute to increased model error and uncertainty. The vertical profile shown in Figure 3.2.12 for instance, depicts a variably-structured, high cloud storm that is unlike the other shallow-cumuliform snowfall events being presented to the model as training data. An examination of historical in situ 2 meter temperature observations also shows that surface temperatures fluctuate between $\pm 1^{\circ}$ Celsius for most of this day, suggesting that mixed-phase precipitation may also have been present and contributing additional uncertainty to an already complex synoptic event. We note that these periods exhibit consistently high retrieval errors for both the RF and $Z_e - S$ relationship (Fig. 3.5.b), with the RF typically showing slightly reduced MSE values compared to the power law relationships. We experimented with various methods of encoding cloud shape information and storm event types as predictors within the RF model based on similar work by Pettersen et al. (2020b) to help the RF better recognize unique snowfall event structures, however model performance remained largely unchanged. A larger and more representative training dataset would be beneficial in teaching the RF to better identify different storm events types. Automated storm-type identification could then improve the overall model skill in predicting surface snow accumulation for a wider array of precipitation events under varying environmental and atmospheric conditions.

Although multiple layers of wind shielding were employed at both GCPEx and ICE-POP, along with our attempts to examine days with only low-to-moderate wind speeds, gauge undercatch may have also been a factor contributing to error in our model as the Pluvio2 data acts as the ground truth we are attempting to predict. This is especially relevant when we consider cases of trace precipitation at one-minute temporal resolution, which have been shown in previous work by Colli et al. (2020) to be challenging to accurately quantify and correct for as a result of its low intensity. Errors in our response variable (which represents what we consider as the true accumulation

occurring on the surface) would contribute to further biases in our ML regression model and further obfuscate the locations of the most important bins in the vertical profile. Uncertainties in the VertiX radar estimates should also be considered as potential contributors to model error. As previously highlighted, Figure 3.2 displays a relatively uniform low-intensity near-surface reflectivity band over all days at GCPEx. It is unclear as to the origins of this feature from the VertiX retrievals, but its presence may impact the accuracy of the RF model (along with the relative importance of these bins) when predicting near-surface snowfall in this region. However, the reflectivity data from ICE-POP does not display these phenomena and the RF continues to display similar levels of skill when tested against this unseen data, which suggests that it is not training itself on this signal or statistical noise, and that its impact therefore appears negligible.

With these uncertainties in mind, we suggest that a followup study be completed which builds on the proof-of-concept methodology introduced in this work and incorporates radar data from multiple sites over an extended time period. An increased training sample would significantly contribute towards building a more robust model which could then be applied to independent, spaceborne reflectivity profiles in a wider context to derive a novel, ML-based hemispheric snow accumulation product.

Chapter 4

DeepPrecip: A deep neural network for precipitation retrievals

4.1 Overview

Remotely-sensed precipitation retrievals are critical for advancing our understanding of global energy and hydrologic cycles in remote regions. Radar reflectivity profiles of the lower atmosphere are commonly linked to precipitation through empirical power laws, but these relationships are tightly coupled to particle microphysical assumptions that do not generalize well to different regional climates. Here, we develop a robust, highly generalized precipitation retrieval algorithm from a deep convolutional neural network (DeepPrecip) to estimate 20-minute average surface precipitation accumulation using near-surface radar data inputs. DeepPrecip displays high retrieval skill and can accurately model total precipitation accumulation, with a mean square

error (MSE) 160% lower, on average, than current methods. DeepPrecip also outperforms a less complex machine learning retrieval algorithm, demonstrating the value of deep learning when applied to precipitation retrievals. Predictor importance analyses suggest that a combination of both near-surface (below 1 km) and higher-altitude ($1.5-2$ km) radar measurements are the primary features contributing to retrieval accuracy. Further, DeepPrecip closely captures total precipitation accumulation magnitudes and variability across nine distinct locations without requiring any explicit descriptions of particle microphysics or geospatial covariates. This research reveals the important role for deep learning in extracting relevant information about precipitation from atmospheric radar retrievals.

4.2 Introduction

Accurate estimates of surface precipitation are highly sought-after as they inform flood forecasting operations, water resource management practices and energy planning (Buttle et al., 2016; Gergel et al., 2017). Due to the sparse nature of in situ precipitation measurement networks, remote sensing has become a prominent alternative source of observations for deriving surface precipitation estimates (Liu, 2008b). Ground-based scanning radars are valuable resources as they provide estimates of precipitation over a wider area and at a higher temporal resolution compared to traditional in situ gauges (Lemonnier et al., 2019). Additionally, the size and availability of both vertically pointing and space-borne remote sensing datasets have expanded greatly in recent decades as a result of technological instrument improvements and new satellite missions (Quirita et al., 2017).

Remotely-sensed radar observations used in empirical, power-law relationships can relate

radar reflectivity (RFL) estimates (Z_e) to surface snowfall (S) or rainfall (R) rates (Eq. 4.1) (Matrosov et al., 2008; Kulie and Bennartz, 2009; Schoger et al., 2021).

$$Z_e = a \times (S/R)^b \tag{4.1}$$

These radar-based retrievals are powerful tools for filling current observational gaps and have been applied to great effect in previous literature (Levizzani et al., 2011; Hiley et al., 2010). However, these relationships demonstrate an inability to generalize well to unseen validation data as a consequence of the microphysical particle assumptions (e.g. shape, diameter, particle size distribution (PSD), terminal fall velocity and mass) used in each relationship's unique derivation (Jameson and Kostinski, 2002).

Recent machine learning (ML) approaches have demonstrated improvements in estimating surface precipitation from remotely-sensed data compared to traditional nowcasting methods (Shi et al., 2017; Kim and Bae, 2017). Deep learning models have benefited greatly from the increased observational sample provided by remote sensing missions and have shown skill in learning complex spatiotemporal characteristics of the underlying datasets (Chen et al., 2020b). However, a deep learning convolutional surface precipitation retrieval using vertical column radar data with no spatiotemporal covariates has yet to be developed to our knowledge. Previous ML studies have typically focused on passive microwave and infrared datasets which lack a detailed analysis of the vertical column structure, or suffer from a limited sample for model training across multiple, distinct regional climates (Xiao et al., 1998; Adhikari et al., 2020b; Ehsani et al., 2021).

In this work, we evaluate the abilities of a novel deep learning precipitation retrieval algorithm trained on vertically pointing radar (up to 3 km above the surface). The regression model we

present (DeepPrecip) is a hybrid deep learning neural network consisting of a feature extraction convolutional neural network (CNN) front-end and a regression feedforward multilayer perceptron (MLP) back-end. The combination of these two architectures allows DeepPrecip to recognize and learn the nonlinear relationships between different layers in the vertical column of radar observations and produce an accurate surface precipitation estimate. Through an analysis of feature input combinations, DeepPrecip performance is examined to identify regions within the vertical column that contain the most important contributions to retrieval accuracy (Lundberg and Lee, 2017). The relationships that exist between different layers of the vertical profile (and each atmospheric covariate) can be used to help inform current and future active radar retrievals of surface precipitation.

4.3 Data

4.3.1 Study Sites

In situ data were collected from 9 study sites (Fig. 4.1.a) from 2012-2020 (Table 4.1). Colored markers in Figure 4.1.b indicate periods where non-zero surface precipitation was recorded. Study sites were selected based on the required presence of a micro rain radar (MRR) and collocated Pluvio2 weighted precipitation gauge. Rain, snow and mixed-phase precipitation were recorded, with each site's precipitation phase and intensity distribution of observations differing based on the regional climate. For instance, Marquette experienced strong lake-effect snowfall while Cold Lake received mostly light, shallow snowfall. Further, due to the warmer temperatures recorded at OLYMPEx, these sites were classified as primarily experiencing liquid precipitation, while

ICE-POP received only solid precipitation.

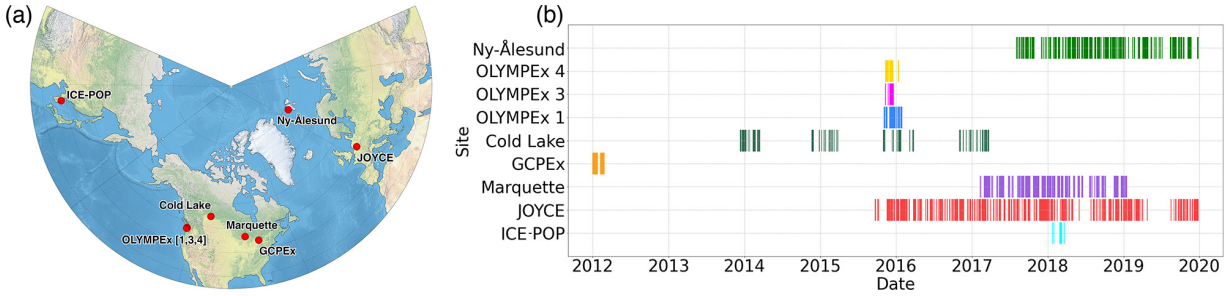

Figure 4.1: **Observational input data locations and temporal coverage periods. a),** Geographic study site locations. **b),** timeline of observational coverage (periods of active precipitation) for each site from 2012 to 2020.

4.3.2 Pluvio2 precipitation weighing gauge

Reference surface precipitation observations were collected by OTT Pluvio2 weighted gauges at each site. The Pluvio2 gauge records the precipitation accumulation from falling hydrometeors with a minimum time resolution of 1 minute (Colli et al., 2014). It includes a 200 cm^2 heated surface orifice (400 cm^2 at Ny-Å lesund) to prevent snow and ice buildup, along with site-specific wind shielding implemented as described in Table 4.1. These fence setups include a Double Fence Intercomparison Reference (DFIR) shield, which is a large, double fenced wooden structure which helps significantly reduce the impact of wind on surface precipitation measurements (Rasmussen et al., 2012; Kochendorfer et al., 2022). The Alter shield system consists of multiple freely hanging, spaced metal slats around the gauge top opening, which also helps mitigate undercatch issues during strong winds (Colli et al., 2014). Sensitivity analyses of different rolling temporal windows indicated an optimal temporal resolution of 20-minute non-real time accumulation (measurement results 5 minutes after precipitation accumulation), with minimum observational

thresholds of at least 0.2 mm over the course of an hour from the Pluvio2 gauge.

Table 4.1: Summary of in situ study site locations, identifiers, and observational details.

Site	ID	Lat	Lon	Elev.	Sample (N)	Shielding	Source
Ny-Ålesund	0	78.92	11.92	11	19068	Alter	(Schoger et al., 2021)
ICE-POP	1	37.67	128.7	789	1705	DFIR	(Kim et al., 2021)
GCPEx	2	44.23	-79.78	252	2314	DFIR	(Skofronick-Jackson et al., 2015)
Marquette	3	46.53	-87.55	430	8369	Alter	(Pettersen et al., 2020b)
OLYMPEx 4	4	47.39	-123.87	2155	6444	None	(Houze et al., 2017)
OLYMPEx 1	5	47.5	-123.58	3340	9114	None	(Houze et al., 2017)
OLYMPEx 3	6	47.68	-123.38	2100	5727	None	(Houze et al., 2017)
JOYCE	7	50.9	6.4	95	43579	Alter	(Lahnert et al., 2015)
Cold Lake	8	54.4	-110.26	541	1692	Alter	(Boudala et al., 2021)

4.3.3 Micro rain radar

Vertical pointing MRRs (developed by METEK) were located nearby the Pluvio2 gauges at each site to record complementary atmospheric observations. The MRR is a K-band (24 GHz) continuous wave Doppler radar which provides information related to hydrometeor particle activity up to 3.1 km above the surface (or 1 km for Ny-Ålesund) as a function of spectral power backscatter intensity. The MRR provides 29 vertical bins (of size 100 m) spanning 300 m to 3100 m above the surface as shown for each site in Figure 4.2.a. Raw radar measurements were preprocessed using Maahn's improved MRR processing tool (IMProToo) for noise removal, dealiasing and for extending the minimum detectable dBZ to -14 which allows for improved measurements of solid precipitation. This data was then temporally averaged to align to the same 20-minute windows generated for the Pluvio2 observations and used as a model input (Maahn and Kollias, 2012).

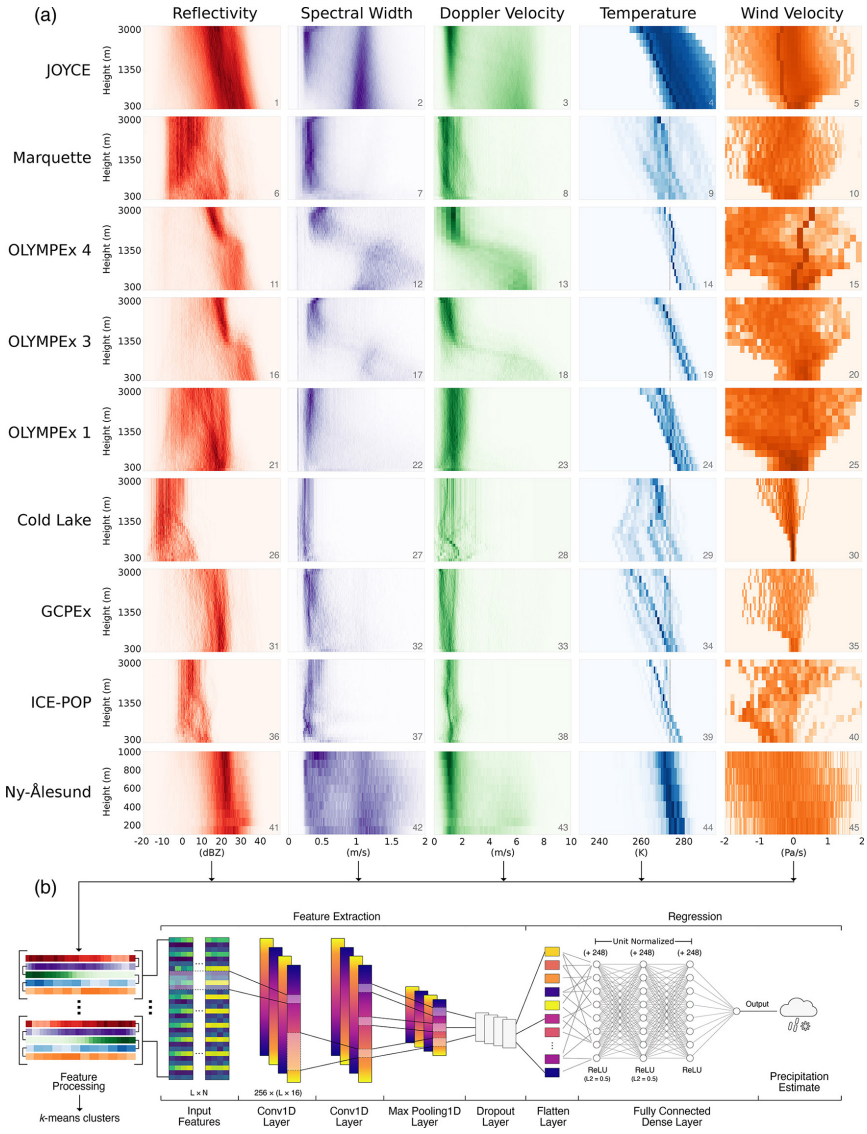

Figure 4.2: **DeepPrecip input covariates, feature processing pipeline and model architecture. a),** Site-predictor matrix of normalized MRR and ERA5 observational frequency histograms used in model training and testing. Note that darker colors in the 2D heatmaps indicate a higher frequency of observations. **b),** DeepPrecip convolutional neural network diagram for L inputs with N predictors.

4.3.4 ERA5

European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5) hourly temperature (TMP) and vertical wind velocity (WVL) on pressure levels from 0 to 3 km were also included as additional input covariates to DeepPrecip (Hersbach et al., 2020). These inputs allow the model to more accurately recognize different precipitation event structures, large-scale atmospheric dynamics and hydrometeor phases during training. Note that WVL units (Pa/s) are defined using the ECMWF Integrated Forecasting System (IFS) which adopts a pressure based vertical co-ordinate system (i.e. negative values indicate upwards air motion, since pressure decreases with height). Each of these variables were linearly interpolated to align with the MRR data over 20 minute intervals and at 100 m vertical resolution.

4.3.5 Surface meteorology

Collocated surface temperature (degrees Celsius (° C)) and 10-meter wind speed (m/s) meteorologic observations were also collected from instruments installed at each site and temporally aligned to the Pluvio2 and MRR datasets. Surface wind data acts as an additional observational constraint for mitigating the effects of undercatch on unshielded measurement gauges (Rasmussen et al., 2012). Undercatch occurs when precipitation falling in the presence of wind can cause hydrometeors to pass over the gauge top orifice. This effect has been shown to bias reported precipitation quantities by up to 10% (Ehsani and Behrangi, 2022). We therefore limit the available training dataset to periods when surface wind speeds are $<$ 5 m/s, as this restricts the analysis to low-medium wind speed events at each location to maintain a high gauge-catch efficiency (Yang, 2014). This preprocessing step reduces the average size of our total observational pool by 16%

across all stations, however, we note that maximum intensity precipitation events are not removed using this technique.

Surface meteorologic station temperature data is used for precipitation-phase partitioning at 5° C to allow for $Z_e - S/R$ comparisons with DeepPrecip. Additional dry surface air temperature thresholds of 0°, 1° and 2° C were also examined, but $Z_e - S/R$ performance for both rain and snow appeared optimal when classified using a 5° C threshold (where temperatures < 5° C are considered as solid precipitation and temperatures >= 5° C are considered as rainfall). This simple temperature threshold is an additional source of uncertainty in our comparisons with the $Z_e - S/R$ relationships due to the influence of mixed-phase precipitation on power law accuracy, along with uncertainties in the location of the active melting layer (Jennings et al., 2018). A more sophisticated phase partitioning system (e.g. using wet-bulb temperature as described in Sims and Liu (2015)) could also be linked to DeepPrecip as an additional predictor to further improve classification of mixed-phase precipitation in future work.

4.4 Methods

4.4.1 Radar-precipitation power laws

Relating radar reflectivity observations to surface accumulation has been done extensively in past surface and spaceborne radar missions through $Z_e - S/R$ power law relationships (Skofronick-Jackson et al., 2017; Liu, 2008b). These power law relationships are empirically defined by relating reflectivity values in a near surface bin to observed surface accumulation under a set of assumed particle microphysics (e.g. size, shape, density and fallspeed) (Matrosov et al., 2008).

While these techniques have been used to great success in previous studies by Schoger et al. (2021) and Levizzani et al. (2011), the assumptions about snowfall and rainfall particle microphysics makes the generalization of these power laws less robust, which contributes to high uncertainty when applied across large areas with unique regional climates (Jameson and Kostinski, 2002).

We examine an ensemble of 12 Ka- and K-band $Z_e - S/R$ relationships in this work to compare with model output from DeepPrecip (Table 4.2). As a consequence of the short temporal period (20 minutes) used in this analysis, MSE values are typically small (< 0.1 mm^2). Each $Z_e - S/R$ relationship was applied to a near-surface bin in the reflectivity profile (bin 5 for DP_{full} and DP_{near}, and bin 11 for DP_{far}) to derive a corresponding surface precipitation estimate. These bins were selected based on a sensitivity analysis where we examined the performance of multiple near-surface high-importance regions of the vertical column (not shown). The best performing regions were identified as the above bins (5 and 11) based on the respective region of the vertical column being considered (near or far). More information regarding the derivation of each $Z_e - S/R$ relationship can be found in Table 4.2.

To further evaluate the performance of DeepPrecip, we also include model comparisons to a set of six site-derived $Z_e - P$ (reflectivity precipitation) power law relations. Each $Z_e - P$ relationship is empirically derived from the collocated MRR and Pluvio data at each observational site examined in this work (excluding Cold Lake and Ny-Ålesund due to the limited available sample and vertical extent of each site, respectively). Each $Z_e - P$ relation is fit via a non-linear least-squares approach for finding optimal a and b coefficients in Eq. 4.1 using SciPy's $curve\_fit$ optimization algorithm (Virtanen et al., 2020). Each $Z_e - P$ relationship was then applied to bin 5 reflectivities at each site (i.e. the same process as is used for $Z_e - S/R$ relationships) and compared with in situ observations to assess their general accuracy.

Table 4.2: Details for each multi-phase precipitation power law relationship.

Phase	Name	Source	Power Law	Reference
Solid	AVE_K	K	$Z_e = 77.61 \times S^{1.22}$	(Schoger et al., 2021)
	KB09sp	Ka	$Z_e = 19.66 \times S^{1.47}$	(Kulie and Bennartz, 2009)
	KB09ag	Ka	$Z_e = 313.29 \times S^{1.85}$	(Kulie and Bennartz, 2009)
	KB09br	Ka	$Z_e = 24.04 \times S^{1.51}$	(Kulie and Bennartz, 2009)
	M07	Ka	$Z_e = 56.00 \times S^{1.20}$	(Matrosov, 2007)
	S17	K	$Z_e = 18.00 \times S^{1.10}$	(Souverijns et al., 2017)
Liquid	BP09h	K	$Z_e = 32.00 \times R^{3.30}$	(Van Baelen et al., 2009)
	BP09m	K	$Z_e = 324.00 \times R^{2.40}$	(Van Baelen et al., 2009)
	MP48	–	$Z_e = 200.00 \times R^{1.60}$	(Marshall and Palmer, 1948)
	J19bb	K	$Z_e = 367.00 \times R^{1.37}$	(Jash et al., 2019)
	J19nbb	K	$Z_e = 211.00 \times R^{1.44}$	(Jash et al., 2019)
	J19hr	K	$Z_e = 168.00 \times R^{1.40}$	(Jash et al., 2019)

4.4.2 Neural network architecture

DeepPrecip is a feedforward convolutional neural network that takes as input a vector of 115 atmospheric covariates (Table 4.3), performs a feature extraction of the vertical column and outputs a single surface precipitation estimate using a fully connected multilayer perceptron. While the structure of this final version of DeepPrecip is complex, the retrieval evolved from a much simpler initial state based on a multiple linear regression (MLR) model. Due to clear nonlinearities between observed reflectivity data and surface precipitation accumulation, the MLR model was unable to capture in situ variability and provided estimates near the mean accumulation value. Similar radar-based precipitation retrieval studies by Chen et al. (2020a) and Choubin et al. (2016) have demonstrated much better performance using an ML-based approach, which led to the development of a random forest (RF) model, an MLP and finally the CNN.

Table 4.3: Summary of DeepPrecip full vertical column model input covariates.

Predictor	Abbreviation	Count	Units	Source	Type
Reflectivity	RFL	29	dBZ	MRR	float64
Doppler velocity	DOV	29	m/s	MRR	float64
Spectral width	SPW	29	m/s	MRR	float64
Temperature	TMP	12	K	ERA5	float64
Wind velocity	WVL	12	P_a/s	ERA5	float64
Profile group	PG	4	Indicator	K-mean	Boolean

The 1D convolutional layers perform a feature extraction of the vertical column of inputs to reduce the total number of parameters being fed into DeepPrecip's fully connected dense layers. This 1D-CNN structure can identify relationships within the vertical column, save on memory and lower computational training time requirements. To estimate a surface precipitation rate, the forward propagation step between the previous layer $(l-1)$ to the input neurons of the current layer (l) in the MLP are expressed in Eq. 4.2 (Abdeljaber et al., 2017).

$$x_k^l = f(b_k^l + w_{ik}^{l-1} \times \sum_{i=1}^{N_{l-1}} Conv1d\ (s_i^{l-1})) \tag{4.2}$$

Where k and l refer to the k^{th} neuron for layer l with x as the resulting input and b as the scalar bias. s and w terms represent the neuron output and kernel weight matrix, respectively, from the i^{th} neuron of layer $l-1$ (and to the k^{th} neuron of layer l for w). The function '$f()$' represents the activation function used to transform the weighted sum into an output to be used in the following network layer, and the $Conv1d$ term represents the 1D convolved output from the CNN.

The RF model tested in this study was based on previous work from King et al. (2022a) where

an RF was used to retrieve surface snow accumulation from a collocated X-band and Pluvio2 instrument at a single experiment site (GCPEx). The RF developed in said study demonstrated good skill in estimating surface accumulation, and so we incorporate the same model here (retrained on the MRR and ERA5 data from this study) as a baseline comparison to other ML retrieval methods (i.e. DeepPrecip).

The final DeepPrecip model structure is outlined in Figure 4.2.b. It includes two 1d-convolutional layers, a 1d max pooling layer, dropout layer, flattening layer and concludes in a dense MLP regressor with 3 hidden layers. The total number of trainable model parameters in DeepPrecip is 3,937,793. Model training and testing was performed using a 90/10 (non-shuffled) split on each site to generate training and testing datasets for each location. As an additional preprocessing step, we standardize all input covariates to remove the mean and by scaling inputs to unit variance. The non-shuffled nature of this splitting process allows for DeepPrecip estimates to be validated against unseen data and prevents overfitting from training on temporally autocorrelated vertical column inputs. Additionally, this stratified selection process guarantees that an equal percentage of data is included from each site during training.

Retrieval accuracy is primarily assessed using a mean squared error (MSE) skill metric calculated between each model's estimated surface accumulation values and the total Pluvio2 non-real-time reference accumulation observations over 20 minutes. Performance statistics are reported from the average skill of the test portion of a non-shuffled 90/10 train/test CV split (i.e. DeepPrecip trained and tested 10 times on different contiguous portions of the full available sample). Note that each split is stratified to include 10% of each station's sample in every test split. Uncertainty estimates are calculated from running each CV split 50 times using dropout to gain additional insight into model variability (resulting in 500 total model instances). The dropout

layers simulate training numerous models with differing architectures in a highly parallelized manner by randomly deactivating (or dropping) a certain fraction of nodes within the network to provide a distribution of retrieval estimates.

4.4.3 Hyperparameter optimization

DeepPrecip was developed, trained and optimized on Graphcore intelligence processing units (IPUs) MK2 Classic IPU-POD4 which significantly sped up the training time by a factor of 6.5 compared to a state-of-the-art Nvidia Tesla V100 GPU (Louw and McIntosh-Smith, 2021). Additional training throughput comparisons are included in Table 4.4. Training was completed using a combination of open-source Python packages including Keras, Tensorflow and scikit-learn. An extension of stochastic gradient descent known as Adam optimization (adaptive moment estimation) is used to continually update internal network weights in the model during training to minimize a standard MSE loss function (Eq. 4.3) and track model learning over time.

$$L(x, y) = \frac{1}{D} \sum_{i=1}^{D} (x_i - y_i)^2 \tag{4.3}$$

Hyperparameters do not change value during training (in contrast to model parameters like internal node weights), but they play a critical role in the neural network learning process to map input features to an output. Selecting optimal hyperparameter values is an important part in constructing a model which minimizes loss, improves model efficiency and quality, and mitigates overfitting. Multiple steps were taken to address concerns of model overfitting. In addition to the use of non-shuffled training, we employ multiple regularization methods including early stopping,

Table 4.4: DeepPrecip model training throughput comparisons running on Tensorflow (v2.4.3) using a batch size of 128 samples on different hardware. Note that 2 IPUs were used in comparison to 1 GPU/TPU to equalize average computation costs when training DeepPrecip using each piece of hardware.

Hardware	Processors	Samples/second
Graphcore Intelligence Processing Unit (IPU)	2	500
Nvidia Tesla V100 Tensor Core GPU	1	77
Google Tensor Processing Unit (TPU)	1	56
Nvidia Tesla K80 GPU	1	23

dropout, the application of layer weight constraints and L2 regularization (details in Table 4.5). L2 regularization (or ridge regression) adds another penalty term to the MSE loss function, which helps to create less complex models when dealing with many input features to improve model generalization.

To select the optimal values for the aforementioned hyperparameters, and to optimize DeepPrecip's general structure, we use a form of hyperparameterization known as hyperband optimization (Li et al., 2017). Hyperband is a variation of Bayesian optimization which intelligently samples the parameter space to find hyperparameter values that minimize loss while learning from previous selections. Hyperband adds another component to the analysis by also slowly increasing the number of epochs run during each phase of the optimization process to sample in a more efficient manner. DeepPrecip hyperparameters were derived by running a 10-fold CV hyperband optimization continuously on a single Graphcore IPU for approximately two weeks. The final hyperparameter values (and their respective parameter search spaces) can be found in Table 4.5.

Table 4.5: DeepPrecip hyperparameters optimization details.

Hyperparameter	Value	Parameter Space
Activation	ReLU	['relu', 'tanh', 'sigmoid']
Batch Size	128	[64, 128, 256, 512]
Dropout Rate	0.1	[0.001, 0.01, 0.1, 0.25, 0.5, 0.75]
Early Stop Patience	8	[4, 8, 16, 32]
Epochs	512	[64, 128, 256, 512, 1024]
Filters	256	[4, 16, 64, 128, 256]
Hidden Layers	3	[1, ..., 20]
Kernel Size	16	[2, 4, 8, 16, 32]
L2 Regularization	0.5	[0.001, 0.01, 0.1, 0.5]
Learning Rate	1e-7	[0.001, 0.0001, 1e-5, 1e-7]
Loss Function	MSE	['MSE']
Neurons	256	[64, 128, 256, 512, 1024]
Optimizer	Adam	['Adam']
Pool Size	2	[2]

4.4.4 Unsupervised classification layer

An unsupervised k-means clustering preprocessing step is also applied using MRR reflectivity profiles as input to provide DeepPrecip with insights into distinct profile group (PG) vertical column structures (Fig. 4.2.b). Minimizing within-cluster sum of squares between each vertical column radar estimate results in $k = 4$ PGs being selected using the within-cluster-sum of squared errors elbow criterion method (Fig. 4.3). The elbow method is a clustering heuristic which allows for an optimal number of clusters to be selected as a function of diminishing returns of explained variation (i.e. finding the elbow or "knee of the curve"). K-means clustering was applied using Python's scikit-learn package on all input reflectivity data to generate four profile clusters, which were included as additional input parameters to DeepPrecip. These clusters are

useful for partitioning the precipitation data into groups based on different precipitation intensity-classes (trace, low, medium and high intensity) to identify where DeepPrecip finds the most important contributors to high retrieval accuracy for each category of storm intensity. Derived cluster groups are useful for interpreting feature importance values from model output (Section 4.5.2).

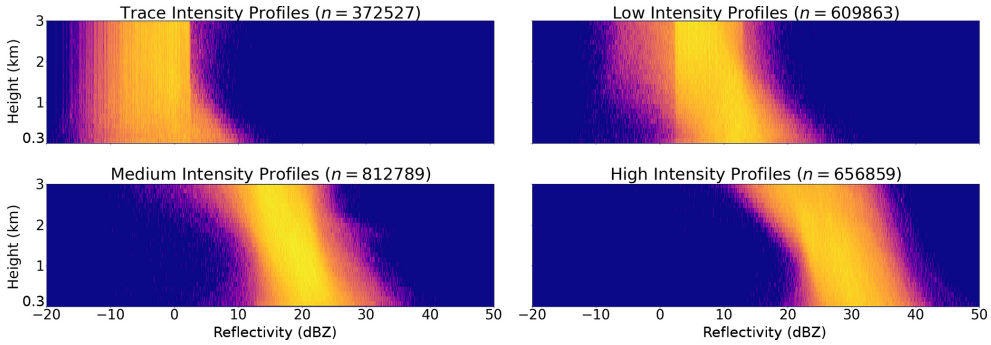

Figure 4.3: **K-means cluster reflectivity intensity-classes of vertical profiles from the MRR instruments at all sites.** A total of 2452038 vertical profiles are organized by reflectivity intensity (dBZ) into $k = 4$ precipitation intensity subsets.

4.5 Results

4.5.1 DeepPrecip retrieval performance

We first examine the differences in performance between DeepPrecip and an RF that has demonstrated good performance in our previous work (not shown) to assess the capabilities of a less-sophisticated ML-based approach over a CNN. DeepPrecip demonstrates improved skill in capturing most of the peaks and troughs in observed precipitation variability (Fig. 4.4.a). These differences are most clearly demonstrated in Figure 4.4.a at OLYMPEx and JOYCE, where DP

more accurately predicts Pluvio2 precipitation extremes compared to the RF. Both models appear to struggle in capturing accumulation intensities during periods of mixed-phase precipitation when temperatures are near 0° C (i.e. Marquette, JOYCE and the tail end of OLYMPEx 1) due to a lack of training data with similar climate conditions and the complex nature of such events. DP does demonstrate improved skill at capturing light intensity precipitation at the beginning of the JOYCE period (compared to the RF), however this is with some uncertainty as noted by the wider shaded region (1 standard deviation). Performance statistics (Fig. 4.4.b) summarize these improvements, with DeepPrecip showing MSE values 21% lower and r^2 values 34% higher (significant at $\alpha < 0.05$) compared to the RF.

Total cumulative surface accumulation comparisons between DeepPrecip and each $Z_e - S/R$ relationship are then examined in Figure 4.4.c for both rain and snow. To examine model skill across different precipitation phases, a simple temperature threshold is imposed, where retrievals recorded during periods with temperatures below 5° C are classified as snow and periods equal to or warmer than 5° C as rain. DeepPrecip more accurately captures surface precipitation quantities when compared to the $Z_e - S/R$ estimates, with a total accumulation curve similar in shape to that of in situ, indicating that DeepPrecip more closely captures the observed precipitation variability and magnitude. Log-scale MSE statistics are calculated between each model and in situ records in Figure 4.4.d and indicate that DeepPrecip consistently outperforms traditional $Z_e - S/R$ power-law methods by 187% on average. As a general precipitation retrieval algorithm, we do not explicitly train a DP_{snow} and DP_{rain} model for different precipitation phases with unique regional atmospheric microphysical conditions. While the $Z_e - S/R$ models shown in Figure 4.4.c/d are bespoke for rain or snow, DeepPrecip is trained on all data with no a priori knowledge of the underlying physical precipitating particle state.

DeepPrecip estimates of accumulated rain display a lower MSE than that of snow (Fig. 4.4.d). We believe these differences to be twofold: 1) the larger sample of rainfall events in the training data (3 times that of snowfall); and 2) the more complex nature of snow particle microphysics. Unlike the uniform properties of a rain droplet, the shape, size and fallspeed of solid precipitation is much more dynamic and challenging to model (Wood et al., 2013). Continued issues with interference from wind may have also impacted the accuracy of in situ measurements of snow accumulation, leading to higher uncertainty and error (further discussions on these uncertainties in Section 4.6) (Kochendorfer et al., 2017). To visualize the range in uncertainty from the CNN model estimates, we display confidence intervals showing 1 standard deviation in Figure 4.4.b/d from 50 DeepPrecip model realizations using dropout. Both ML-based models exhibit the highest uncertainty during periods of mixed-phase precipitation at GCPEx and Marquette, along with high intensity precipitation at OLYMPEx.

To further evaluate DeepPrecip's retrieval skill over traditional methods, we compare model performance to a set of six custom $Z_e - P$ site-derived power laws (derivation details in Section 4.4). While $Z_e - P$ relationships typically perform well in the regional climate under which they were derived, they do not generalize well outside said climate. This lack of robustness is visible in the differences between in situ and $Z_e - P$ estimates of accumulation in Figure 4.5.a, where each $Z_e - P$ (light grey line) displays consistent positive or negative biases and no single power law captures the high variability in accumulation across multiple sites. For instance, OLYMPEx 1 and OLYMPEx 3-derived relationships produce a strong positive bias at JOYCE, and the JOYCE-derived $Z_e - P$ power law is quite negatively biased when applied at OLYMPEx. The mean of all six custom power laws is shown in bold grey, and while it closely captures total mean accumulation across all sites, it is unable to model the high variability in precipitation

intensity.

The resulting MSE from the application of each custom $Z_e - P$ relationship to each site (along with DeepPrecip) further demonstrates DeepPrecip's improved robustness (Fig. 4.5.b). In all other cases, DeepPrecip either outperforms all $Z_e - P$ power laws or is only slightly worse than the power law derived for the site in which it is being tested. On average, DeepPrecip retrievals result in 187% lower MSE values than all $Z_e - P$ site-derived power laws estimates when applied to the testing data across the full spatiotemporal domain (Table 4.6). Figure 4.5.b also displays a model intercomparison of each $Z_e - P$ relation, where we can clearly see how $Z_e - P$ relations like those derived at OLYMPEx 1 and 3 are clearly unable to capture the vastly different snowfall regimes at sites like ICE-POP, GCPEx and JOYCE with their much larger MSE values for these sites.

Table 4.6: MSE values (in e^{-3} mm^2) for all vertical extent experiments across all models for both solid and liquid precipitation.

Phase	Model	Mean Squared Error (e^{-3} mm^2)		
		Full Column	< 1 km	$1-3$ km
All	DeepPrecip	0.7	0.94	1.2
	RF	1.1	0.92	1.5
	$\overline{Z_e - P}$	20.3	20.3	21.4
Solid	DeepPrecip	1.2	1.5	2.2
	RF	2.9	1.5	4.2
	$\overline{Z_e - S}$	31	31	85
Liquid	DeepPrecip	0.43	0.47	0.85
	RF	0.5	0.53	0.6
	$\overline{Z_e - R}$	16.9	16.9	19.7

The robustness of DeepPrecip was further evaluated using a leave-one-out cross validation

(CV) for each site of training observations. This approach tests the skill of DeepPrecip at predicting precipitation for a location that was not included in the training data, which is a strong indicator of the generalizability of the model. Log-scale MSE results of this test for each site are shown in Figure 4.6 for each precipitation-phase subset, along with the corresponding average $Z_e - P/S/R$ estimate when applied at that site. These findings demonstrate similar performance to the baseline DeepPrecip model skill, which continues to outperform all traditional power law techniques on average. The large range in skill in the power law relationships at most sites (wide error bars) further demonstrates the relative lack of generalizability of $Z_e - P/S/R$ relationships to different regional climates. Further, the site-derived power law fits (grey dots) perform worse on average than DeepPrecip for locations that are close in proximity (i.e. the OLYMPEx sites).

Predictably, DeepPrecip performance degrades compared to the baseline model when the testing site is left out, since the model is no longer trained using data representing the regional climate of the site being tested. This difference in performance is most notable at the set of OLYMPEx sites, and while DeepPrecip performance is still improved over the $Z_e - S/R$ relationships, we note a substantial percentage increase in MSE (375% on average) at these locations. OLYMPEx measurements were the only observational datasets without any gauge shielding and which is a likely source of uncertainty, further contributing to this increase in error when the site is removed from the training set (Kochendorfer et al., 2022).

4.5.2 Quantifying sources of retrieval accuracy

Identifying regions within the vertical column that are the most important contributors towards retrieval accuracy is critical for informing future satellite-based radar precipitation retrievals.

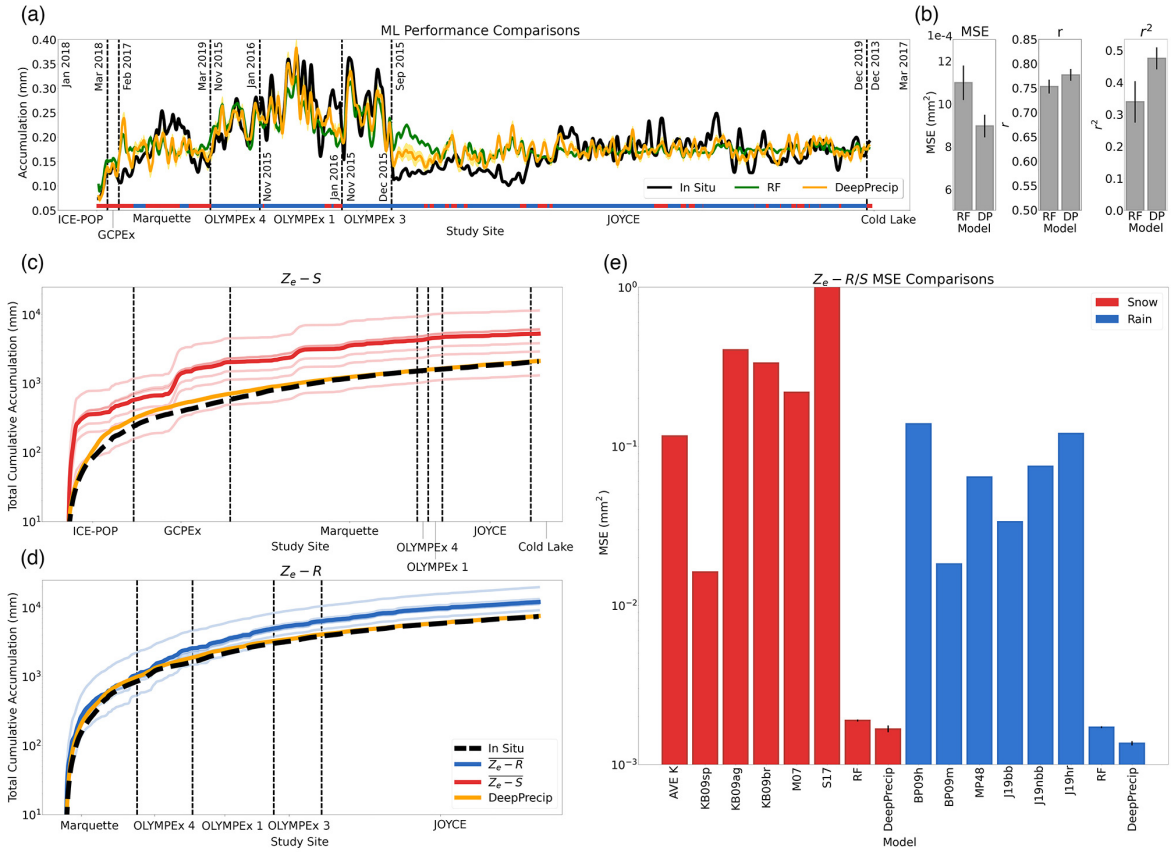

Figure 4.4: **Performance comparisons between DeepPrecip (DP), an RF and an ensemble of power law-derived retrievals of surface precipitation. a),** Running mean (window size 500 time steps) of accumulation for all sites with Pluvio2 measurements in black, RF estimates in green and DeepPrecip in yellow. Data is sorted by station and then time, with each station separated by a dashed vertical line. 1 standard deviation from 50 dropout runs per cross-validated instance is shown in the shaded regions (most notable at the start of JOYCE). **b),** performance statistics for RF/DeepPrecip accuracy including MSE, Pearson correlation (r) and r^2 with error bars showing 1 standard deviation. **c),** Timeseries of total accumulation estimates over the full observation period for all $Z_e - S$ relationships (individual red lines) and DeepPrecip. The mean of the $Z_e - S$ relationships is shown in bold. **d),** The same as in c) but for $Z_e - R$. **e),** Phase-partitioned log-scale MSE values between each model and in situ observations from 50 model realizations. Note that S17 MSE values extend beyond the top of the graph to 10^1 mm^2.

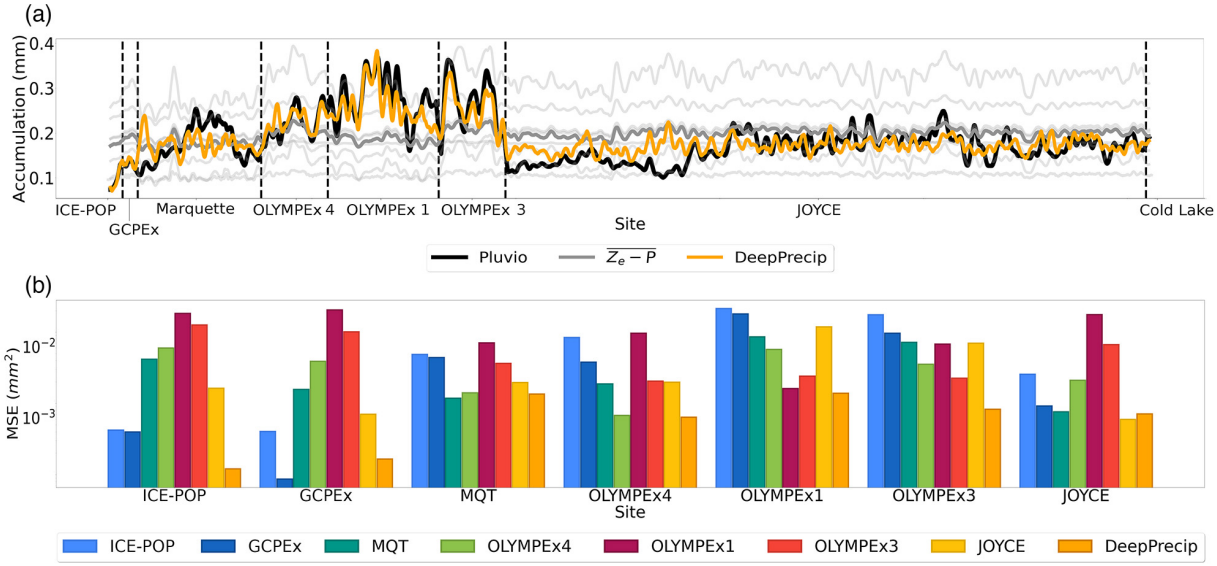

Figure 4.5: **Site-derived empirical $Z_e - P$ power law performance comparisons. a),** The same as Figure 4.4.a, except now using $Z_e - P$ relationships derived at each study site. **b),** MSE values for DeepPrecip and each $Z_e - P$ relationship when tested on each site.

The ground-based radar instruments used in this work do not suffer from the same ground clutter contamination issues typical of satellite-based radar observations, and we are therefore able to quantify the contributions to model skill arising from the included boundary layer reflectivity measurements in DeepPrecip. Separating the training data into three subsets based on vertical extent and generating new models with this data, allows us to examine changes in performance as a function of information availability. These subsets include: DP_{full} (all 29 vertical bins, i.e. the baseline model), DP_{near} (the lowest 1 km; 8 bins), and DP_{far} (1 – 3 km; 21 bins). DeepPrecip MSE results (Table 4.6) for each subset suggest that the information provided by a combination of both near-surface and far-profile data results in the highest accuracy.

Since Ny-Ålesund MRR observations were recorded with a maximal vertical extent of 1 km, they are only included in DP_{near}. Model skill when including/excluding Ny-Ålesund training data

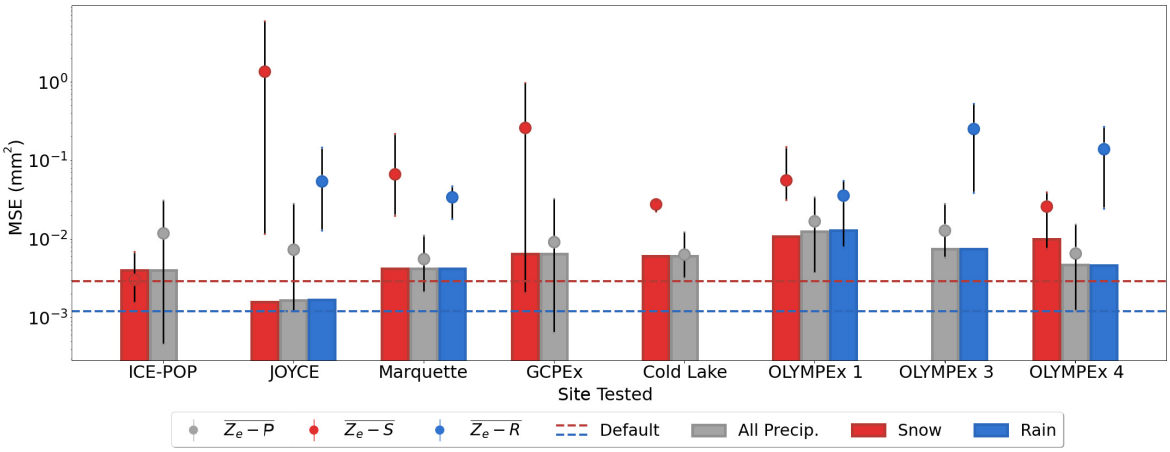

Figure 4.6: **Leave-site-out full column DeepPrecip performance robustness analysis.** Each bar represents a DeepPrecip full column log-scale MSE value when trained on all precipitation data excluding the noted site, and then validated against said excluded site (dashed line is the default DeepPrecip model with all sites). Each red and blue dot represents the average $Z_e - S/R$ relationship estimate tested in the same manner (error bars represent the min and max ensemble values). Gray dots represent the mean, min and max ensemble values from all site-derived $Z_e - P$ relationships (excluding the relationship derived from the site being tested), when applied to each site.

(19,000 samples) was examined to determine whether it was confounding comparisons between the aforementioned vertical profile subset models. The results of these tests suggested that the impact on overall performance is negligible across both precipitation phases when Ny-Ålesund is included or excluded in the training set.

Distributions of surface precipitation anomalies appear distinct for rain and snow (Fig. 4.7), with the full column model more closely capturing accumulation recorded by in situ gauges. Anomaly frequencies are derived by removing the mean accumulation estimate for each phase at each site. We attribute the structural differences between the anomaly distributions of of snow and rain to the more complex particle size distributions (PSDs) of snowfall coupled with the more variable particle water content of snow compared to that of rain (Yu et al., 2020). Additional uncertainties in the surface Pluvio2 measurement gauge observational records of snowfall due to gauge undercatch is another likely contributor to increased error (Kochendorfer et al., 2022). In Figure 4.7.a, both DP_{far} and DP_{near} exhibit higher anomaly values with a flattened curve top and heavy tails. Using a combination of information from both near and far bins reduce these biases and tightens each accumulation anomaly distribution around zero. A similar trend is also present for rain in 4.7.b, where we again most closely capture the in situ anomaly distribution using DP_{full}.

A major challenge in deep learning is interpreting model output. SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017), is a game theory approach to artificial intelligence model interpretability based on Shapley values that has previously been used to great effect in the Geosciences (Maxwell and Shobe, 2022; Li et al., 2022). Shapley values quantify the contributions from all permutations of input features on retrieval accuracy to identify which are the most meaningful. While computationally expensive (with exponential time complexity), this process

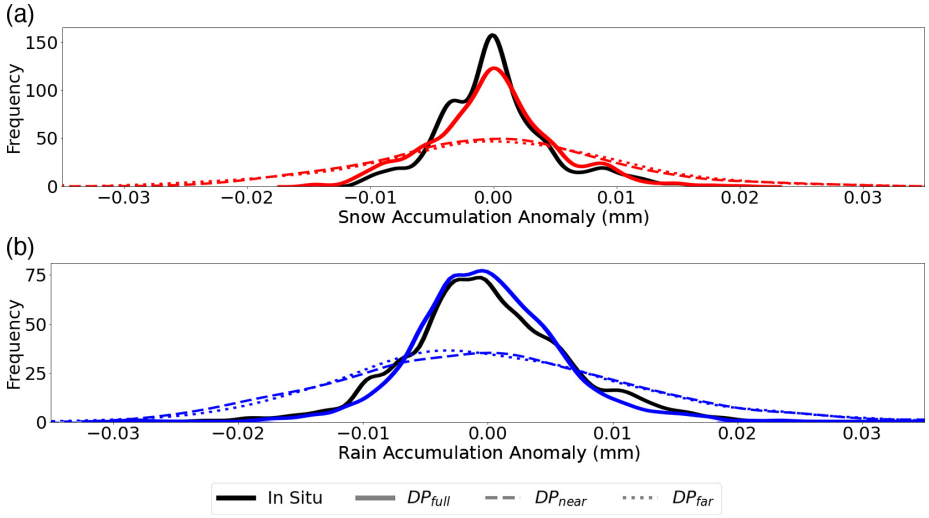

Figure 4.7: **Phase-partitioned surface precipitation accumulation anomaly frequency distributions.** DeepPrecip is trained and tested on three subsets of bins from the vertical column: DP_{near} (< 1 km), DP_{far} ($1-3$ km) and DP_{full} (the entire vertical column) for **a),** solid and **b),** liquid precipitation.

provides local interpretability within the model by examining how each possible combination of all input features impacts model accuracy (Jia et al., 2020). Here, the calculated Shapley values give insight into the regions of the vertical column that are contributing the most useful radar information in the precipitation retrieval.

Shapley values for the entire dataset used in DP_{full} indicate that the most important model predictors comprise a combination of both near-surface and far profile bins (Fig. 4.8). Reanalysis variable model inputs are generally the least influential, except for the trace precipitation case, where low-mid level TMP and WVL bins appear highly important (Fig. 4.8). In all cases, TMP and WVL decrease in importance as a function of height above the surface. DeepPrecip typically considers MRR-derived bins in the $1.5-2.5$ km range as the most important predictors. In non-trace intensity profiles, it is the 2 km region Doppler velocity (DOV) observations which

are the dominant contributing predictor. When we consider all profiles, reflectivity (the input to $Z_e - S/R$ relationships) is not necessarily the dominant feature, and it is a combination of $1.5-2$ km profile information from reflectivity, Doppler velocity and spectral width (SPW) that results in the highest model skill. Combinations of these regions within the vertical column appear to allow DeepPrecip to better understand precipitation events with complex cloud structures which would not necessarily be recognized by conventional $Z_e - S/R$ relations that primarily rely on information from a small subset of near-surface bins.

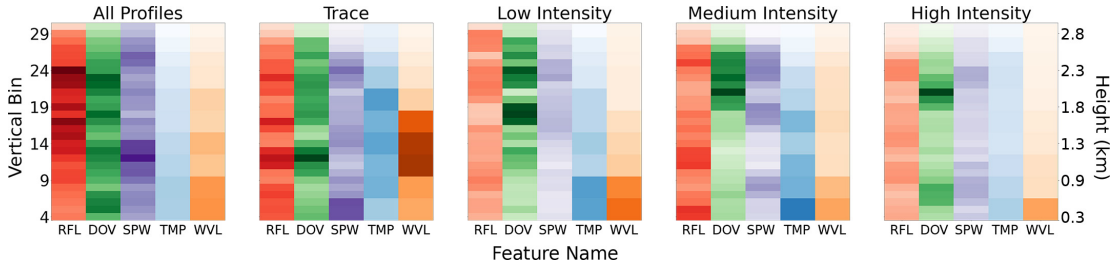

Figure 4.8: **Normalized vertical column Shapley global feature importance values (i.e. $|SHAP_{DP}|$).** Shapley output values are calculated for different subsets of vertical column reflectivities separated into all profiles, trace intensity, low intensity, medium intensity, and high intensity precipitation events based on a k-means clustering of input data (more in Section 4.4.2). Areas of dark color indicate a high feature importance at that location within the vertical column.

4.6 Discussion and Conclusions

DeepPrecip not only demonstrates considerable retrieval accuracy without the need for physical assumptions about hydrometeors or spatio-temporal information, but also provides insight into the regions of the vertical column which are most important for improving predictive accuracy. The results from Section 4.5.2 suggest that while the exact altitudes providing predictive information from the vertical column may shift up or down under different precipitation intensities, there

exists a consistent combination of both near-surface and far profile bins that always appear as highly important contributors to model skill. Furthermore, while RFL is typically considered as the most important predictor in radar-based precipitation retrievals (Stephens et al., 2008; Skofronick-Jackson et al., 2015), we find that contributions from RFL, DOV and SPW provide a near-equal level of importance, with respective average percent contributions to model output of 30%, 31% and 28%, while ERA5 TMP and WVL variables have a total combined importance of 10%.

The combined insights from DeepPrecip's multi-model vertical extent evaluations and feature importance analyses demonstrate a potential to influence current and future remote sensing precipitation retrievals using deep learning. Instruments like CloudSat's Cloud Profiling Radar (CPR), or the Global Precipitation Measurement (GPM) mission's Dual-frequency Precipitation Radar (DPR) also use active radar systems to perform similar, radar-based precipitation retrievals based on data from vertical column reflectivities (Stephens et al., 2008). While CPR and GPM-derived products use a more sophisticated Bayesian retrieval to the $Z_e - S/R$ relationships evaluated here, the resulting precipitation estimates are still tightly coupled to a priori physical assumptions of particle shape, size and fallspeed which is a substantial source of uncertainty (Hiley et al., 2010; Wood et al., 2013). Additionally, the results of this study further support prior inference regarding the existence of regions of high importance in the < 1 km (near-surface) region of the vertical column relating to shallow-cumuliform precipitation strongly influencing retrieval accuracy. This is an area that is typically masked in satellite-based products (i.e. the radar "blind-zone") due to surface clutter contamination, and has been shown in previous work to likely be a major source of underestimation from missing shallow cumuliform precipitation (Maahn et al., 2014; Bennartz et al., 2019). This work motivates the importance of continued

research towards obtaining high-quality, non-cluttered near surface radar data to use as additional model inputs in future space-based retrievals of precipitation.

DeepPrecip is not without uncertainty and error, which will reduce its accuracy when tested against new data. Uncertainties present in the training data (stemming from the MRR, ERA5 or Pluvio2 observations), will propagate through the model and bias the output estimates (Kochendorfer et al., 2022; Jakubovitz et al., 2019). We have taken steps to mitigate the impact of these uncertainties through multiple data alignment and preprocessing decisions (details in Section 4.4), however precipitation gauge undercatch, wind shielding configurations, MRR attenuation and differences in site-specific vertical extent cannot be eliminated as contributors to retrieval error. While 60% of the power laws examined in this work were MRR-derived K-band relationships, the remaining 40% were either Ka-band or the Marshall-Palmer (MP) Rayleigh relationship. While K and Ka are similar radar frequencies, the differences between the two can bias the resulting precipitation estimate when a Ka-derived power law is applied to K-band data (especially during periods of intense precipitation). Furthermore, while the collection of data from multiple sites provides us with a robust training set under multiple regional climates, due to the unique experimental setups at each site, calibration biases between study locations may further reduce DeepPrecip's skill when applied to new data. As the MRR instrument has a limited 3 km maximum vertical range, we also miss possible precipitation events occurring outside this region, which may contribute to further surface precipitation underestimation. Internal CNN model uncertainty is likely driven, in part, by a combination of the high variability that is typical of precipitation and the limited sample from nine measurement sites over 8 years, which does not fully capture all different forms of possible precipitation structure and occurrence.

Chapter 5

Conclusions

5.1 Summary

The primary goal of this collective body of work was to advance our understanding of Arctic snow accumulation using a novel combination of remote sensing data and machine learning. Each chapter focused on a separate sub-topic or methodology in this area to evaluate uncertainties in traditional snowfall retrievals and provide insight into future techniques. Multidisciplinary evaluations such as these (combining machine learning and remote sensing in this case) are increasingly important as Arctic temperatures continue to rise, and climate model estimates of Arctic SWE_{max} still display biases of up to 100% over the next 50 years (with similar uncertainties in reanalysis estimates of NH snow mass) (Brown et al., 2017; Mudryk et al., 2015).

The ability to reduce uncertainties in current snow accumulation estimates would be extremely beneficial in constraining current snow products and would facilitate less uncertainty

in future climate model projections (Cassotta et al., 2022). This assessment was performed by first evaluating the capabilities of a spaceborne remote sensing dataset for constraining current gridded SWE products. An assessment of traditional snowfall retrieval methodologies (i.e. power law relations) was then performed, and their accuracy was compared to increasing sophisticated machine learning-based retrievals. Additional emphasis was placed on machine learning interpretability in these comparisons, using multiple feature extraction methods and importance scoring criteria to gain insight into regions within the vertical column of reflectivity data which appear as the most significant contributors to low uncertainty estimates of snowfall. Altogether, these findings comprise a thesis which facilitates an important step towards low uncertainty and highly generalized snowfall retrieval algorithms using machine learning and radar data.

The first research objective, in Section 1.3, asked how remote sensing information could be used as an independent, observation-based constraint for enhancing current gridded SWE products. This topic was addressed in King and Fletcher (2021), where gridded snowfall data from the CloudSat-CPR was used to quantitatively identify statistically inconsistent accumulation estimates (i.e. outliers) from a set of four, independent, gridded SWE datasets. This methodology revealed that the majority (79%) of outliers occurred during periods of high accumulation in September, October, November (SON). Spatiotemporally masking outliers in each gridded product and re-blending them using an unweighted averaging approach to form a new, adjusted dataset, leads to a 14.9% decrease in total NH SWE over SON (equating to $-310,000$ kg m^{-3} of surface water). This overestimation in gridded product records of snow accumulation is consistent with our earlier findings in King and Fletcher (2019), suggesting that models and reanalysis products are overestimating SWE in the CAA. In general, the quality flags generated using this methodology are a quantitative approach to finding regions of potentially low quality data in

gridded SWE products. However, as mentioned at the end of chapter 2, CloudSat-derived snow accumulation estimates are not without their own biases and uncertainties, and future work should continue to assess, and aim to enhance, radar-based retrievals of snowfall.

Chapter 3 focused on addressing the second objective of Section 1.3 by examining uncertainties in traditional reflectivity-snowfall ($Z_e - S$) relationships, and analyzing the potential of machine learning in remote sensing-based retrievals of snowfall (King et al., 2022a). Trained on vertically pointing surface radar data from a single site (GCPEx), the RF model consistently outperformed a set of five other K-band $Z_e - S$ relationships when tested on unseen data at the same site. RF model MSE increased by only 5% when trained on data from GCPEx and tested on data from a completely new location (ICE-POP), while the average $Z_e - S$ MSE values increased by 40%. Additionally, RF feature importance scores and vertical extent sensitivity analysis indicated that multiple, adjacent near surface (< 2 km) observations were the most influential for accurately retrieving snowfall. This was a significant result, as current spaceborne radar retrievals are often blind to the lowest 1.5 km of the atmosphere due to surface clutter contamination (Maahn et al., 2014). While the overall performance of the RF demonstrated improved accuracy over traditional $Z_e - S$ relationships, the available training sample was quite limited ($n = 391$ observations) and the RF had difficulty retrieving snowfall when tested on storm structures not well represented in the training data.

Chapter 4 iterated upon what was learned using the RF retrieval from chapter 3 to address the first half of the final objective in Section 1.3. This objective was focused on the development of a highly generalized NH precipitation retrieval using deep learning (King et al., 2022b). The model (DeepPrecip) was extensively trained on both snowfall and rainfall data from 9 locations over 8 years using a 1D convolutional neural network (resulting in a deep learning

model composed of 3.9 million internal network parameters and 100 thousand training samples). DeepPrecip displayed improved MSE values, which were 187% lower than the traditional power law relationships derived at each location. Further, DeepPrecip outperformed the RF model from chapter 3 (retrained on the same set of data), with MSE values 21% lower and r^2 values 34% higher. DeepPrecip also demonstrated high levels of robustness when tested using a leave-one-site-out cross validation methodology, with lower MSE values (often by orders of magnitude) than any power law relationship examined in this work.

The final portion of chapter 4 focused on the second half of objective 3 related to ML model interpretability. Deep learning algorithms are often considered black box models, wherein an input is provided, the model does some work, and an output is given, with little insight into the internal model decision-making process (Rudin, 2019). Similar to the RF feature importance analysis, a major component of the DeepPrecip project was to analyze model behaviour to gain insight into statistically significant portions of the vertical column of input covariates (King et al., 2022b). Using a Shapely analysis of model inputs, derived feature importance values suggested that reflectivity information is not necessarily the most important predictors in all cases. Inputs like spectral width and Doppler velocity are also beneficial sources of information, which become increasingly important during high intensity precipitation events. This analysis also revealed that a combination of both near-surface (below 1 km) and far ($1.5-2.5$ km) portions of the vertical column are consistently ranked as the most significant contributors to high model accuracy. The combined importance of both near-surface and far bins was also demonstrated in a sensitivity analysis using multiple versions of DeepPrecip trained on different portions of the vertical column, where the model trained on the full column ($0-3$ km) significantly outperformed models trained on solely near-surface (<1 km) or far ($1-3$ km) subsets of the vertical profile.

5.2 Limitations

The assumption that CloudSat-derived snow accumulation estimates can be used as an acceptable ground truth value for constraining gridded products involves substantial uncertainty. CloudSat has been shown to produce similar seasonal cycles and monthly means to in situ at some high Arctic locations, but areas below 70° N suffer from much higher levels of error and uncertainty as overpass granules diverge and the CloudSat sampling frequency substantially decreases (King and Fletcher, 2019; Hiley et al., 2011). As a consequence of the radar blind zone (which masks the lowest 1.5 km of the atmosphere), CloudSat also misses some shallow cumuliform precipitation, which has been found to equate to an approximate 10% underestimation in total annual snowfall (Maahn et al., 2014). Findings from Milani and Wood (2021) suggest that following the extended 2011 battery failure and the activation of Daylight-Only Operations (DO-Op), CloudSat's sampling frequency was further reduced below 70° N and snowfall quantities dropped by 8.21% on average (further compounding the underestimation problem). Resolution also plays a major role in the accuracy of gridded CloudSat snowfall estimates, with seasonal SWE estimate median error percentages only dropping below 50% at 2° by 2° resolution (Duffy et al., 2021). A priori assumptions of hydrometeor microphysics used in the 2C-SNOW-PROFILE retrieval add additional uncertainty to the comparisons with in situ and gridded products (Hiley et al., 2011). Finally, CloudSat snowfall retrieval performance over alpine areas with complex topography has been evaluated over high latitude regions and was found to display strong positive biases; adding further uncertainty to our comparisons of CloudSat-derived accumulation to gridded SWE products over the mountains regions of NA (Palerme et al., 2019).

Errors and uncertainty in the accuracy of the in situ snow accumulation record is another major

source of uncertainty in our analyses. Since the collocated in situ gauge data is the response variable being predicted by the ML models in chapters 3 and 4, any errors in their observations will be learned by the ML retrievals being developed. Wind shielding was a major potential source of this error at many of the sites from chapter 4, some of which (i.e. OLYMPEx 1, 3 and 4) had no shielding installed at the site (Houze et al., 2017). As wind speeds increase, falling hydrometers are pushed over and above the gauge top orifice due to turbulent air fluxes, and are therefore not recorded (Smith, 2008). This undercatch effect can lead to gauge catch efficiency reductions of 50% or more at wind speeds of just 5 m/s (Goodison, 1978). Adding shielding to the measurement site (i.e. Alter shields or Double Fence Intercomparison Reference (DFIR) shields) can substantially reduce this undercatch effect, however underestimation can still occur (Yang, 2014). Transfer functions are one solution for addressing known biases in shielded and unshielded gauge estimates by manually correcting well-defined, known undercatch errors (Buisán et al., 2017). These functions have demonstrated skill at reducing mean absolute errors (MAE) to 2% (from 34%) at multiple sites from the World Meteorological Organization Solid Precipitation Intercomparison Experiment (WMO SPICE) (Kochendorfer et al., 2017). Similar to the assumptions required for deriving $Z_e - S$ relationships, these transfer functions are tightly coupled to the regional climates in which they are derived and do not necessarily generalize well to new locations (Buisán et al., 2017).

Finally, there exist numerous uncertainties and limitations in the ML-based snowfall retrievals performed in chapters 3 and 4. The highest levels of error and uncertainty in both models (i.e. the RF and DeepPrecip) occurred during periods of near-zero °C surface temperatures, and therefore during periods of likely mixed-phase precipitation (Wang et al., 2019b). Recent work from Shin et al. (2022) has demonstrated the effectiveness of ML for precipitation classification,

and highlighted the importance of covariates like wet bulb temperature in accurately predicting precipitation phase. The inclusion of a preprocessing step that calculates phase using another ML classifier and then feeding this into DeepPrecip as an additional predictor, would likely be beneficial in further reducing model uncertainty. Due to the time and training constraints on the DeepPrecip project, this was not attempted, but is discussed more in Section 5.3. The limited vertical extent of the surface radar instruments (10 km for the VertiX system and only 3 km for the MRR) adds further uncertainty in the ML retrieval estimates if precipitation occurs above these regions in the atmosphere. While the majority of medium-to-high intensity snowfall (i.e. > 0.1 mm h^{-1}) generally occurs in the lowest 3 km of the atmosphere, large quantities of precipitation still occur above this threshold (Liu, 2020). As DeepPrecip uses the MRR as its primary source of training data, the model has no knowledge of precipitating cloud layers above the 3 km mark, which may bias its estimates when applied in regions with higher frequencies of deep convective precipitation. Finally, both the RF and DeepPrecip assimilate data from ERA5 which, while being a well-validated temperature and wind product, is still subject to biases and uncertainties in its estimates (Bell et al., 2021). As a provider of input covariates to both models, these errors will be learned by each ML algorithm and propagate through to the final output estimate of precipitation.

5.3 Future work

The continued growth of earth observing system databases, combined with the ease-of-access to ML libraries and cloud computing resources, has facilitated the creation of a new, fast-growing sub-field in the Geosciences (Karpatne et al., 2019). The number of ML-related papers submitted to journals in the Geosciences has increased by orders of magnitude since 2012 for instance,

bringing forth new perspectives and questions for future researchers (Dramsch, 2020). While the work presented here was focused on the objectives in Section 1.3, the findings from each chapter also raise new questions which should be considered in forthcoming studies.

As previously mentioned, uncertainties surrounding mixed-phase precipitation have demonstrated an association with periods of high error in our models (King et al., 2022a). Research from Wang et al. (2019b), Shin et al. (2022) and Behrangi et al. (2018) has demonstrated the significance of wet-bulb temperature (T_w) on accurately predicting precipitation phase. In addition to surface temperature, T_w provides critical information regarding the cooling effect from surface evaporation, which is tightly coupled to hydrometeor phase (Shin et al., 2022). The inclusion of hydrometeor microphysical information (i.e. shape, size, fallspeed, density) as additional input covariates in model training, along with precipitation characteristics from Precipitation Imaging Package (PIP) instrumentation would also be a valuable area of future research, as this information has been shown to be closely related to precipitation rates and has also been used for phase discrimination (Pettersen et al., 2020a; Kneifel et al., 2015; Dudhia et al., 2008). Followup work using large eddy simulations for modelling atmospheric turbulence should also be investigated as another source of information describing boundary layer hydrometeor movement in future deep learning retrievals.

ML-based prediction of southern hemisphere (SH) snowfall should also be investigated. Many of the observational sources of information discussed in this thesis (i.e. CloudSat, MRR records and Pluvio gauges) also have coverage and data records across much of Antarctica, and have been used in validation campaigns in previous studies (Souverijns et al., 2018; Palerme et al., 2017). Applying DeepPrecip to Antarctic station data (at Dumont d'Urville, for instance, which is equipped with an MRR/Pluvio), and comparing these results to reanalysis in the region would

be a beneficial comparison for further evaluating model skill. Incorporating additional Antarctic sites into the set of training data and retraining the model to see how skill metrics change would also help in understanding the generalizability of ML-based retrievals under different climatic regimes.

Comparisons between gridded ML-based snow accumulation estimates and climate model outputs should also be examined in followup work. Examining the spatiotemporal differences in accumulation predictions between these two techniques in the Arctic, could provide valuable new insights into climate model deficiencies over critically important (but poorly constrained) remote regions. Many studies are already using a combination of machine learning and climate modelling for tasks like precipitation downscaling, parameter emulation, and bias correction, however large research gaps still remain at the intersection of these two fields for precipitation prediction (Trinh et al., 2021; Kajbaf et al., 2022; Ahmed et al., 2020; Sachindra et al., 2018; Fletcher et al., 2022; King et al., 2020).

This work also acts as a motivator for future studies related to the spaceborne radar blind zone. It has been suggested that the radar blind zone contributes to a high underestimation in CloudSat's total annual snowfall estimates (up to 10%) (with similar underestimation also suggested from GPM) due to missing shallow cumuliform snowfall (Maahn et al., 2014; Valdivia et al., 2022). The results of both chapter 3 and 4, highlight this portion of the vertical column (between 0 and 2 km) as being an area that is extremely important. Improved observational coverage (or an accurate simulation) of the radar blind zone is, therefore, highly important for improving future radar-based retrievals of snowfall, and reducing biases in current state-of-the-art remote sensing retrievals of precipitation.

Finally, the ability for ML to effectively make use of large datasets suggests that models like DeepPrecip would strongly benefit from the vast amount of spaceborne remote sensing radar data currently available (i.e. from satellites like CloudSat or GPM). Performing a similar analysis to chapter 4 using spaceborne data would result in a training dataset that is orders of magnitude larger and would, in theory, provide additional context into synoptic storm structures. Issues with radar signal attenuation would need to be addressed, since this change would require the development of a retrieval using downwards (instead of upwards) pointing radar. However, the extended height of the vertical profile (up to 30 km for CloudSat) would allow for the ML model to capture light precipitation events occurring further above the 3 km range of the MRR (Stephens, 2017). ML models like DeepPrecip may then operate as an alternative technique (or supplementary component) to current retrieval methods to reduce uncertainties in current and future spaceborne retrievals of snowfall.

Chapter 6

Copyright Permissions

The contents of chapter 2 are taken from King and Fletcher (2021) (i.e. *Using CloudSat-Derived Snow Accumulation Estimates to Constrain Gridded Snow Water Equivalent Products*), which is an open access article distributed under the terms of the Creative Commons CC 4.0 BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The contents of chapter 3 are taken from King et al. (2022a) (i.e. *A Centimeter-Wavelength Snowfall Retrieval Algorithm Using Machine Learning*), with copyright permissions for use in this thesis provided by Erin Gumbel (Senior Permissions Specialist) at the American Meteorological Society.

The contents of chapter 4 are taken from King et al. (2022b) (i.e. *DeepPrecip: A deep neural network for precipitation retrievals*), which is an open access article distributed under the terms of the Creative Commons CC 4.0 BY license, which permits unrestricted use, distribution, and

References

Abadi, M. et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., and Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388:154–170.

Adhikari, A., Ehsani, M. R., and Behrangi, A. (2020a). Snowfall Retrieval from Satellite-based Microwave Humidity Sounders using Machine Learning Methods. *American Geophysical Union, Fall Meeting*, 2020:H075–09. Conference Name: AGU Fall Meeting Abstracts ADS Bibcode: 2020AGUFMH075...09A.

Adhikari, A., Ehsani, M. R., Song, Y., and Behrangi, A. (2020b). Comparative Assessment of Snowfall Retrieval From Microwave Humidity Sounders Using Machine Learning Methods. *Earth and Space Science*, 7(11):e2020EA001357.

Adhikari, A., Liu, C., and Kulie, M. S. (2018). Global Distribution of Snow Precipitation Features and Their Properties from 3 Years of GPM Observations. *Journal of Climate*, 31(10):3731–3754. Publisher: American Meteorological Society Section: Journal of Climate.

Agapiou, A. (2017). Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine applications. *International Journal of Digital Earth*, 10(1):85–102.

Ahmed, K., Sachindra, D. A., Shahid, S., Iqbal, Z., Nawaz, N., and Khan, N. (2020). Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmospheric Research*, 236:104806.

Akbari, A., Ng, L., and Solnik, B. (2021). Drivers of economic and financial integration: A machine learning approach. *Journal of Empirical Finance*, 61:82–102.

Ali, J., Khan, R., Ahmad, N., and Maqsood, I. (2012). Random Forests and Decision Trees.

Amari, S.-i. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4):185–196.

Anderton, S. P., White, S. M., and Alvera, B. (2004). Evaluation of spatial variability in snow water equivalent for a high mountain catchment. *Hydrological Processes*, 18(3):435–453.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., and Prat, O. P. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, 96(1):69–83. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Authorities, C. (1985). *Snow Surveying Manual, Standards and Procedures*. Conservation Authorities and Water Management Branch.

Babichev, S. A., Ries, J., and Lvovsky, A. I. (2002). Quantum scissors: teleportation of single-mode optical states by means of a nonlocal single photon.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, 19(1):389–407.

Banadkooki, F. B., Ehteram, M., Ahmed, A. N., Fai, C. M., Afan, H. A., Ridwam, W. M., Sefelnasr, A., and El-Shafie, A. (2019). Precipitation Forecasting Using Multilayer Neural Network and Support Vector Machine Optimization Based on Flow Regime Algorithm Taking into Account Uncertainties of Soft Computing Models. *Sustainability*, 11(23):6681. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066):303–309. Number: 7066 Publisher: Nature Publishing Group.

Barrie, L. A. (1991). Snow Formation and Processes in the Atmosphere that Influence its Chemical Composition. In Davies, T. D., Tranter, M., and Jones, H. G., editors, *Seasonal Snowpacks*, NATO ASI Series, pages 1–20, Berlin, Heidelberg. Springer.

Battaglia, A., Ajewole, M. O., and Simmer, C. (2007). Evaluation of radar multiple scattering effects in Cloudsat configuration. *Atmospheric Chemistry and Physics*, 7(7):1719–1730. Publisher: Copernicus GmbH.

Becker, E. and Schmitz, G. (2003). Climatological effects of orography and land–sea heating contrasts on the gravity wave–driven circulation of the mesosphere. *J. Atmos. Sci.*, 60:103–118.

Behrangi, A., Yin, X., Rajagopal, S., Stampoulis, D., and Ye, H. (2018). On distinguishing snowfall from rainfall using near-surface atmospheric information: Comparative analysis, uncertainties and hydrologic importance. *Quarterly Journal of the Royal Meteorological Society*, 144(S1):89–102.

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., Soci, C., Villaume, S., Bidlot, J.-R., Haimberger, L., Woollen, J., Buontempo, C., and Thépaut, J.-N. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741):4186–4227.

Beneke, M., Buchalla, G., and Dunietz, I. (1997). Mixing induced CP asymmetries in inclusive B decays. *Phys. Lett.*, B393:132–142.

Bennartz, R., Fell, F., Pettersen, C., Shupe, M. D., and Schuettemeyer, D. (2019). Spatial and temporal variability of snowfall over Greenland from CloudSat observations. *Atmospheric Chemistry and Physics Discussions*, pages 1–32.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., and Kirchner, J. W. (2019). The Relative Importance of Different Flood-Generating Mechanisms Across Europe. *Water Resources Research*, 55(6):4582–4593.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., and Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, 43(9):4382–4390.

Besic, N., Figueras i Ventura, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A. (2016). Hydrometeor classification through statistical clustering of polarimetric radar measurements: a semi-supervised approach. *Atmospheric Measurement Techniques*, 9(9):4425–4445. Publisher: Copernicus GmbH.

Betts, A. K., Chan, D. Z., and Desjardins, R. L. (2019). Near-Surface Biases in ERA5 Over the Canadian Prairies. *Frontiers in Environmental Science*, 7:129.

Beucher, A., Rasmussen, C., Moeslund, T., and Greve, M. (2022). Interpretation of Convolutional Neural Networks for Acid Sulfate Soil Classification. *Frontiers in Environmental Science*, 9.

Bokhorst, S., Pedersen, S. H., Brucker, L., Anisimov, O., Bjerke, J. W., Brown, R. D., Ehrich, D., Essery, R. L. H., Heilig, A., Ingvander, S., Johansson, C., Johansson, M., Jónsdóttir, I. S., Inga, N., Luojus, K., Macelloni, G., Mariash, H., McLennan, D., Rosqvist, G. N., Sato, A., Savela, H., Schneebeli, M., Sokolov, A., Sokratov, S. A., Terzago, S., Vikhamar-Schuler, D., Williamson, S., Qiu, Y., and Callaghan, T. V. (2016). Changing Arctic snow cover: A review of recent developments and assessment of future needs for observations, modelling, and impacts. *Ambio*, 45(5):516–537.

Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, 320(5882):1444–1449. Publisher: American Association for the Advancement of Science.

Boucher, R. J. and Wieler, J. G. (1985). Radar Determination of Snowfall Rate and Accumulation. *Journal of Applied Meteorology and Climatology*, 24(1):68–73. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Boudala, F. S., Gultepe, I., and Milbrandt, J. A. (2021). The Performance of Commonly Used Surface-Based Instruments for Measuring Visibility, Cloud Ceiling, and Humidity at Cold Lake, Alberta. *Remote Sensing*, 13(24):5058. Number: 24 Publisher: Multidisciplinary Digital Publishing Institute.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Breiman, L. (2017). *Classification And Regression Trees*. Routledge, New York.

Bromwich, D. H., Wilson, A. B., Bai, L.-S., Moore, G. W. K., and Bauer, P. (2016). A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic. *Quarterly Journal of the Royal Meteorological Society*, 142(695):644–658.

Brown, R., Derksen, C., and Wang, L. (2010). A multi-data set analysis of variability and change in Arctic spring snow cover extent, 1967–2008. *Journal of Geophysical Research: Atmospheres*, 115(D16).

Brown, R. et al. (2017). *Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017*. Arctic Monitoring and Assessment Programme, Oslo, Norway. OCLC: 1013169302.

Brown, R. D. (2000). Northern Hemisphere Snow Cover Variability and Change, 1915–97. *Journal of Climate*, 13(13):2339–2355. Publisher: American Meteorological Society.

Brown, R. D., Brasnett, B., and Robinson, D. (2003). Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere-Ocean*, 41(1):1–14.

Brown, R. D. and Mote, P. W. (2009). The Response of Northern Hemisphere Snow Cover to a Changing Climate. *Journal of Climate*, 22(8):2124–2145. Publisher: American Meteorological Society Section: Journal of Climate.

Broxton, P. D., Zeng, X., and Dawson, N. (2016). Why Do Global Reanalyses and Land Data Assimilation Products Underestimate Snow Water Equivalent? *Journal of Hydrometeorology*, 17(11):2743–2761.

Broy, M. (1992). Software engineering—from auxiliary to key technologies. In Broy, M. and Denert, E., editors, *Software Pioneers*, pages 10–13. Springer, New York.

Brun, E., Martin, E., Simon, V., Gendre, C., and Coleou, C. (1989). An Energy and Mass Model of Snow Cover Suitable for Operational Avalanche Forecasting. *Journal of Glaciology*, 35(121):333–342.

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F., and Morin, S. (2013). Simulation of Northern Eurasian Local Snow Depth, Mass, and Density Using a Detailed Snowpack Model and Meteorological Reanalyses. *Journal of Hydrometeorology*, 14(1):203–219. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

Buisán, S. T., Earle, M. E., Collado, J. L., Kochendorfer, J., Alastrué, J., Wolff, M., Smith, C. D., and López-Moreno, J. I. (2017). Assessment of snowfall accumulation underestimation by tipping bucket gauges in the Spanish operational network. *Atmospheric Measurement Techniques*, 10(3):1079–1091. Publisher: Copernicus GmbH.

Burke, A., Snook, N., Ii, D. J. G., McCorkle, S., and McGovern, A. (2020). Calibration of Machine Learning–Based Probabilistic Hail Predictions for Operational Forecasting. *Weather and Forecasting*, 35(1):149–168. Publisher: American Meteorological Society Section: Weather and Forecasting.

Buttle, J. M., Allen, D. M., Caissie, D., Davison, B., Hayashi, M., Peters, D. L., Pomeroy, J. W., Simonovic, S., St-Hilaire, A., and Whitfield, P. H. (2016). Flood processes in Canada: Regional and special aspects. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 41(1-2):7–30.

Cabaj, A., Kushner, P. J., Fletcher, C. G., Howell, S., and Petty, A. A. (2020). Constraining Reanalysis Snowfall Over the Arctic Ocean Using CloudSat Observations. *Geophysical Research Letters*, 47(4):e2019GL086426.

Campbell, S. L. and Gear, C. W. (1995). The index of general nonlinear DAES. *Numer. Math.*, 72(2):173–196.

Casella, D., Panegrossi, G., Sanò, P., Marra, A. C., Dietrich, S., Johnson, B. T., and Kulie, M. S. (2017). Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat-CPR. *Atmospheric Research*, 197:64–75.

Casellas Masana, E. (2022). Precipitation phase discrimination: diagnosing and nowcasting. *University of Barcelona*.

Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M., Ottersen, G., Pritchard, H., and Schuur, E. (2022). *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 1 edition.

Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M., and McGovern, A. (2022). A Machine Learning Tutorial for Operational Meteorology, Part I: Traditional Machine Learning. *Weather and Forecasting*, 37(8):1509–1529. arXiv:2204.07492 [physics].

Chase, R. J., Nesbitt, S. W., and McFarquhar, G. M. (2020). Evaluation of the Microphysical Assumptions within GPM-DPR Using Ground-Based Observations of Rain and Snow. *Atmosphere*, 11(6):619. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.

Chase, R. J., Nesbitt, S. W., and McFarquhar, G. M. (2021). A Dual-Frequency Radar Retrieval of Two Parameters of the Snowfall Particle Size Distribution Using a Neural Network. *Journal of Applied Meteorology and Climatology*.

Chen, H., Chandrasekar, V., Cifelli, R., and Xie, P. (2020a). A Machine Learning System for Precipitation Estimation Using Satellite and Ground Radar Network Observations. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2):982–994. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Chen, L., Cao, Y., Ma, L., and Zhang, J. (2020b). A Deep Learning-Based Methodology for Precipitation Nowcasting With Radar. *Earth and Space Science*, 7(2):e2019EA000812.

Chen, X.-W. and Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE Access*, 2:514–525. Conference Name: IEEE Access.

Chollet, F. et al. (2015). Keras.

Choubin, B., Khalighi-Sigaroodi, S., and Malekian, A. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, 61(6):1001–1009.

Chung, S. T. and Morris, R. L. (1978). Isolation and characterization of plasmid deoxyribonucleic acid from streptomyces fradiae. Paper presented at the 3rd international symposium on the genetics of industrial microorganisms, University of Wisconsin, Madison, 4–9 June 1978.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., and Jones, J. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, 7(9):627–637. Number: 9 Publisher: Nature Publishing Group.

Colli, M., Lanza, L. G., La Barbera, P., and Chan, P. W. (2014). Measurement accuracy of weighing and tipping-bucket rainfall intensity gauges under dynamic laboratory testing. *Atmospheric Research*, 144:186–194.

Colli, M., Lanza, L. G., Rasmussen, R., and Thériault, J. M. (2016). The Collection Efficiency of Shielded and Unshielded Precipitation Gauges. Part II: Modeling Particle Trajectories. *Journal of Hydrometeorology*, 17(1):245–255. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

Colli, M., Stagnaro, M., Lanza, L. G., Rasmussen, R., and Thériault, J. M. (2020). Adjustments for Wind-Induced Undercatch in Snowfall Measurements Based on Precipitation Intensity. *Journal of Hydrometeorology*, 21(5):1039–1050. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001916.

Das, S., Chakraborty, R., and Maitra, A. (2017). A random forest algorithm for nowcasting of intense precipitation events. *Advances in Space Research*, 60(6):1271–1282.

Davenport, F. V. and Diffenbaugh, N. S. (2021). Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation. *Geophysical Research Letters*, 48(15):e2021GL093787.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Berg, L. v. d., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., Rosnay, P. d., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Derksen, C. and Brown, R. (2012). Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophysical Research Letters*, 39(19).

Dietz, A. J., Kuenzer, C., Gessner, U., and Dech, S. (2012). Remote sensing of snow – a review of available methods. *International Journal of Remote Sensing*, 33(13):4094–4134.

Dingman, S. L. (2015). *Physical Hydrology: Third Edition*. Waveland Press. Google-Books-ID: rUUaBgAAQBAJ.

Doesken, N. J. and Robinson, D. A. (2009). The Challenge of Snow Measurements. In Dupigny-Giroux, L.-A. and Mock, C. J., editors, *Historical Climate Variability and Impacts in North America*, pages 251–273. Springer Netherlands, Dordrecht.

Dozier, J., Bair, E. H., and Davis, R. E. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *WIREs Water*, 3(3):461–474.

Dramsch, J. S. (2019). Machine Learning in 4D Seismic Data Analysis. *Technical University of Denmark*, page 200.

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. *Advances in Geophysics*, 61:1–55.

Dudhia, J., Hong, S.-Y., and Lim, K.-S. (2008). A New Method for Representing Mixed-phase Particle Fall Speeds in Bulk Microphysics Parameterizations. *J-Stage*, 86A:33–44.

Duffy, G. and Bennartz, R. (2018). The Role of Melting Snow in the Ocean Surface Heat Budget. *Geophysical Research Letters*, 45(18):9782–9789.

Duffy, G., King, F., Bennartz, R., and Fletcher, C. G. (2021). Seasonal Estimates and Uncertainties of Snow Accumulation from CloudSat Precipitation Retrievals. *Atmosphere*, 12(3):363. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Dyer, J. L. and Mote, T. L. (2006). Spatial variability and trends in observed snow depth over North America. *Geophysical Research Letters*, 33(16).

Déry, S. J. and Brown, R. D. (2007). Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback. *Geophysical Research Letters*, 34(22).

Egli, L., Jonas, T., and Meister, R. (2009). Comparison of different automatic methods for estimating snow water equivalent. *Cold Regions Science and Technology*, 57(2):107–115.

Ehsani, M. R. and Behrangi, A. (2022). A comparison of correction factors for the systematic gauge-measurement errors to improve the global land precipitation estimate. *Journal of Hydrology*, 610:127884.

Ehsani, M. R., Behrangi, A., Adhikari, A., Song, Y., Huffman, G. J., Adler, R. F., Bolvin, D. T., and Nelkin, E. J. (2021). Assessment of the Advanced Very High Resolution Radiometer (AVHRR) for Snowfall Retrieval in High Latitudes Using CloudSat and Machine Learning. *Journal of Hydrometeorology*, 22(6):1591–1608. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

El Naqa, I. and Murphy, M. J. (2015). What Is Machine Learning? In El Naqa, I., Li, R., and Murphy, M. J., editors, *Machine Learning in Radiation Oncology: Theory and Applications*, pages 3–11. Springer International Publishing, Cham.

Emery, W. J., Thomas, A. C., Collins, M. J., Crawford, W. R., and Mackas, D. L. (1986). An objective method for computing advective surface velocities from sequential infrared satellite images. *J. Geophys. Res.*, 91:12865–12878.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y.-J., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalczyk, E., Nasonova, N. O., Pyles, R. D., Schlosser, A., Shmakin, A. B., Smirnova, T. G., Strasser, U., Verseghy, D., Yamazaki, T., and Yang, Z.-L. (2017). Validation of the energy budget of an alpine snowpack simulated by several snow models (Snow MIP project). *Annals of Glaciology*, 38:150–158. Publisher: Cambridge University Press.

Euskirchen, E. S., Goodstein, E. S., and Huntington, H. P. (2013). An estimated cost of lost climate regulation services caused by thawing of the Arctic cryosphere. *Ecological Applications*, 23(8):1869–1880.

Fassnacht, S. R. (2004). Estimating Alter-shielded gauge snowfall undercatch, snowpack sublimation, and blowing snow transport at six sites in the coterminous USA. *Hydrological Processes*, 18(18):3481–3492.

Feurer, M. and Hutter, F. (2019). Hyperparameter optimization. In *Automated machine learning*, pages 3–33. Springer, Cham.

Flanner, M., Shell, K., Barlage, M., Perovich, D., and Tschudi, M. (2011). Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. *Nature geoscience*, 4:151–155.

Fletcher, C. G., McNally, W., Virgin, J. G., and King, F. (2022). Toward Efficient Calibration of Higher-Resolution Earth System Models. *Journal of Advances in Modeling Earth Systems*, 14(7):e2021MS002836.

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J. (2020). Latest climate models confirm need for urgent mitigation. *Nature Climate Change*, 10(1):7–10. Number: 1 Publisher: Nature Publishing Group.

Foster, J. L., Hall, D. K., Eylander, J. B., Riggs, G. A., Nghiem, S. V., Tedesco, M., Kim, E., Montesano, P. M., Kelly, R. E. J., Casey, K. A., and Choudhury, B. (2011). A blended global snow product using visible, passive microwave and scatterometer satellite data. *International Journal of Remote Sensing*, 32(5):1371–1395. WOS:000288693800010.

Franzke, C. L. E., O'Kane, T. J., Berner, J., Williams, P. D., and Lucarini, V. (2015). Stochastic climate theory and modeling. *WIREs Climate Change*, 6(1):63–78.

Frappart, F., Ramillien, G., Biancamaria, S., Mognard, N. M., and Cazenave, A. (2006). Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002–2004). *Geophysical Research Letters*, 33(2).

Gatlin, P. (2020). GPM Ground Validation ICE-POP Field Campaign Data Collection. Type: dataset.

Geddes, K. O., Czapor, S. R., and Labahn, G. (1992). *Algorithms for Computer Algebra*. Kluwer, Boston.

Gehring, J., Ferrone, A., Billaulta Roux, A.-C., Besic, N., Ahn, K. D., Lee, G., and Berne, A. (2020). Radar and ground-level measurements of precipitation collected by EPFL during the ICE-POP 2018 campaign in South-Korea. preprint, Atmosphere Meteorology.

Geiss, A. and Hardin, J. C. (2021). Inpainting radar missing data regions with deep learning. *Atmospheric Measurement Techniques*, 14(12):7729–7747. Publisher: Copernicus GmbH.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454.

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., and Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, 141(2):287–299.

Gilmore, M. S., Straka, J. M., and Rasmussen, E. N. (2004). Precipitation Uncertainty Due to Variations in Precipitation Particle Parameters within a Simple Microphysics Scheme. *Monthly Weather Review*, 132(11):2610–2627. Publisher: American Meteorological Society Section: Monthly Weather Review.

Goodison, B. E. (1978). Accuracy of Canadian Snow Gage Measurements. *Journal of Applied Meteorology and Climatology*, 17(10):1542–1548. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Gray, D. and Landine, P. (2011). An Energy-Budget Snowmelt Model for the Canadian Prairies. *Canadian Journal of Earth Sciences*, 25:1292–1303.

Grazioli, J., Genthon, C., Boudevillain, B., Duran-Alarcon, C., Del Guasta, M., Madeleine, J.-B., and Berne, A. (2017). Measurements of precipitation in Dumont d'Urville, Adalie Land, East Antarctica. *The Cryosphere*, 11(4):1797–1811. Publisher: Copernicus GmbH.

Groisman, P. and Davies, T. (2001). Snow cover and the climate system. *Snow Ecolog*, pages 1–44.

Groisman, P. Y., Karl, T. R., Knight, R. W., and Stenchikov, G. L. (1994). Changes of Snow Cover, Temperature, and Radiative Heat Balance over the Northern Hemisphere. *Journal of Climate*, 7(11):1633–1656. Publisher: American Meteorological Society.

Guan, B., Molotch, N. P., Waliser, D. E., Jepsen, S. M., Painter, T. H., and Dozier, J. (2013). Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations. *Water Resources Research*, 49(8):5029–5046.

Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., and Liang, D. (2017). Big Earth Data: a new challenge and opportunity for Digital Earth's development. *International Journal of Digital Earth*, 10(1):1–12.

Hamburger, C. (1995). Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. *Ann. Mat. Pura. Appl.*, 169(2):321–354.

Hanson, C. L., Pierson, F. B., and Johnson, G. L. (2004). Dual-Gauge System for Measuring Precipitation: Historical Development and Use. *Journal of Hydrologic Engineering*, 9(5):350–359. Publisher: American Society of Civil Engineers.

Hao, Z., AghaKouchak, A., Nakhjiri, N., and Farahmand, A. (2014). Global integrated drought monitoring and prediction system (gidmaps) data sets. https://doi.org/10.6084/m9.figshare.853801.

Haynes, J. M., L'Ecuyer, T. S., Stephens, G. L., Miller, S. D., Mitrescu, C., Wood, N. B., and Tanelli, S. (2009). Rainfall retrieval over the ocean with spaceborne W-band radar. *Journal of Geophysical Research: Atmospheres*, 114(D8).

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Hiley, M. J., Kulie, M. S., and Bennartz, R. (2010). Uncertainty Analysis for CloudSat Snowfall Retrievals. *Journal of Applied Meteorology and Climatology*, 50(2):399–418.

Hiley, M. J., Kulie, M. S., and Bennartz, R. (2011). Uncertainty Analysis for CloudSat Snowfall Retrievals. *Journal of Applied Meteorology and Climatology*, 50(2):399–418. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Holte, R. C. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. *Machine Learning*, 11(1):63–90.

Hong, G. (2007). Radar backscattering properties of nonspherical ice crystals at 94 GHz. *Journal of Geophysical Research: Atmospheres*, 112(D22).

Hosaka, M., Nohara, D., and Kitoh, A. (2005). Changes in Snow Cover and Snow Water Equivalent Due to Global Warming Simulated by a 20km-mesh Global Atmospheric Model. *Sola*, 1:93–96.

Houze, R. A., McMurdie, L. A., Petersen, W. A., Schwaller, M. R., Baccus, W., Lundquist, J. D., Mass, C. F., Nijssen, B., Rutledge, S. A., Hudak, D. R., Tanelli, S., Mace, G. G., Poellot, M. R., Lettenmaier, D. P., Zagrodnik, J. P., Rowe, A. K., DeHart, J. C., Madaus, L. E., Barnes, H. C., and Chandrasekar, V. (2017). The Olympic Mountains Experiment (OLYMPEX). *Bulletin of the American Meteorological Society*, 98(10):2167–2188. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501.

Hudak, D., Rodriguez, P., and Donaldson, N. (2008). Validation of the CloudSat precipitation occurrence algorithm using the Canadian C band radar network. *Journal of Geophysical Research: Atmospheres*, 113(D8).

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S. (2013). The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., Le, M., Chandrasekar, V., Brodzik, S., and Kubota, T. (2010). GPM/DPR Level-2 Algorithm Theoretical Basis Document. *NASA*, page 127.

Im, E., Durden, S., Wu, C., and Livermore, T. (2001). The 94-GHz cloud profiling radar for the CloudSat mission. In *2001 IEEE Aerospace Conference Proceedings (Cat. No.01TH8542)*, volume 4, pages 4/1803–4/1809 vol.4.

Jakubovitz, D., Giryes, R., and Rodrigues, M. R. D. (2019). Generalization Error in Deep Learning. In Boche, H., Caire, G., Calderbank, R., Kutyniok, G., Mathar, R., and Petersen, P., editors, *Compressed Sensing and Its Applications: Third International MATHEON Conference 2017*, pages 153–193. Springer International Publishing, Cham.

Jameson, A. R. and Kostinski, A. B. (2002). Spurious power-law relations among rainfall and radar parameters. *Quarterly Journal of the Royal Meteorological Society*, 128(584):2045–2058.

Jash, D., Resmi, E. A., Unnikrishnan, C. K., Sumesh, R. K., Sreekanth, T. S., Sukumar, N., and Ramachandran, K. K. (2019). Variation in rain drop size distribution and rain integral parameters during southwest monsoon over a tropical station: An inter-comparison of disdrometer and Micro Rain Radar. *Atmospheric Research*, 217:24–36.

Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P. (2018). Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, 9(1):1148. Number: 1 Publisher: Nature Publishing Group.

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gurel, N. M., Li, B., Zhang, C., Song, D., and Spanos, C. (2020). Towards Efficient Data Valuation Based on the Shapley Value. *arXiv:1902.10275 [cs, stat]*. arXiv: 1902.10275.

Jiménez-Muñoz, J. C. and Sobrino, J. A. (2003). A generalized single-channel method for retrieving land surface temperature from remote sensing data. *Journal of Geophysical Research: Atmospheres*, 108(D22):2003JD003480.

Jogin, M., Mohana, Madhulika, M. S., Divya, G. D., Meghana, R. K., and Apoorva, S. (2018). Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning. In *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 2319–2323.

Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., and Carslaw, K. S. (2015). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, 7(1):162–187.

Kajbaf, A. A., Bensi, M., and Brubaker, K. L. (2022). Temporal downscaling of precipitation from climate model projections using machine learning. *Stochastic Environmental Research and Risk Assessment*, 36(8):2173–2194.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440. Number: 6 Publisher: Nature Publishing Group.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

Kasischke, E. S. and Turetsky, M. R. (2006). Recent changes in the fire regime across the North American boreal region—Spatial and temporal patterns of burning across Canada and Alaska. *Geophysical Research Letters*, 33(9).

Kelly, R., Chang, A., Tsang, L., and Foster, J. (2003). A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):230–242. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Kerr, Y. H., Waldteufel, P., Wigneron, J., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M., Font, J., Reul, N., Gruhier, C., Juglea, S. E., Drinkwater, M. R., Hahne, A., Martín-Neira, M., and Mecklenburg, S. (2010). The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. *Proceedings of the IEEE*, 98(5):666–687. Conference Name: Proceedings of the IEEE.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., and Kirschbaum, D. B. (2017). So, How Much of the Earth Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, 98(1):69–78. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Kim, H.-U. and Bae, T.-S. (2017). Preliminary Study of Deep Learning-based Precipitation. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 35(5):423–430. Publisher: Korean Society of Surveying, Geodesy, Photogrammetry and Cartography.

Kim, K., Bang, W., Chang, E.-C., Tapiador, F. J., Tsai, C.-L., Jung, E., and Lee, G. (2021). Impact of wind pattern and complex topography on snow microphysics during International Collaborative Experiment for PyeongChang 2018 Olympic and Paralympic winter games (ICE-POP 2018). *Atmospheric Chemistry and Physics*, 21(15):11955–11978. Publisher: Copernicus GmbH.

Kim, M.-J., Kulie, M. S., O'Dell, C., and Bennartz, R. (2007). Scattering of Ice Particles at Microwave Frequencies: A Physically Based Parameterization. *Journal of Applied Meteorology and Climatology*, 46(5):615–633. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Kinar, N. J. and Pomeroy, J. W. (2015). Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, 53(2):481–544.

King, F., Duffy, G., and Fletcher, C. G. (2022a). A Centimeter-Wavelength Snowfall Retrieval Algorithm Using Machine Learning. *Journal of Applied Meteorology and Climatology*, -

1(aop). Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

King, F., Duffy, G., Milani, L., Fletcher, C. G., Pettersen, C., and Ebell, K. (2022b). DeepPrecip: A deep neural network for precipitation retrievals. *EGUsphere*, pages 1–24. Publisher: Copernicus GmbH.

King, F., Erler, A. R., Frey, S. K., and Fletcher, C. G. (2020). Application of machine learning techniques for regional bias correction of snow water equivalent estimates in Ontario, Canada. *Hydrology and Earth System Sciences*, 24(10):4887–4902. Publisher: Copernicus GmbH.

King, F. and Fletcher, C. (2019). Validating CloudSat-CPR retrievals for the estimation of snow accumulation in the Canadian Arctic. *University of Waterloo*.

King, F. and Fletcher, C. G. (2020). Using CloudSat-CPR Retrievals to Estimate Snow Accumulation in the Canadian Arctic. *Earth and Space Science*, 7(2):e2019EA000776.

King, F. and Fletcher, C. G. (2021). Using CloudSat-Derived Snow Accumulation Estimates to Constrain Gridded Snow Water Equivalent Products. *Earth and Space Science*, 8(10):e2021EA001835.

King, F., Kelly, R., and Fletcher, C. G. (2022c). Evaluation of LiDAR-Derived Snow Depth Estimates From the iPhone 12 Pro. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5. Conference Name: IEEE Geoscience and Remote Sensing Letters.

Kneifel, S., von Lerber, A., Tiira, J., Moisseev, D., Kollias, P., and Leinonen, J. (2015). Observed relations between snowfall microphysics and triple-frequency radar measurements. *Journal of Geophysical Research: Atmospheres*, 120(12):6034–6055.

Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meeh, G. A., Meinshausen, M., Murphy, J. M., Plattne, G.-K., Raper, S. C. B., Stocker, T. F., Stott, P. A., Teng, H., , and Wigley, T. M. L. (2008). A review of uncertainties in global temperature projections over the twenty-first century. *J. Climate*, 21:2651–2663.

Kochendorfer, J., Earle, M., Rasmussen, R., Smith, C., Yang, D., Morin, S., Mekis, E., Buisan, S., Roulet, Y.-A., Landolt, S., Wolff, M., Hoover, J., Thériault, J. M., Lee, G., Baker, B., Nitu, R., Lanza, L., Colli, M., and Meyers, T. (2022). How Well Are We Measuring Snow Post-SPICE? *Bulletin of the American Meteorological Society*, 103(2):E370–E388. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Kochendorfer, J., Nitu, R., Wolff, M., Mekis, E., Rasmussen, R., Baker, B., Earle, M. E., Reverdin, A., Wong, K., Smith, C. D., Yang, D., Roulet, Y.-A., Buisan, S., Laine, T., Lee, G., Aceituno, J. L. C., Alastrué, J., Isaksen, K., Meyers, T., Brækkan, R., Landolt, S., Jachcik, A., and Poikonen, A. (2017). Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. *Hydrology and Earth System Sciences*, 21(7):3525–3542. Publisher: Copernicus GmbH.

Kodamana, R. and Fletcher, C. G. (2021). Validation of CloudSat-CPR Derived Precipitation Occurrence and Phase Estimates across Canada. *Atmosphere*, 12(3):295. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Kollias, P. and University, M. (2013). GPM Ground Validation McGill Vertical Pointing X-Band (VertiX) Radar GCPEx. Type: dataset.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres*, 105(D20):24809–24822.

Krumbein, W. C. and Dacey, M. F. (1969). Markov chains and embedded Markov chains in geology. *Journal of the International Association for Mathematical Geology*, 1(1):79–96.

Kuji, M. and Nakajima, T. (2002). Retrieval of cloud geometrical parameters using remote sensing data. In *11th Conf. on Cloud Physics*, page JP1.7, Ogden, UT. Amer. Meteor. Soc.

Kulie, M. S. and Bennartz, R. (2009). Utilizing Spaceborne Radars to Retrieve Dry Snowfall. *Journal of Applied Meteorology and Climatology*, 48(12):2564–2580.

Kulie, M. S., Bennartz, R., Greenwald, T. J., Chen, Y., and Weng, F. (2010). Uncertainties in Microwave Properties of Frozen Precipitation: Implications for Remote Sensing and Data Assimilation. *Journal of the Atmospheric Sciences*, 67(11):3471–3487.

Kulie, M. S. and Milani, L. (2018). Seasonal variability of shallow cumuliform snowfall: A CloudSat perspective. *Quarterly Journal of the Royal Meteorological Society*, 144(S1):329–343.

Kulie, M. S., Milani, L., Wood, N. B., and L'Ecuyer, T. S. (2020a). Global Snowfall Detection and Measurement. In Levizzani, V., Kidd, C., Kirschbaum, D. B., Kummerow, C. D., Nakamura, K., and Turk, F. J., editors, *Satellite Precipitation Measurement*, volume 69, pages 699–716. Springer International Publishing, Cham. Series Title: Advances in Global Change Research.

Kulie, M. S., Milani, L., Wood, N. B., and L'Ecuyer, T. S. (2020b). Global Snowfall Detection and Measurement. In Levizzani, V., Kidd, C., Kirschbaum, D. B., Kummerow, C. D., Nakamura, K., and Turk, F. J., editors, *Satellite Precipitation Measurement: Volume 2*, Advances in Global Change Research, pages 699–716. Springer International Publishing, Cham.

Kulie, M. S., Pettersen, C., Merrelli, A. J., Wagner, T. J., Wood, N. B., Dutter, M., Beachler, D., Kluber, T., Turner, R., Mateling, M., Lenters, J., Blanken, P., Maahn, M., Spence, C., Kneifel, S., Kucera, P. A., Tokay, A., Bliven, L. F., Wolff, D. B., and Petersen, W. A. (2021). Snowfall in the Northern Great Lakes: Lessons Learned from a Multisensor Observatory. *Bulletin of the American Meteorological Society*, 102(7):E1317–E1339. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Lahnert, U., Schween, J. H., Acquistapace, C., Ebell, K., Maahn, M., Barrera-Verdejo, M., Hirsikko, A., Bohn, B., Knaps, A., OConnor, E., Simmer, C., Wahner, A., and Crewell, S. (2015). JOYCE: Jaelich Observatory for Cloud Evolution. *Bulletin of the American Meteorological Society*, 96(7):1157–1174. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Lange, S., Rockel, B., Volkholz, J., and Bookhagen, B. (2015). Regional climate model sensitivities to parametrizations of convection and non-precipitating subgrid-scale clouds over South America. *Climate Dynamics*, 44(9):2839–2857.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444. Number: 7553 Publisher: Nature Publishing Group.

L'Ecuyer, T. S., Wood, N. B., Haladay, T., Stephens, G. L., and Stackhouse, P. W. (2008). Impact of clouds on atmospheric heating based on the R04 CloudSat fluxes and heating rates data set. *Journal of Geophysical Research: Atmospheres*, 113(D8).

Lemke, P., Ren, J., Alley, R. B., Allison, I., Carrasco, J., Flato, G., Fujii, Y., Kaser, G., Mote, P., Thomas, R. H., Zhang, T., Box, J., Bromwich, D., Brown, R., Cogley, J. G., Comiso, J., Dyurgerov, M., Fitzharris, B., Hagen, J. O., Harris, C., Hinzman, L., Hock, R., Hoelzle, M., Huybrechts, P., Isaksen, K., Jansson, P., Jenkins, A., Joughin, I., Kottmeier, C., Kwok, R., Laxon, S., Liu, S., MacAyeal, D., Melling, H., Ohmura, A., Payne, A., Prowse, T., Raup, B. H., Raymond, C., Rignot, E., Rigor, I., Robinson, D., Rothrock, D., Scherrer, S. C., Smith, S., Solomina, O., Yamada, T., Zhao, L., Barry, R., Koike, T., Ren, J., Alley, R. B., Allison, I., Carrasco, J., Flato, G., Fujii, Y., Kaser, G., Mote, P., Thomas, R. H., and Zhang, T. (2007). Observations: Changes in snow, ice and frozen ground. in: Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the

intergovernmental panel on climate change. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*, page 48.

Lemonnier, F., Madeleine, J.-B., Claud, C., Genthon, C., Durán-Alarcón, C., Palerme, C., Berne, A., Souverijns, N., van Lipzig, N., Gorodetskaya, I. V., L'Ecuyer, T., and Wood, N. (2019). Evaluation of CloudSat snowfall rate profiles by a comparison with in situ micro-rain radar observations in East Antarctica. *The Cryosphere*, 13(3):943–954. Publisher: Copernicus GmbH.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, 51(9):7309–7342.

Levizzani, V., Laviola, S., and Cattani, E. (2011). Detection and Measurement of Snowfall from Space. *Remote Sensing*, 3(1):145–166. Number: 1 Publisher: Molecular Diversity Preservation International.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. (2017). Hyperband: a novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816.

Li, L., Qiao, J., Yu, G., Wang, L., Li, H., Liao, C., and Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. *Water Research*, 211:118078.

Li, L., Shangguan, W., Deng, Y., Mao, J., Pan, J., Wei, N., Yuan, H., Zhang, S., Zhang, Y., and Dai, Y. (2020). A Causal Inference Model Based on Random Forests to Identify the Effect of Soil Moisture on Precipitation. *Journal of Hydrometeorology*, 21(5):1115–1131. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., and Dai, Y. (2021). A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology*, 601:126636.

Lindsay, R., Wensnahan, M., Schweiger, A., and Zhang, J. (2014). Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic. *Journal of Climate*, 27(7):2588–2606.

Liston, G. E. (2004). Representing Subgrid Snow Cover Heterogeneities in Regional and Global Models. *Journal of Climate*, 17(6):1381–1397.

Liston, G. E. and Hiemstra, C. A. (2011). The Changing Cryosphere: Pan-Arctic Snow Trends (1979–2009). *Journal of Climate*, 24(21):5691–5712. Publisher: American Meteorological Society Section: Journal of Climate.

Liu, G. (2008a). A Database of Microwave Single-Scattering Properties for Nonspherical Ice Particles. *Bulletin of the American Meteorological Society*, 89(10):1563–1570. Publisher: American Meteorological Society.

Liu, G. (2008b). Deriving snow cloud characteristics from CloudSat observations. *Journal of Geophysical Research: Atmospheres*, 113(D8).

Liu, G. (2020). Radar Snowfall Measurement. In Levizzani, V., Kidd, C., Kirschbaum, D. B., Kummerow, C. D., Nakamura, K., and Turk, F. J., editors, *Satellite Precipitation Measurement: Volume 1*, Advances in Global Change Research, pages 277–295. Springer International Publishing, Cham.

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. Technical Report arXiv:1804.07723, arXiv. arXiv:1804.07723 [cs] type: article.

Lops, Y., Pouyaei, A., Choi, Y., Jung, J., Salman, A. K., and Sayeed, A. (2021). Application of a Partial Convolutional Neural Network for Estimating Geostationary Aerosol Optical Depth Data. *Geophysical Research Letters*, 48(15):e2021GL093096.

Louw, T. and McIntosh-Smith, S. (2021). Using the Graphcore IPU for Traditional HPC Applications. *AccML*. Number: 4896 Publisher: EasyChair.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67. Number: 1 Publisher: Nature Publishing Group.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 4768–4777, Red Hook, NY, USA. Curran Associates Inc.

Maahn, M., Burgard, C., Crewell, S., Gorodetskaya, I. V., Kneifel, S., Lhermitte, S., Van Tricht, K., and van Lipzig, N. P. M. (2014). How does the spaceborne radar blind zone affect derived surface snowfall statistics in polar regions? *Journal of Geophysical Research (Atmospheres)*, 119:13,604–13,620. ADS Bibcode: 2014JGRD..11913604M.

Maahn, M. and Kollias, P. (2012). Improved Micro Rain Radar snow measurements using Doppler spectra post-processing. *Atmospheric Measurement Techniques*, 5(11):2661–2673. Publisher: Copernicus GmbH.

Maahn, M., Turner, D. D., Löhnert, U., Posselt, D. J., Ebell, K., Mace, G. G., and Comstock, J. M. (2020). Optimal Estimation Retrievals and Their Uncertainties: What Every Atmospheric Scientist Should Know. *Bulletin of the American Meteorological Society*, 101(9):E1512–E1523. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Mao, Y. and Sorteberg, A. (2020). Improving Radar-Based Precipitation Nowcasts with Machine Learning Using an Approach Based on Random Forest. *Weather and Forecasting*, 35(6):2461–2478. Publisher: American Meteorological Society Section: Weather and Forecasting.

Marchand, R., Mace, G. G., Ackerman, T., and Stephens, G. (2008). Hydrometeor Detection Using Cloudsat—An Earth-Orbiting 94-GHz Cloud Radar. *Journal of Atmospheric and Oceanic Technology*, 25(4):519–533.

Marks, C. J. and Rodgers, C. D. (1993). A retrieval method for atmospheric composition from limb emission measurements. *Journal of Geophysical Research: Atmospheres*, 98(D8):14939–14953.

Marshall, J. S. and Palmer, W. M. K. (1948). THE DISTRIBUTION OF RAINDROPS WITH SIZE. *Journal of the Atmospheric Sciences*, 5(4):165–166. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.

Matrosov, S. Y. (2007). Modeling Backscatter Properties of Snowfall at Millimeter Wavelengths. *Journal of the Atmospheric Sciences*, 64(5):1727–1736. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.

Matrosov, S. Y. and Battaglia, A. (2009). Influence of multiple scattering on CloudSat measurements in snow: A model study. *Geophysical Research Letters*, 36(12).

Matrosov, S. Y., Campbell, C., Kingsmill, D., and Sukovich, E. (2009). Assessing Snowfall Rates from X-Band Radar Reflectivity Measurements. *Journal of Atmospheric and Oceanic Technology*, 26(11):2324–2339. Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology.

Matrosov, S. Y., Clark, K. A., and Kingsmill, D. E. (2007). A Polarimetric Radar Approach to Identify Rain, Melting-Layer, and Snow Regions for Applying Corrections to Vertical Profiles of Reflectivity. *Journal of Applied Meteorology and Climatology*, 46(2):154–166. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Matrosov, S. Y., Shupe, M. D., and Djalalova, I. V. (2008). Snowfall Retrievals Using Millimeter-Wavelength Cloud Radars. *Journal of Applied Meteorology and Climatology*, 47(3):769–777.

Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Maxwell, A. and Shobe, C. (2022). Land-surface parameters for spatial predictive mapping and modeling. *Earth-Science Reviews*, page 103944.

McIlhattan, E. A., Pettersen, C., Wood, N. B., and L'Ecuyer, T. S. (2020). Satellite observations of snowfall regimes over the Greenland Ice Sheet. *The Cryosphere*, 14(12):4379–4404. Publisher: Copernicus GmbH.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J. (2000). The Coupled Model Intercomparison Project (CMIP). *Bulletin of the American Meteorological Society*, 81(2):313–318. Publisher: American Meteorological Society.

Meixner, T., Bastidas, L. A., Gupta, H. V., and Bales, R. C. (2002). Multicriteria parameter estimation for models of stream chemical composition. *Water Resour. Res.*, 38.

Mekis, E., Donaldson, N., Reid, J., Zucconi, A., Hoover, J., Li, Q., Nitu, R., and Melo, S. (2018). An Overview of Surface-Based Precipitation Observations at Environment and Climate Change Canada. *Atmosphere-Ocean*, 56(2):71–95.

Mekis, E. and Vincent, L. A. (2011). An Overview of the Second Generation Adjusted Daily Precipitation Dataset for Trend Analysis in Canada. *Atmosphere-Ocean*, 49(2):163–177. Publisher: Taylor and Francis.

Metcalfe, J. R., Routledge, B., and Devine, K. (1997). Rainfall measurement in canada: Changing observational methods and archive adjustment procedures. *Journal of Climate*, 10(1):92–101.

Milani, L., Kulie, M., Casella, D., Dietrich, S., L'Ecuyer, T., Panegrossi, G., Porcu, F., Sanò, P., and B. Wood, N. (2018). CloudSat snowfall estimates over Antarctica and the Southern Ocean: An assessment of independent retrieval methodologies and multi-year snowfall analysis. *Atmospheric Research*, 213.

Milani, L. and Wood, N. B. (2021). Biases in CloudSat Falling Snow Estimates Resulting from Daylight-Only Operations. *Remote Sensing*, 13(11):2041. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Mudryk, L. R., Derksen, C., Kushner, P. J., and Brown, R. (2015). Characterization of Northern Hemisphere Snow Water Equivalent Datasets, 1981–2010. *Journal of Climate*, 28(20):8037–8051.

Mudryk, L. R., Kushner, P. J., Derksen, C., and Thackeray, C. (2017). Snow cover response to temperature in observational and climate model ensembles. *Geophysical Research Letters*, 44(2):919–926.

Mullan, D., Swindles, G., Patterson, T., Galloway, J., Macumber, A., Falck, H., Crossley, L., Chen, J., and Pisaric, M. (2017). Climate change and the long-term viability of the World's busiest heavy haul ice road. *Theoretical and Applied Climatology*, 129(3):1089–1108.

Munchak, S. J., Schrom, R. S., Helms, C. N., and Tokay, A. (2022). Snow microphysical retrieval from the NASA D3R radar during ICE-POP 2018. *Atmospheric Measurement Techniques*, 15(5):1439–1464. Publisher: Copernicus GmbH.

Musselman, K. N., Addor, N., Vano, J. A., and Molotch, N. P. (2021). Winter melt trends portend widespread declines in snow water resources. *Nature Climate Change*, 11(5):418–424. Number: 5 Publisher: Nature Publishing Group.

Navon, I. M. (2009). Data Assimilation for Numerical Weather Prediction: A Review. In Park, S. K. and Xu, L., editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, pages 21–65. Springer, Berlin, Heidelberg.

Newendorp, P. D. (1976). *Decision analysis for petroleum exploration*. Planning Press. Publisher: Penn Well Books,Tulsa, OK.

Nummelin, A., Li, C., and Hezel, P. J. (2017). Connecting ocean heat transport changes from the midlatitudes to the Arctic Ocean. *Geophysical Research Letters*, 44(4):1899–1908.

Oozeer, Y., Fletcher, C. G., and Champagne, C. (2020). Evaluation of Satellite-Derived Surface Soil Moisture Products over Agricultural Regions of Canada. *Remote Sensing*, 12(9):1455. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Orescanin, M., Petković, V., Powell, S. W., Marsh, B. R., and Heslin, S. C. (2022). Bayesian Deep Learning for Passive Microwave Precipitation Type Detection. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5. Conference Name: IEEE Geoscience and Remote Sensing Letters.

Orsolini, Y. J., Senan, R., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., Carrasco, A., and Benestad, R. E. (2013). Impact of snow initialization on sub-seasonal forecasts. *Climate Dynamics*, 41(7):1969–1982.

Osborn, G. (1982). Handbook of Snow: Principles, Processes, Management & Use, edited by D.M. Gray and D.H. Male. *ARCTIC*, 35(2):346–346. Number: 2.

Palerme, C., Claud, C., Dufour, A., Genthon, C., Wood, N. B., and L'Ecuyer, T. (2017). Evaluation of Antarctic snowfall in global meteorological reanalyses. *Atmospheric Research*, 190:104–112.

Palerme, C., Claud, C., Wood, N. B., L'Ecuyer, T., and Genthon, C. (2019). How Does Ground Clutter Affect CloudSat Snowfall Retrievals Over Ice Sheets? *IEEE Geoscience and Remote Sensing Letters*, 16(3):342–346. Conference Name: IEEE Geoscience and Remote Sensing Letters.

Palerme, C., Kay, J. E., Genthon, C., L'Ecuyer, T., Wood, N. B., and Claud, C. (2014). How much snow falls on the Antarctic ice sheet? *Cryosphere*, 8(4):1577–1587. WOS:000341622700032.

Pan, B., Hsu, K., AghaKouchak, A., and Sorooshian, S. (2019). Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resources Research*, 55(3):2301–2321.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011a). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011b). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peng, S., Piao, S., Ciais, P., Fang, J., and Wang, X. (2010). Change in winter snow depth and its impacts on vegetation in China. *Global Change Biology*, 16(11):3004–3013.

Petersen, W. A., Hudak, D., Bringi, V. N., Siqueira, P., Tokay, A., Chandrasekar, V., Bliven, L. F., Cifelli, R., Rutledge, S., Skofronick-Jackson, G., and Schwaller, M. (2007). NASA GPM/PMM PARTICIPATION IN THE CANADIAN CLOUDSAT/CALIPSO VALIDATION PROJECT (C3VP): PHYSICAL PROCESS STUDIES IN SNOW. *33rd International Conference on Radar Meteorology, American Meteorological Society*, page 7.

Petersen, W. A. and NASA (2013). GPM Ground Validation Pluvio Precipitation Gauge GCPEx. Type: dataset.

Pettersen, C., Bliven, L. F., von Lerber, A., Wood, N. B., Kulie, M. S., Mateling, M. E., Moisseev, D. N., Munchak, S. J., Petersen, W. A., and Wolff, D. B. (2020a). The Precipitation Imaging Package: Assessment of Microphysical and Bulk Characteristics of Snow. *Atmosphere*, 11(8):785. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Pettersen, C., Kulie, M. S., Bliven, L. F., Merrelli, A. J., Petersen, W. A., Wagner, T. J., Wolff, D. B., and Wood, N. B. (2020b). A Composite Analysis of Snowfall Modes from Four Winter Seasons in Marquette, Michigan. *Journal of Applied Meteorology and Climatology*, 59(1):103–124. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Pomeroy, J. and Brun, E. (2001). Physical properties of snow. *Snow Ecology*.

Prasad, A. M., Iverson, L. R., and Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, 9(2):181–199.

Pulliainen, J., Grandell, J., and Hallikainen, M. (1999). HUT snow emission model and its applicability to snow water equivalent retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1378–1390. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., and Norberg, J. (2020). Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018. *Nature*, 581(7808):294–298. Number: 7808 Publisher: Nature Publishing Group.

Putman, W. M. and Lin, S.-J. (2007). Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, 227:55–78.

Pörtner, H., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegria, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N. (2019). Ipcc, 2019: IPCC special report on the ocean and cryosphere in a changing climate. *Cambridge University Press*, page 755.

Quirita, V. A. A., da Costa, G. A. O. P., Happ, P. N., Feitosa, R. Q., Ferreira, R. d. S., Oliveira, D. A. B., and Plaza, A. (2017). A New Cloud Computing Architecture for the Classification of Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):409–416. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., and Gutmann, E. (2012). How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bulletin of the American Meteorological Society*, 93(6):811–829. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Rasmussen, R. M., Hallett, J., Purcell, R., Cole, J., and Tryhane, M. (2002). The hot plate snow gauge. In *Preprints, 11th Conf. on Cloud Physics, Ogden, UT, Amer. Meteor. Soc. P*, volume 1.

Rasmussen, R. M. and Rasmussen, R. M. (2014). Examination of the Performance of Single Alter Shielded and Unshielded Snow Gauges Using Observations from the Marshall Field Site during the SPICE WMO Field Program and Numerical Model Simulations. In *Examination of the Performance of Single Alter Shielded and Unshielded Snow Gauges Using Observations from the Marshall Field Site during the SPICE WMO Field Program and Numerical Model Simulations*. AMS.

Rathore, M. M. U., Paul, A., Ahmad, A., Chen, B.-W., Huang, B., and Ji, W. (2015). Real-Time Big Data Analytical Architecture for Remote Sensing Application. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4610–4621. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S. (2021). Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677. Number: 7878 Publisher: Nature Publishing Group.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204. Number: 7743 Publisher: Nature Publishing Group.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, New York, NY, USA. Association for Computing Machinery.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24(14):3624–3648.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*. World Scientific. Google-Books-ID: p3b3ngEACAAJ.

Rogers, D. C. (1973). *The Aggregation of Natural Ice Crystals*. Department of Atmospheric Resources, University of Wyoming. Google-Books-ID: SoiqtgAACAAJ.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215. Number: 5 Publisher: Nature Publishing Group.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B. J. C. (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric Research*, 212:240–258.

Saha, R., Testik, F. Y., and Testik, M. C. (2021). Assessment of OTT Pluvio2 Rain Intensity Measurements. *Journal of Atmospheric and Oceanic Technology*, 38(4):897–908. Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.

Schoger, S. Y., Moisseev, D., Lerber, A. v., Crewell, S., and Ebell, K. (2021). Snowfall-Rate Retrieval for K- and W-Band Radar Measurements Designed in Hyyti, Finland, and Tested at Ny-alesund, Svalbard, Norway. *Journal of Applied Meteorology and Climatology*, 60(3):273–289. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Schuur, E. a. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E. (2015). Climate change and the permafrost carbon feedback. *Nature*, 520(7546):171–179. Number: 7546 Publisher: Nature Publishing Group.

Sevruk, B., Hertig, J. A., and Spiess, R. (1991). The effect of a precipitation gauge orifice rim on the wind field deformation as investigated in a wind tunnel. *Atmospheric Environment. Part A. General Topics*, 25(7):1173–1179.

Seymour, R. S., editor (1981). *Conductive Polymers*. Plenum, New York.

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, 54(11):8558–8593.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. arXiv:1506.04214 [cs].

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. *Conference on Neural Information Processing Systems*.

Shin, K., Kim, K., Song, J. J., and Lee, G. (2022). Classification of Precipitation Types Based on Machine Learning Using Dual-Polarization Radar Measurements and Thermodynamic Fields. *Remote Sensing*, 14(15):3820. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

Shrestha, D. L. and Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19(2):225–235.

Shupe, M. D., Kollias, P., Matrosov, S. Y., and Schneider, T. L. (2004). Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra. *Journal of Atmospheric and Oceanic Technology*, 21(4):660–670. Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology.

Simmons, A., Soci, C., Nicolas, J., Bell, B., Berrisford, P., Dragani, R., Flemming, J., Haimberger, L., Healy, S., Hersbach, H., Horányi, A., Inness, A., Muñoz-Sabater, J., Radu, R., and Schepers, D. (2020). *Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1*. ECMWF.

Sims, E. M. and Liu, G. (2015). A Parameterization of the Probability of Snow–Rain Transition. *Journal of Hydrometeorology*, 16(4):1466–1477. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.

Skofronick-Jackson, G., Hudak, D., Petersen, W., Nesbitt, S. W., Chandrasekar, V., Durden, S., Gleicher, K. J., Huang, G.-J., Joe, P., Kollias, P., Reed, K. A., Schwaller, M. R., Stewart, R., Tanelli, S., Tokay, A., Wang, J. R., and Wolde, M. (2015). Global Precipitation Measurement Cold Season Precipitation Experiment (GCPEX): For Measurement Sake Let It Snow. *Bulletin of the American Meteorological Society*, 96(10):1719–1741. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Skofronick-Jackson, G., Kulie, M., Milani, L., Munchak, S. J., Wood, N. B., and Levizzani, V. (2019). Satellite Estimation of Falling Snow: A Global Precipitation Measurement (GPM) Core Observatory Perspective. *Journal of Applied Meteorology and Climatology*, 58(7):1429–1448. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini,

R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., and Wilheit, T. (2017). The Global Precipitation Measurement (GPM) Mission for Science and Society. *Bulletin of the American Meteorological Society*, 98(8):1679–1695. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Slifka, M. K. and Whitton, J. L. (2000). Clinical implications of dysregulated cytokine production. *J. Mol. Med.*, 78:74–80.

Smith (2009). The Relationships Between Snowfall Catch Efficiency and Wind Speed for the Geonor T-200B Precipitation Gauge Utilizing Various Wind Shield Configurations. Place: Canmore, AB Publisher: Western Snow Conference Series: Proceedings of the 77th Annual Western Snow Conference.

Smith, D. C. (2008). Correcting the wind bias in snowfall measurements made with a geonor t-200b precipitation gauge and alter wind shield. *Canadian Meteorological and Oceanographic Society*, pages 162–167.

Smith, S. E. (1976). Neuromuscular blocking drugs in man. In Zaimis, E., editor, *Neuromuscular junction. Handbook of experimental pharmacology*, volume 42, pages 593–660, Heidelberg. Springer.

Souverijns, N., Gossart, A., Lhermitte, S., Gorodetskaya, I. V., Grazioli, J., Berne, A., Duran-Alarcon, C., Boudevillain, B., Genthon, C., Scarchilli, C., and van Lipzig, N. P. M. (2018). Evaluation of the CloudSat surface snowfall product over Antarctica using ground-based precipitation radars. *The Cryosphere*, 12(12):3775–3789. Publisher: Copernicus GmbH.

Souverijns, N., Gossart, A., Lhermitte, S., Gorodetskaya, I. V., Kneifel, S., Maahn, M., Bliven, F. L., and van Lipzig, N. P. M. (2017). Estimating radar reflectivity - Snowfall rate relationships and their uncertainties over Antarctica by combining disdrometer and radar observations. *Atmospheric Research*, 196:211–223.

Stephens, G. (2017). 2C-SNOW-PROFILE-PR05. *NASA*, page 21.

Stephens, G., Vane, D., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., and Mitrescu, C. (2002). The cloudsat mission and the a-train. *Bulletin of the American Meteorological Society*, 83(12):1771–1790.

Stephens, G., Vane, D. G., Tanelli, S., Im, E., Durden, S., Rokey, M., Reinke, D., Partain, P., Mace, G. G., Austin, R., L'Ecuyer, T., Haynes, J., Lebsock, M., Suzuki, K., Waliser, D., Wu, D., Kay, J., Gettelman, A., Wang, Z., and Marchand, R. (2008). CloudSat mission: Performance and

early science after the first year of operation. *Journal of Geophysical Research: Atmospheres*, 113(D8).

Stephens, G., Winker, D., Pelon, J., Trepte, C., Vane, D., Yuhas, C., L'Ecuyer, T., and Lebsock, M. (2018). CloudSat and CALIPSO within the A-Train: Ten Years of Actively Observing the Earth System. *Bulletin of the American Meteorological Society*, 99(3):569–581. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

Stone, P. H. and Risbey, J. S. (1990). On the limitations of general circulation climate models. *Geophysical Research Letters*, 17(12):2173–2176.

Sturm, M., Goldstein, M. A., and Parr, C. (2017). Water and life from snow: A trillion dollar science question. *Water Resources Research*, 53(5):3534–3544.

Sturm, M. and Stuefer, S. (2013). Wind-blown flux rates derived from drifts at arctic snow fences. *Journal of Glaciology*, 59(213):21–34. Publisher: Cambridge University Press.

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., and Lea, J. (2010). Estimating Snow Water Equivalent Using Snow Depth Data and Climate Classes. *Journal of Hydrometeorology*, 11(6):1380–1394.

Sun, C., Walker, J. P., and Houser, P. R. (2004). A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research: Atmospheres*, 109(D8).

Sun, W., Hu, Y., Lin, B., Liu, Z., and Videen, G. (2011). The impact of ice cloud particle microphysics on the uncertainty of ice water content retrievals. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(2):189–196.

Szyrmer, W. and Zawadzki, I. (2010). Snow Studies. Part II: Average Relationship between Mass of Snowflakes and Their Terminal Fall Velocity. *Journal of the Atmospheric Sciences*, 67(10):3319–3335. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.

Tait, A. B. (1998). Estimation of Snow Water Equivalent Using Passive Microwave Radiation Data. *Remote Sensing of Environment*, 64(3):286–291.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, 115(12):3517–3529.

Tarek, M., Brissette, F. P., and Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24(5):2527–2544. Publisher: Copernicus GmbH.

Tei, S., Sugimoto, A., Liang, M., Yonenobu, H., Matsuura, Y., Osawa, A., Sato, H., Fujinuma, J., and Maximov, T. (2017). Radial Growth and Physiological Response of Coniferous Trees to Arctic Amplification. *Journal of Geophysical Research: Biogeosciences*, 122(11):2786–2803.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456:12–29. Place: Amsterdam Publisher: Elsevier WOS:000308060100002.

Thackeray, C. W. and Fletcher, C. G. (2016). Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions. *Progress in Physical Geography: Earth and Environment*, 40(3):392–408. Publisher: SAGE Publications Ltd.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1-2):123–138.

Trinh, T., Do, N., Nguyen, V. T., and Carr, K. (2021). Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam. *Climate Dynamics*, 57(9):2713–2735.

Troy, T. J., Sheffield, J., and Wood, E. F. (2012). The role of winter precipitation and temperature on northern Eurasian streamflow trends. *Journal of Geophysical Research: Atmospheres*, 117(D5).

Umargono, E., Suseno, J., and Gunawan, S. (2020). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. In *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*.

Uzair, M. and Jamil, N. (2020). Effects of Hidden Layers on the Efficiency of Neural networks. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6. ISSN: 2049-3630.

Valdivia, J. M., Gatlin, P. N., Kumar, S., Scipión, D., Silva, Y., and Petersen, W. A. (2022). The GPM-DPR Blind Zone Effect on Satellite-Based Radar Estimation of Precipitation over the Andes from a Ground-Based Ka-band Profiler Perspective. *Journal of Applied Meteorology and Climatology*, 61(4):441–456. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Van Baelen, J., Tridon, F., and Pointin, Y. (2009). Simultaneous X-band and K-band study of precipitation to derive specific ZR relationships. *Atmospheric Research*, 94(4):596–605.

Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24:774–780.

Vaughan, D., Comiso, J., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Steffen, K., and Zhang, T. (2013). *Observations: Cryosphere*, book section 4, page 317–382. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Vicente-Serrano, S. M., García-Herrera, R., Peña-Angulo, D., Tomas-Burguera, M., Domínguez-Castro, F., Noguera, I., Calvo, N., Murphy, C., Nieto, R., Gimeno, L., Gutierrez, J. M., Azorin-Molina, C., and El Kenawy, A. (2022). Do CMIP models capture long-term observed annual precipitation trends? *Climate Dynamics*, 58(9):2825–2842.

Vincent, L. A., Zhang, X., Brown, R. D., Feng, Y., Mekis, E., Milewska, E. J., Wan, H., and Wang, X. L. (2015). Observed Trends in Canada's Climate and Influence of Low-Frequency Variability Modes. *Journal of Climate*, 28(11):4545–4560. Publisher: American Meteorological Society Section: Journal of Climate.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Vogl, T., Maahn, M., Kneifel, S., Schimmel, W., Moisseev, D., and Kalesse-Los, H. (2022). Using artificial neural networks to predict riming from Doppler cloud radar observations. *Atmospheric Measurement Techniques*, 15(2):365–381. Publisher: Copernicus GmbH.

von Lerber, A., Moisseev, D., Bliven, L. F., Petersen, W., Harri, A.-M., and Chandrasekar, V. (2017). Microphysical Properties of Snow and Their Link to Ze–S Relations during BAECC 2014. *Journal of Applied Meteorology and Climatology*, 56(6):1561–1582. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Wang, C., Graham, R. M., Wang, K., Gerland, S., and Granskog, M. A. (2019a). Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic

sea ice: effects on sea ice thermodynamics and evolution. *The Cryosphere*, 13(6):1661–1679. Publisher: Copernicus GmbH.

Wang, Y.-H., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., and Niu, G.-Y. (2019b). A Wet-Bulb Temperature-Based Rain-Snow Partitioning Scheme Improves Snowpack Prediction Over the Drier Western United States. *Geophysical Research Letters*, 46(23):13825–13835.

Wellmann, C., Barrett, A. I., Johnson, J. S., Kunz, M., Vogel, B., Carslaw, K. S., and Hoose, C. (2020). Comparing the impact of environmental conditions and microphysics on the forecast uncertainty of deep convective clouds and hail. *Atmospheric Chemistry and Physics*, 20(4):2201–2219. Publisher: Copernicus GmbH.

Wen, Y., Vicol, P., Ba, J., Tran, D., and Grosse, R. (2018). Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches. arXiv:1803.04386 [cs, stat].

Witschas, B., Lemmerz, C., Geiß, A., Lux, O., Marksteiner, U., Rahm, S., Reitebuch, O., and Weiler, F. (2020). First validation of Aeolus wind observations by airborne Doppler wind lidar measurements. *Atmospheric Measurement Techniques*, 13:2381–2396.

Wolfe, J. P. and Snider, J. R. (2012). A Relationship between Reflectivity and Snow Rate for a High-Altitude S-Band Radar. *Journal of Applied Meteorology and Climatology*, 51(6):1111–1128. Publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology.

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., and Berne, A. (2021). RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland. *Atmospheric Measurement Techniques*, 14(4):3169–3193. Publisher: Copernicus GmbH.

Wolff, M. A., Isaksen, K., Petersen verleir, A., demark, K., Reitan, T., and Braskkan, R. (2015). Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19(2):951–967. Publisher: Copernicus GmbH.

Wood, N. B. and L'Ecuyer, T. S. (2021). What millimeter-wavelength radar reflectivity reveals about snowfall: an information-centric analysis. *Atmospheric Measurement Techniques*, 14(2):869–888. Publisher: Copernicus GmbH.

Wood, N. B., L'Ecuyer, T. S., Bliven, F. L., and Stephens, G. L. (2013). Characterization of video disdrometer uncertainties and impacts on estimates of snowfall rate and radar reflectivity. *Atmospheric Measurement Techniques*, 6(12):3635–3648.

Wood, N. B., L'Ecuyer, T. S., Heymsfield, A. J., Stephens, G. L., Hudak, D. R., and Rodriguez, P. (2014). Estimating snow microphysical properties using collocated multisensor observations. *Journal of Geophysical Research: Atmospheres*, 119(14):8941–8961.

Wood, N. B., L'Ecuyer, T. S., Heymsfield, A. J., and Stephens, G. L. (2015). Microphysical Constraints on Millimeter-Wavelength Scattering Properties of Snow Particles. *Journal of Applied Meteorology and Climatology*, 54(4):909–931. Publisher: American Meteorological Society.

Wrzesien, M. L., Durand, M. T., Pavelsky, T. M., Kapnick, S. B., Zhang, Y., Guo, J., and Shum, C. K. (2018). A New Estimate of North American Mountain Snow Accumulation From Regional Climate Model Simulations. *Geophysical Research Letters*, 45(3):1423–1432.

Xiao, R., Chandrasekar, V., and Liu, H. (1998). Development of a neural network based algorithm for radar snowfall estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3):716–724. Conference Name: IEEE Transactions on Geoscience and Remote Sensing.

Xue, M., Hang, R., Liu, Q., Yuan, X.-T., and Lu, X. (2021). CNN-based near-real-time precipitation estimation from Fengyun-2 satellite over Xinjiang, China. *Atmospheric Research*, 250:105337.

Yang, D. (2014). Double Fence Intercomparison Reference (DFIR) vs. Bush Gauge for "true" snowfall measurement. *Journal of Hydrology*, 509:94–100.

Yu, T., Chandrasekar, V., Xiao, H., and Joshil, S. S. (2020). Characteristics of Snow Particle Size Distribution in the PyeongChang Region of South Korea. *Atmosphere*, 11(10):1093. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

Yucel, I., Onen, A., Yilmaz, K. K., and Gochis, D. J. (2015). Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *Journal of Hydrology*, 523:49–66.

Zahmatkesh, Z., Tapsoba, D., Leach, J., and Coulibaly, P. (2019). Evaluation and bias correction of SNODAS snow water equivalent (SWE) for streamflow simulation in eastern Canadian basins. *Hydrological Sciences Journal*, 64(13):1541–1555.

Zhang, L., Zhang, L., and Du, B. (2016). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40. Conference Name: IEEE Geoscience and Remote Sensing Magazine.