Quantifying seasonal snow using very-high-resolution stereo optical satellite images

Jyun-Yi Michelle Hu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

David E. Shean, Chair

Jessica D. Lundquist

Ellyn M. Enderlin

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Remote sensing of seasonal snow cover and depth using very-high-resolution satellite images

Jyun-Yi Michelle Hu

Chair of the Supervisory Committee:
David Shean
Department of Civil and Environmental Engineering

Seasonal snow is one of the most dynamic components of the cryosphere. This water resource serves multiple roles in ecohydrology, climate modulation, and nutrient cycling. Measuring snow distributions is critical to estimating water storage and to predict melt volumes and timing. Interconnected, multi-scalar processes control snow distributions, and at the mountain ridge scale and smaller, wind redistribution, terrain, and vegetation interactions drive heterogeneity. With a changing climate affecting meteorological conditions and shifting the timing of events on a regional basis, statistical relationships (provided enough data has been collected to generate these relationships) break down, making real-time region-specific monitoring important.

In the last two decades, the ability to accurately make fine-scale observations of snow has greatly increased, enabling us to better understand and model snow depth spatial heterogeneity and temporal variability. However, most of these methods involve aerial or terrestrial platforms, are limited in coverage, and require significant technical expertise and investment of resources to acquire instruments and collect data. Very-high-resolution satellite remote sensing is a promising alternative, providing broad coverage, meter-scale resolution, and little cost to end users.

However, rigorously evaluated retrieval algorithms are needed for operational adoption of satellite products.

This dissertation investigates the potential of spaceborne stereo photogrammetry using commercial optical imagery to make fine-scale measurements of seasonal snow across a variety of snow climates including the maritime snowpack of the Washington North Cascades, the continental snowpack of the Colorado Rockies, and the boreal forest / taiga of interior Alaska. Comparison with an expansive collection of reference datasets (e.g., data from SNOTEL network; ground, air, and satellite datasets from SnowEx campaigns) across a range of scales provide extensive validation of stereo snow depth mapping approaches. Assessments of stereo snow depth illustrate how stereo snow depth measurement accuracy is affected by stereo pair collection conditions (viewing geometry, illumination), surface properties (land cover, topography), and processing choices (stereo reconstruction and co-registration). The large spatial coverage of stereo snow depth maps also highlights the strengths and limitations of other measurement approaches, such as swath alignment in airborne lidar products. Residual artifacts with decimeter magnitudes are the largest remaining source of error in stereo snow depth products. Though this work focused on select snow climates and vegetated regions in the United States, it has global applications for quantifying changes in surface elevation with sub-meter-scale magnitudes. With decimeter-scale accuracy (RMSE 0.3 m), satellite stereo collections may offer accurate basin-scale estimates of snow distributions.

ACKNOWLEDGEMENTS

I must first begin by extending my gratitude to my advisor, David Shean, for bringing me to Seattle and UW. This experience has given me the opportunity of getting to know and working with some amazing people. David's meticulous attention to detail has played a pivotal role in helping me to sharpen my eye and ensuring my work stands on its own. I would also like to thank the rest of my committee for their time, energy, perspectives, and feedback throughout the years. Thank you to Jessica Lundquist, for being an amazing mentor, a source of strength and compassion. This dissertation would not have come together without her. Thank you to Ellyn Enderlin for her candidness, which continues to inspire me to give voice to the parts of myself which do not neatly fit into academia's boxes. Thank you to David Perkel for his guidance and generosity of time –his words of encouragement, reality checks, and advice were buoys of light in a sometimes dark and stormy sea.

I am truly appreciative for the financial support that has enabled me to pursue and finish my doctoral studies. The Allison/Knudson Fellowship funded my first year, and in subsequent years, I was supported by funding from NASA (80NSSC18K1405, 80NSSC22K0682) and the Bureau of Reclamation (R21AC10446).

Thank you to the current and former members of the Terrain Analysis and Cryosphere Observations at the University of Washington. Your enthusiasm, and eagerness to provide opportunities for procrastination and distraction have been a delight and getting to know you all has been one of my favorite parts of this experience. Thank you to the Hydrology and Hydrodynamics graduate students and alumni and eScience affiliates who formed my academic family, you've infused this journey with an unparalleled balance of levity and depth. Thank you especially to the crew at Wilcox for bringing light and laughter to a very dingy building. Thank you to my larger academic adjacent community – the members of the CEE Action Planning group, the representatives of the Graduate Student Advisory Board (2021–2023) and the stewards and members of the UAW 4121. Your passion, energy, and commitment to making our community a better place give me hope for the future.

A heartfelt thank you to my friends and family, who have helped ground me when I felt adrift, disconnected, and demotivated. Your reminders about the vast world beyond this graduate experience kept me from sinking too deeply into the weeds. I am especially indebted to my parents for their unwavering support. Thank you for your trust and for providing me with such a solid foundation – I'm blown away by the sacrifices you've made for your children and all the ways in which you show us how deeply we are loved. Finally, I would not be where I am or who I am without the support of my partner, Ash. Thank you for signing up for this adventure neither of us was fully prepared for. Your patience, kindness, humor, and love have been the wind in my sails. I can't wait to start the next chapter with you.

PUBLISHED MATERIAL

At the time of writing, Chapter 2 and Chapter 3 of this research have been published. Chapter 2 was accepted to *Remote Sensing* and Chapter 3 was accepted to *Geophysical Research Letters.* I would like to acknowledge MDPI, publisher of *Remote Sensing*, and AGU, publisher of *Geophysical Research Letters*, for granting permission to reproduce the published and submitted articles.

The citations for the published articles are:

Hu, J. M. and D. E. Shean. (2022). Improving mountain snow and land cover mapping using very- high-resolution optical satellite imagery and Random Forest machine learning models. *Remote Sensing.* https://www.mdpi.com/2072-4292/14/17/4227

Hu, J. M., Shean, D. E., and S. Bhushan. (2023). Six consecutive seasons of high-resolution mountain snow depth maps from satellite stereo imagery. *Geophysical Research Letters.* https://doi.org/10.1029/2023GL104871

Datasets accompanying the preceding articles have also been published with the following citations:

Hu, J. M. and D. Shean. (2023). Land Cover Classification, Snow Cover, and Fractional Snow-Covered Area Maps from Maxar WorldView Satellite Images, Version 1 [Data Set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/USXB6X9CD4Q2.

Hu, J. M. and D. Shean. (2023). SnowEx20 3m Snow Depth DEM Time Series from High-Resolution Satellite Image Pairs, Version 1 [Data Set]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/7QCNCHVQMCI8.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

1.1 MOTIVATION

1.1.1 Seasonal snow and a changing climate

Seasonal snow is a critical component of earth's climatic, hydrologic, and ecological systems. Snow cover regulates energy balance and exchange with its low thermal conductivity and high albedo surface. Snow accumulation serves as climate refuge for over-wintering animals, with its distribution affecting phenological and plant species' distributions and controlling the timing, quantity, and rate of snowmelt and biogeochemical cycling. With altered precipitation regimes and warming temperatures, widespread declines in snow cover and snowmelt water storage/availability and the corresponding impacts on downstream flow are expected to continue, especially in high elevation areas (e.g., Cho et al., 2021; Gudmundsson et al., 2021; Hale et al., 2023; Marshall et al., 2019; Stewart, 2009). However, regional variability in seasonal snow changes makes measurements of snow depth even more critical for estimating seasonal snow volume, snow water equivalent, and intra- and inter-seasonal changes. Accurate measurement of these snow variables is needed to make useful forecasts and effective management decisions for hydropower, irrigation, and water supply.

1.2 BACKGROUND

1.2.1 Dynamics and variability of seasonal snow

Highly dynamic in space and time, the distribution of seasonal snow is modified by preferential wind deposition, redistribution, and differential melt (e.g., Lehning et al., 2008; Mott et al., 2011). These factors are in turn affected by topography, vegetation, and meteorology at varying spatial scales (e.g., Elder et al., 1991) to generate highly localized, site-specific accumulation distributions. Although snow depth generally increases with elevation due to orographic lift, precipitation is more strongly tied to wind (i.e., speed, direction) and slope

(Pomeroy & Brun, 2001). After snow falls, it can be eroded by the wind from low surface roughness areas and preferentially deposited in high roughness or wind-sheltered areas, resulting in lee-side accumulation of objects perpendicular to prevailing winds (Mott et al., 2018). Provided that prevailing winds and precipitation patterns are consistent, terrain-forced convergence can result in high intra- and inter-annual persistence of snow depth distributions (Grünewald et al., 2013; Schirmer et al., 2011) and allow for basin-wide extrapolation (Pflug et al., 2021).

While networks of sensors, like the Snow Telemetry (SNOTEL) network, collect temporally dense data, they are spatially sparse and typically established at lower elevations. The preferential spatial distribution of sensors effectively undersamples higher elevations and rugged environments, where large volumes of seasonal snow accumulate each year and where snow depth distributions can vary widely from initial snowfall distributions (Pomeroy & Brun, 2001). Many water managers rely on empirical relationships between snow water equivalent at index sites and expected streamflow to make decisions and generate water forecasts, but these statistical relationships may break down in abnormal years.

Establishing statistical relationships are limited to regions that possess sufficient data, which may not be attainable for many data-scarce, snow-dominant watersheds whose communities lack access to the resources (e.g., financial, technological, personnel, or otherwise) needed to collect the requisite data. Indeed, in many parts of the world, seasonal snow is deposited in remote, treacherous locations posing additional hazards for data collection. Enter remote sensing approaches for widespread coverage and monitoring.

1.2.2 Key remote sensing approaches for large-scale snow mapping

Several techniques have been developed to map snow cover and estimate snow water equivalent across large extents. Optical approaches leverage the distinct spectral characteristics of snow (high reflectance in the visible range and the low reflectance in the infrared range) to help distinguish from other types of land cover (Valovcin, 1976). In the satellite age, operational snow cover maps that employ this approach with a computed normalized difference snow index (NDSI) are limited by the spatial resolution of SWIR sensors and revisit times (3-5 days at 20-30

m or daily at 500 m) in addition to cloud obstruction. Despite efforts to address these issues (Gascoin et al., 2020), these scales of observation are unlikely to either capture the true distribution of snow or to accurately represent driving interacting processes that control distribution (Blöschl, 1999) at and below the hillslope scale (<100 m).

Multifrequency passive microwave remote sensing approaches use condition-dependent retrieval algorithms to estimate snow depth by disentangling detected radiation (a combination of emission from the ground, the snow, and scattering within the snowpack). While these approaches can operate day or night and are weather-robust, due to the low level of energy emitted at these wavelengths (0.1 to 100 cm), large areas (~1 km resolution) are needed for accurate estimates. Passive microwave products are coarse and limited in spatial coverage and accuracy is sensitive to dense forest cover, heterogeneous snow grain size, deep snow, snow state (i.e., wet or dry), soil moisture, and complex topography (Ly, 2020; Tanniru & Ramsankaran, 2023). Active microwave sensing approaches, including interferometric synthetic aperture radar, can attain much finer spatial resolution (~3 m), but are similarly challenged by dense forest cover, changes in atmospheric water vapor, deep snow, and wet snow.

In airborne light detection and ranging (lidar), pulses of light are emitted towards a surface, and by recording precise position of emission and the distance and time of flight of pulses, can retrieve elevation measurements and produce point clouds and digital elevation models (DEM). With a snow-free collection, corresponding point clouds and DEMs alike can be aligned in space (i.e., co-registered) and differenced to produce highly accurate (vertical RMSE ~0.1 m; Hodgson & Bresnahan, 2004), very-high-resolution (< 3 m) maps of elevation change between the dates of collection. With further processing, snow depth signals can be extracted from other elevation change signals. While airborne lidar currently represents the most precise and accurate method for large coverage snow mapping, instruments and personnel are costly and flight planning and weather windows are needed for data collection.

1.2.3 Mapping snow with stereo images

Remote sensing has entered a new era in the past two decades, with an ever-increasing number of commercial satellites delivering submeter-scale panchromatic (PAN) and meter-scale

multispectral (MS) imagery. This new generation of very-high-resolution (VHR) satellite sensors can provide data at spatial resolutions previously limited to airborne campaigns. With accurate image geolocation of ~3-4 m, VHR products can make meter-scale measurements of snow distributions in regions of complex land cover and terrain, enabling the study of snow processes across a range of observation scales, without sacrificing spatial coverage.

Stereo approaches reconstruct surface elevation by capturing two or more images of an area of interest from two or more different perspectives. These images are usually taken in a short span of time to minimize differences in the surface and atmosphere. Common pixels in these images are matched with correlation algorithms and the 3-dimensional position of each matched pixel can be triangulated by leveraging information about precise satellite positions at the time of image capture and pixel location in the images. Point clouds and DEMs are generated through this process of pixel correlation and triangulation and like airborne lidar, two collections (one snow-on, one snow-free) may be used to produce a fine resolution (1–3 m) maps of elevation change and snow depth.

Though spaceborne stereo snow depth mapping may pose a "viable alternative to airborne campaigns" (Deschamps-Berger et al., 2020) in extending fine-scale snow depth mapping in complex terrain, substantial refinement, testing, and validation is needed to assess available sensors and stereo products across environmental snow conditions.

1.3 RESEARCH OBJECTIVES AND OVERVIEW

The primary objective of this dissertation is to investigate the capacity (and limitations) of VHR satellite stereo techniques to reconstruct surface elevation and retrieve seasonal snow properties across a variety of snow climates. The following chapters are presented as standalone manuscripts. Machine learning models and big data approaches are used to develop stereo snow mapping techniques. Records from long-term monitoring networks, and ground, air, and satellite measurements from NASA's SnowEx campaign are used to evaluate stereo snow depth products.

Chapter 2 presents an assessment of WorldView imagery products using Random Forest machine learning models to classify snow and a broad subset of land cover classes during the accumulation season over the Washington North Cascades, and the Grand Mesa in Colorado.

Direct measurements of VHR snow covered area allowed for evaluation of coarser operational fractional snow cover products. Results show that coarser products overestimate completely snow-covered and completely snow-free pixels in open areas, demonstrating the impact of observation scale on retrieved snow cover distributions.

Chapter 3 details the use of multi-view stereo collections over the Grand Mesa in Colorado. These collections are used to produce an optimized workflow that processes stereo images, leveraging contemporaneous land cover maps from Chapter 2, and integrates lidar snow-free references to generate maps of snow depth. Over a six-season range of collections near the timing of peak SWE, including anomalously low and high snow years, I investigate intra-annual and interannual snow depth variability.

Chapter 4 extends the workflow developed in Chapter 3 to investigate stereo snow depth mapping with a wide assortment of stereo pairs over a shallow, boreal forest snowpack in Fairbanks, Alaska. The assortment of stereo pairs enables an assessment of viewing geometries and sensor pairings to identify optimal cross-track configurations that may approximate in-track accuracy and precision. Assessments of alternative DSM products and sensors are provided to characterize error and residual artifacts which limit vertical accuracy.

Chapter 2. Improving Mountain Snow and Land Cover Mapping Using Very-High-Resolution (VHR) Optical Satellite Images and Random Forest Machine Learning Models

2.1 Abstract

Very-high-resolution (VHR) optical imaging satellites can offer precise, accurate, and direct measurements of snow-covered areas (SCA) with sub-meter to meter-scale resolution in regions of complex land cover and terrain. We explore the potential of Maxar WorldView-2 and WorldView-3 in-track stereo images (WV) for land and snow cover mapping at two sites in the Western U.S. with different snow regimes, topographies, vegetation, and underlying geology. We trained random forest models using combinations of multispectral bands and normalized difference indices (i.e., NDVI) to produce land cover maps for priority feature classes (snow, shaded snow, vegetation, water, and exposed ground). We then created snow-covered area products from these maps and compared them with coarser resolution satellite fractional snow-covered area (fSCA) products from Landsat (~30 m) and MODIS (~500 m). Our models generated accurate classifications, even with limited combinations of available multispectral bands. The models were trained on a single-image-demonstrated limited model transfer, with the best results found for in-region transfers. The coarser-resolution Landsat and MODSCAG fSCA products identified many more pixels as completely snow-covered (100% fSCA) than WV fSCA. However, while MODSCAG fSCA products also identified many more completely snow-free pixels (0% fSCA) than WV fSCA, Landsat fSCA products only slightly underestimated the number of completely snow-free pixels. Overall, our results demonstrate that strategic image observations with VHR satellites such as WorldView-2 and WorldView-3 can complement the existing operational snow data products to map the evolution of seasonal snow cover.

2.2 Introduction

As one of the most dynamic components of the cryosphere, seasonal snow is an integral part of hydrologic, ecologic, economic, cultural, and climatic systems. Snow accumulates during the

cool season and releases meltwater during warmer and drier periods (Bales et al., 2006; Viviroli et al., 2007). The timing and quantity of the accumulation and melt can have consequential impacts on downstream processes, phenology, and water supply (Jones, 1999; Steltzer et al., 2009; Walker et al., 1993; Winkler et al., 2018), especially within mountain watersheds. Seasonal snowpack acts as a natural reservoir, providing a critical freshwater resource to over one billion people worldwide (Mankin et al., 2015).

Despite the significant role of seasonal snow in many different Earth systems, understanding and measuring its distribution and water content can be challenging. Snow cover varies across a range of spatiotemporal scales based on meteorology, topography, and vegetation. Localized sources of redistribution, such as wind transport, combine with topography and vegetation processes to generate additional spatial variability after snow has fallen (Elder et al., 1991; Winstral et al., 2002). This highly heterogeneous distribution of snow suggests that in situ point measurements (e.g., snow depth from automatic weather stations) may not be representative of broader snowpack characteristics (Molotch & Bales, 2005), especially for remote, topographically complex, and inaccessible areas. Though useful for hydrological models, the existing snow fraction estimates are insufficient for process-based ecohydrological (Broxton et al., 2015; Mazzotti et al., 2020) and ecological research that requires more fine-scale and spatially distributed information (Berman et al., 2019; Cosgrove et al., 2021; Essery & Pomeroy, 2004; Thibault & Ouellet, 2005).

2.2.1 Satellite Snow Mapping and Mixed Pixels

Satellite remote sensing can offer measurements over large spatial extents with variable temporal coverage. Passive microwave ‘all-weather day-or-night’ approaches are capable of expansive daily coverage but with a limited spatial resolution of ~25 km and a variable performance for wet vs. dry snow. These issues limit the ability of passive microwave sensors to observe fine-scale processes affecting the highly heterogeneous snow conditions in complex terrain. Active microwave approaches (e.g., synthetic aperture radar) can offer much finer spatial resolutions (0.25–20 m), but its interpretations are affected by wet snow, vegetation, and steep topography (increased layover and shadow effects).

The widely available optical remote-sensing earth observation instruments (e.g., Moderate Resolution Imaging Spectroradiometer [MODIS] on Aqua/Terra, Enhanced Thematic Mapper Plus [ETM+] on Landsat 7, and Operational Land Imager [OLI] on Landsat 8) have been essential for mapping regional to global snow cover (Hall et al., 2002). Snow-covered area (SCA) can be extracted from optical data using relatively simple approaches (Nolin, 2010) given the difference in albedo between snow and other common land cover types. However, the presence of optically thick clouds and variable illumination in complex terrain still present challenges for optical remote-sensing land cover classification (Giles, 2001). For a more detailed discussion of the methods and issues associated with optical remote sensing for global snow cover mapping, we refer the reader to (Dozier & Painter, 2004; Giles, 2001; Hall et al., 1995; Rittger et al., 2020; Vikhamar & Solberg, 2003).

Optical snow-cover-mapping approaches (Valovcin, 1976) routinely leverage the relatively low reflectance of snow in the shortwave infrared (SWIR) and relatively high reflectance in the visible spectrum to calculate a normalized difference snow index (NDSI), which can more effectively distinguish snow from clouds. The operational workflows to derive snow-cover products from remote-sensing satellites employ reflectance thresholding and a rule-based approach using this index for pixel-based snow classification [17,22,25]. These products are dependent on SWIR band(s) and are typically limited to a binary snow vs. “not snow” classification at the lower resolution of SWIR sensors. Thus, the spatial resolution of most publicly available snow products is relatively coarse (~15–500 m), resulting in mixed radiance values from multiple land cover features in each pixel. In one study, mixed pixels were found to comprise 25–93% of all pixels at a 40 m resolution and 67–100% at 500 m (D. J. Selkowitz et al., 2014). The WorldView-3 satellite can also collect SWIR image data, but the resolution is coarser than the multispectral bands (~3.7 m vs. ~1.2 m, respectively) and must be requested during tasking, with additional product costs.

Subpixel snow-mapping techniques were developed to address mixed pixels by separating snow endmembers from non-snow endmembers using field and laboratory spectral libraries (Dozier & Painter, 2004). These spectral unmixing models use linear spectral mixture algorithms to deconstruct constituent signals of different land-cover features. They then estimate fractional

snow cover for each pixel and aggregate the fractional snow-covered area (fSCA) (Painter et al., 2003, 2009; Rosenthal & Dozier, 1996; Vikhamar & Solberg, 2003).

While unmixing provides estimates of the snow fraction within a pixel (with post-processing yielding values between 0–100% or 0–1), these approaches rely on well-calibrated data, both in terms of spectral libraries and radiometrically calibrated satellite instruments such as MODIS. Even with this calibration, the coarser products tend to overestimate the snow at boundaries with other feature classes (Hao et al., 2019) and are not able determine the true spatial distribution of snow at the subpixel level. To address these limitations, approaches involving downscaling, histogram matching, and data fusion techniques can potentially be used to extract finer spatial resolution fSCA products (Cristea et al., 2017; Walters et al., 2014).

2.2.2 Very-High-Resolution Snow Mapping and Machine Learning

Despite improvements in subpixel fractional snow mapping, these techniques cannot identify small-scale snow features such as drifts or snow patches, which can be important for late-season water resources (Parr et al., 2020). Very-high-resolution (VHR) image products (< 1–2 m resolution) can capture precise and accurate measurements of snow and land cover in regions of complex terrain and can be used to resolve snow cover in forest gaps and even between individual trees. This enables a more detailed study of snow deposition (Lundquist & Lott, 2008; Mott et al., 2017; Thorn & Hall, 2002) and snowmelt processes (Currier et al., 2022). The percentage of mixed pixels at a fine resolution is also considerably smaller than at coarser Landsat or MODIS resolutions (Marston et al., 2017).

However, this boost in spatial resolving power comes with tradeoffs—large data volumes and larger- than-memory datasets present major challenges for standard processing approaches. While GUI-based unsupervised and semi-supervised classification tools (e.g., ESRI, ArcGIS) may be sufficient for some studies, those approaches can be cumbersome and inefficient for large data volumes.

Modern machine learning approaches, combined with enhanced computational resources, can be used to extract information from large data volumes with limited supervision (i.e., manual

intervention). The random forest (RF) algorithm (Breiman, 2001), a pixel-based method, is often used for land cover mapping due to its computational efficiency, interpretability, ability to extract feature importance metrics, and relatively low requirements for its training data volume (Liu et al., 2020; Maxwell et al., 2018) compared to other machine learning approaches such as neural networks. Though machine learning models have demonstrated substantial skill in classification accuracy, the spatiotemporal transferability (i.e., generalization) of such models is variable (Jin et al., 2018; Revuelto et al., 2020; Tsai et al., 2019; Wu et al., 2021).

Beyond operational snow-monitoring applications, accurate, fine-resolution maps of snow and land cover are needed to identify static, exposed control surfaces for the precise co-registration of "snow-on" and "snow-off" DEMs to derive snow depth products from stereo VHR images (Deschamps-Berger et al., 2020; Marti et al., 2016). However, stereo images inherently require larger off-nadir viewing angles, often > 20–30°, to provide sufficient parallax for accurate stereo triangulation of the snow surface. While not ideal compared to sensors with nadir orientation, we attempt to use these off-nadir stereo images to produce accurate landcover maps, with the goal of combining them with contemporaneous stereo DEMs for accurate snow depth mapping.

In this work, we use WorldView-2 and WorldView-3 stereo panchromatic and multispectral images and a semi-automated workflow to train land cover models to produce fine-scale snow and land cover maps over mountainous areas at two study sites in the Western U.S. Though other studies have demonstrated the uneven transferability of random forest models, we also investigated whether our simple land cover models could transfer to other images, specifically to classify snow. To assess how the resulting VHR snow cover maps may complement publicly available, operational snow cover datasets such as those derived from MODIS and Landsat, the VHR maps were downsampled and compared with corresponding coarser-resolution operational fSCA products. These VHR snow maps can also be used for the calibration and validation of coarser products (Bair et al., 2020), despite differences in instrument architectures (e.g., whiskbroom vs. pushbroom), illumination geometries, viewing geometries, and methodological processing approaches.

Using varying combinations of spectral bands and band ratio indices, we developed simple random forest models to classify a broad subset of priority land cover classes (i.e., illuminated snow, shaded snow, vegetation, exposed surfaces, surface water, and clouds) common to mountainous areas and needed for stereo snow depth mapping. With these models and land cover classifications, we investigated the following questions:

1. Can random forest models and minimally processed VHR stereo multispectral images be used to accurately classify snow cover at the meter scale without SWIR bands or more complex atmospheric, topographic, and BRDF corrections?
    a. What combination(s) of input layers provide the best model performance?
    b. Can a single model trained for one region be used to accurately classify snow when applied to out-of-region images?
2. How do coarser resolution operational fSCA products from the spectral unmixing of Landsat (30 m) and MODIS (500 m) images compare with the VHR snow cover products?

This paper is organized as follows: Section 2.3 describes the study sites and data. Section 2.4 details the data preprocessing steps, the machine learning model training, the mode of assessment, the model generalization tests, and the procedure for the snow cover product comparison. Section 2.5 summarizes the key results of the classification, generalization, and snow cover comparison tests. In Section 2.6, we discuss machine learning model performance, provide insight into classification challenges, examine the differences with coarser resolution snow cover products, and provide a commentary on the operational potential.

2.3 Study Sites and Data

2.3.1 Study Sites

We selected two snow monitoring sites in the Western US: the Washington North Cascade Range and Grand Mesa in Colorado (Figure 2.1). These two sites have distinct climates, topography, and land cover that produce different snow conditions. The North Cascades site is more challenging for optical snow observation compared to the Grand Mesa site, with increased cloud cover, terrain shadowing, and the presence of glacial ice and firn.

The North Cascade Range of the Washington state (Figure 2.1) site spans an elevation range of 430–2703 m a.s.l. from the river valley bottom to the mountain peaks. The exposed areas in this region are primarily schist, orthogneiss, and plutonic rocks (Haugerud & Tabor, 2009). Western hemlock, red cedar, and Douglas fir are the predominant tree species at lower elevations (below 800 m) while silver fir is found between 600–1300 m (Kruckeberg, 1991). At higher elevations, mountain hemlock is mixed with subalpine meadows between 1200–1600 m (Kruckeberg, 1991), with alpine lakes between 1270–1749 m a.s.l.

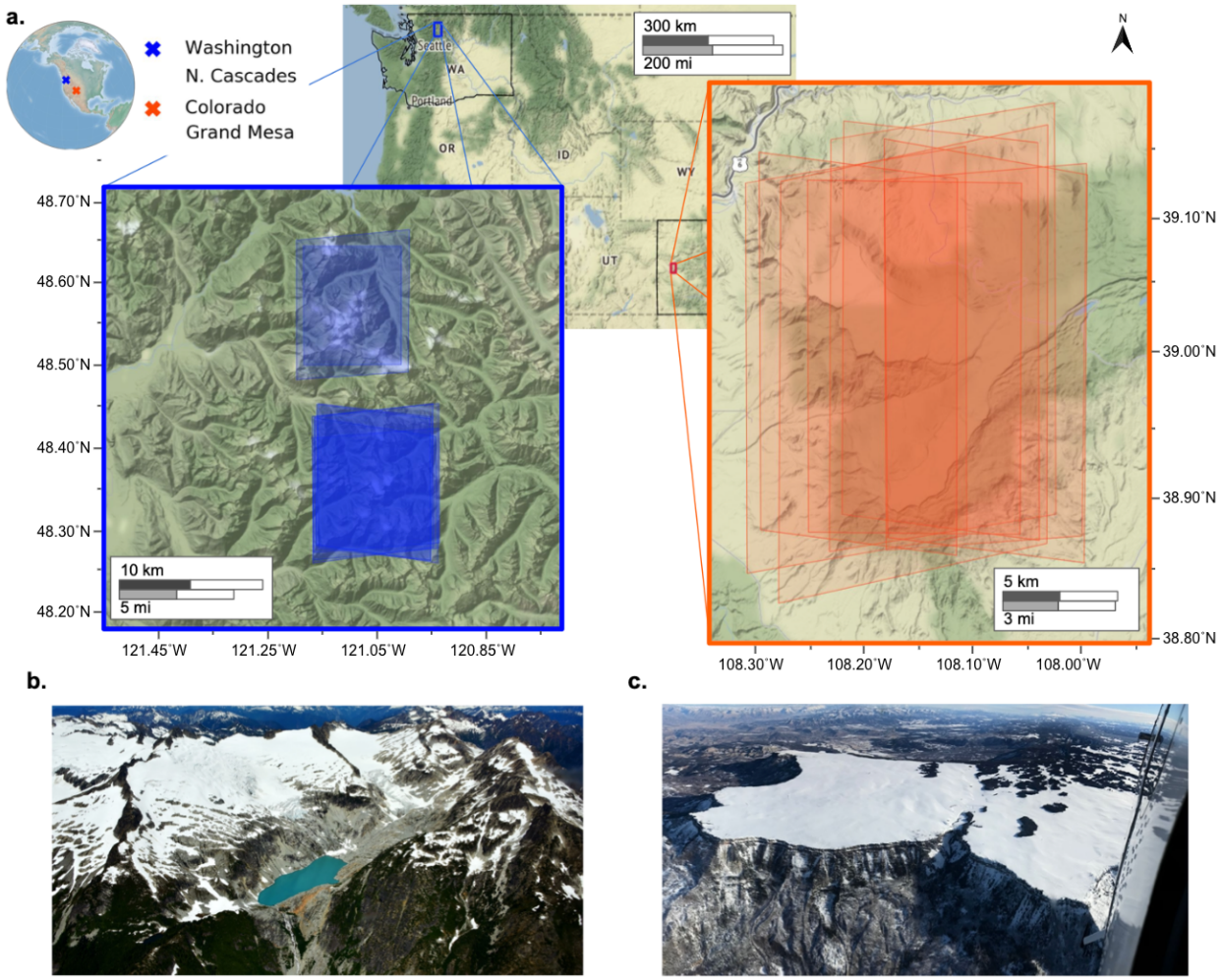

Figure 2.1. (a) Study site locations with VHR image footprints over shaded relief basemap: North Cascades, WA, USA (left in blue, southern footprints show images over South Cascade Glacier) and Grand Mesa, CO, USA (right in orange). Oblique aerial context images of (b) mountain peaks and an alpine lake in the North Cascades site (July 2017, photo courtesy of Long Bach Nguyen) and (c) the north arm of the Grand Mesa site (February 2017, photo courtesy of Chris Chickadel).

This site is geologically complex with perennial snowfields and dozens of active glaciers, including the USGS benchmark South Cascade Glacier (O'Neel et al., 2019). Winters in the Washington North Cascades are generally mild, with moisture-laden air from the Pacific Ocean (Bach, 2002; Rasmussen & Tangborn, 1976) helping to generate warm, dense maritime snow (Trujillo & Molotch, 2014) and some of the deepest snowpack in the Western U.S., with mean annual snowfall exceeding 15 m in some locations (Leffler et al., 2001).

Grand Mesa, located in western Colorado, is one of the largest high-elevation, flat-topped mountains in the world (Figure 2.1). Chosen as the primary site for NASA's SnowEx campaigns, the mesa spans an elevation range of 3000–3400 m a.s.l., with a relief of 1800 m above the surrounding valley floor (Kim et al., 2017). The mesa is dotted with lakes and reservoirs as well as isolated stands of Engelmann spruce and subalpine fir (Webb et al., 2020), which increase in density and coverage from west to east. The snow water equivalent (SWE) and elevation also increase from west to east (Kim et al., 2017). The geology of the mesa includes a cap of volcanic basalt, with exposures of the Green River shale and Wasatch sandstone formations (Austin, 2008; Yeend, 1969) on the surrounding hillslopes. During the accumulation season, prevailing maritime air masses coming from the Pacific generate cold, continental snowpack (Kulakowski & Veblen, 2006), with a mean annual snowfall of ~5–6 m (Austin, 2008).

2.3.2 Data

We obtained archived "System-ready" Level-1B (L1B) Maxar WorldView-2 and WorldView-3 satellite images acquired over our study sites between 2015 and 2019 under the NGA NextView license (Table 2.1). All the images were collected as in-track stereo pairs, with both panchromatic (PAN, 450–800 nm) and 8-band multispectral (MS, 397–1040 nm) images. The off-nadir viewing angles ranged between 6.9° and 33.5° (Table 2.1) with corresponding ground sample distance (GSD) values between 0.31–0.65 m for PAN and 1.25–2.59 m for MS images. Some of the images (e.g., 24 April 2018) of illuminated snow-covered surfaces suffered from limited detector saturation due to overexposure. Supplementary Table S1 includes additional information on commercial VHR sensors including the available products, spectral coverage, and GSD for data products.

Table 2.1. Metadata for WorldView-2 (WV-2) and WorldView-3 (WV-3) images analyzed in this study. Ground sample distance (GSD) is a measure of the ground-projected distance between the center of two adjacent pixels.

Location	Date	Sensor	Catalog ID	Mean Sun Elevation	Mean sun Azimuth	Mean Satellite Elevation	Mean satellite Azimuth	Mean off-Nadir Viewing Angle	PAN GSD [m]	MS GSD [m]
WA N Cascades	20 May 2015	WV-3	104001000C1BB800	60.0°	157.2°	58.5°	196.4°	28.3°	0.39	1.55
			104001000CB3D400	60.0°	156.7°	82.7°	6.6°	6.9°	0.31	1.25
WA N Cascades: South Cascade Glacier	24 April 2018	WV-3	104001003B034600	53.8°	162.7°	63.5°	145.3°	23.8°	0.36	1.45
			104001003B7AC300	53.8°	162.3°	61.4°	54.2°	26.0°	0.37	1.50
	27 May 2018	WV-3	104001003D88B900	63.0°	171.7°	61.7°	318.2°	25.7°	0.37	1.49
			104001003DD34200	63.0°	172.2°	58.9°	246.9°	28.0°	0.39	1.54
	5 May 2019	WV-3	104001004C8CF300	56.6°	158.2°	58.1°	132.6°	28.7°	0.39	1.57
			104001004CBC0600	56.6°	157.8°	58.8°	72.2°	28.3°	0.39	1.55
CO: Grand Mesa	1 February 2017	WV-3	1040010026C28A00	31.9°	160.5°	62.3°	137.0°	24.9°	0.37	1.47
			10400100276B9500	31.9°	160.2°	57.1°	52.7°	29.9°	0.40	1.59
			1040010028192C00	32.0°	160.7°	56.9°	152.4°	29.6°	0.40	1.59
			10400100286A3900	31.9°	160.4°	63.3°	65.7°	24.2°	0.36	1.45
	3 April 2018	WV-2	103001007A0DBD00	53.7°	290.0°	78.9°	109.4°	10.0°	0.48	1.91
			103001007B395800	53.7°	210.2°	52.7°	28.1°	33.5°	0.65	2.59
	26 March 2019	WV-3	1040010048434C00	52.3°	163.4°	62.7°	315.7°	24.9°	0.37	1.47
			104001004918EB00	52.3°	163.8°	55.8°	236.1°	30.7°	0.41	1.62

The WorldView-2 and WorldView-3 instruments include time-delay integration (TDI) linescan sensors with arrays of adjacent CCD detectors providing images across the full ~13–17 km wide swath. Each detector requires relative geometric and radiometric calibration, including dark offset subtraction and relative gain modifications, to produce the self-consistent Level-1B products. More information can be found in the technical note for WorldView-2 and WorldView-3 (Kuester, 2017).

We orthorectified the WorldView images using the ⅓-arcsecond (~10 m) seamless digital elevation model (DEM) from the USGS 3D Elevation Program (3DEP, formerly known as the National Elevation Dataset) (U.S. Geological Survey, 2017a). The 3DEP DEM basemap is a mosaic of terrain models with varying sources (i.e., airborne lidar and digitized contour maps) and collection dates (USGS, 2017). The source of the DEM timestamp for the Grand Mesa site (CO_MesaCo-QL2_2015) was 2016 and between 1958 and 2017 for the North Cascades site.

We adjusted the 3DEP DEM datums to the ellipsoid using the `dem_geoid` utility from the NASA Ames Stereo Pipeline (ASP) (Beyer et al., 2018; D. E. Shean et al., 2016).

2.4 Materials and Methods

2.4.1 Pre-processing

Figure 2.2 shows the general preprocessing workflow. We used the ASP `mapproject` utility with bilinear interpolation to orthorectify all PAN and MS L1B images to a common 1.2 m (for WorldView-3) or 2.0 m (for WorldView-2) grid using the rational polynomial coefficient (RPC) sensor models and the 3DEP DEM basemap with ellipsoid heights.

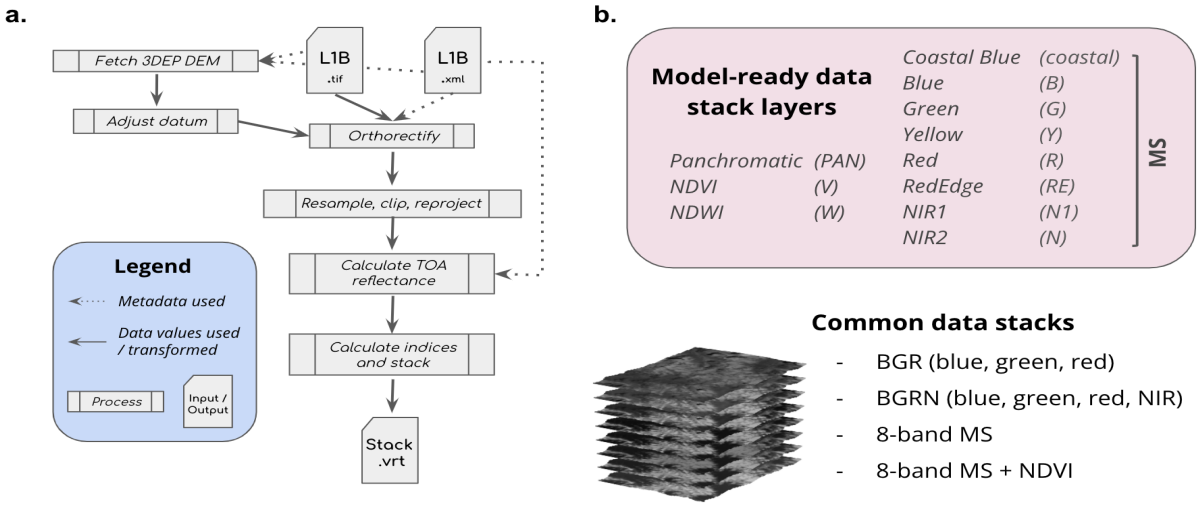

Figure 2.2. (a) Preprocessing workflow from Level-1B products to model input data stacks, (b) list of input data layers used to train random forest models and commonly available data stacks for spectral bands collected by earth-observing sensors. Table 2.2 contains all input data layer combinations used to train models presented in this work.

Additionally, we performed absolute radiometric corrections to convert the 11-bit L1B digital number (DN) values to top-of-atmosphere (TOA) radiance values for each band. We used the absolute calibration factors and effective bandwidths provided with the L1B metadata, along with the 2016v0.int version calibration adjustment factors (i.e., gain and offset) following the methodology outlined in Updike and Comp for WV-2 (Updike & Comp, 2010) and Kuester for WV-3 (Kuester, 2017). We then corrected for solar spectral irradiance to convert TOA radiance

to TOA reflectance values. Any TOA reflectance values less than 0.0 or greater than 1.0 were clamped to 0.0 and 1.0, respectively. TOA values exceeding 1.0 amounted to less than 1.5% of any single image. As mentioned earlier, one of our primary objectives was to evaluate the performance of simple landcover-classification approaches for accurate snow-cover mapping using VHR stereo images. Thus, we did not attempt to convert the TOA reflectance products to higher-level surface reflectance products, which would require more complex models and corrections for atmospheric, topographic, and bidirectional reflectance distribution function effects. See Section 5.2 for further discussion of these issues.

We also calculated common spectral indices and included them in our data stacks (Figure 2.2) to mitigate the impacts of sensor calibration error and shading and atmospheric variability. We calculated normalized difference vegetation index (NDVI) values using the red (626–696 nm) and NIR 1 (765–899 nm) bands, and normalized difference water index (NDWI) (McFeeters, 1996) using the green (507–586 nm) and NIR 2 (857–1039 nm) bands.

2.4.2 Classification

2.4.2.1 Machine Learning Algorithm Selection and Model Implementation

After preliminary exploration and consideration of classification algorithms (e.g., support vector machines, Gaussian Processes, and artificial and convolutional neural networks), we chose the random forest algorithm (Breiman, 2001) implemented in the `scikit-learn` Python package (Pedregosa et al., 2011). A random forest classifier leverages an ensemble of decision tree predictors to determine the class label for a given input. Preliminary testing ranged between 100 and 500 trees, but we observed little to no improvement in accuracy over the 100-tree configuration. As noted in the scikit-learn documentation, to control overfitting, the random forest implementation uses the highest average probability estimate from decision trees rather than majority voting to make class predictions. The final model parameters were 100 trees (*n_estimators*) with a square root maximum feature number (*max_features*) used for splitting and tree-building.

2.4.2.2 Model Input Configurations

We considered several variables when preparing the models: input data layer combination, time of year, and site/physiography of training images. All combinations of input data layers used to build the models are listed in Table 2.2.

Table 2.2. Input data stack layer combinations for all models described.

Data Stack	Input data stack layers
coastal_V	coastal, NDVI
coastal_VW	coastal, NDVI, NDWI
PAN_VW	panchromatic, NDVI, NDWI
BGR	blue, green, red
BGRN	blue, green, red, near infrared 2
BGRN_V	blue, green, red, near infrared 2, NDVI
MS	coastal, blue, green, yellow, red, red edge, near infrared 1, near infrared 2
MS_V	coastal, blue, green, yellow, red, red edge, near infrared 1, near infrared 2, NDVI
PAN_MS	panchromatic, coastal, blue, green, yellow, red, red edge, near infrared 1, near infrared 2
PAN_MS_V	panchromatic, coastal, blue, green, yellow, red, red edge, near infrared 1, near infrared 2, NDVI

Several different combinations of input data layers were created to emulate commonly available optical and near-infrared spectral combinations for earth-observing satellite sensors (e.g., RGB and RGBN, see Figure 2.2). As these models used a subset of bands from the WV MS products, they are referred to as "limited band models." These limited band models were included to assess the feasibility of using only a few spectral inputs to produce accurate land cover maps. Models trained using all eight MS bands (for both WV-2 and WV-3) as inputs are referred to as "8-band MS" models (e.g., MS, MS_V, PAN_MS, and PAN_MS_V). These models were constructed to leverage the full multispectral coverage available from Maxar VHR sensors.

For interpretation, models were also qualitatively subdivided into two categories based on the source of training data: single-scene or multiple-scene models. The single-scene models were trained on a data stack from a single date and location. These models were used to answer questions about models' spatiotemporal transferability (i.e., generalization)—how the site and time of year impact trained model performance (Section 3.2.6). The multiple-scene model

(M101) was trained on two input data stacks with different dates and locations (both North Cascades and Grand Mesa sites) with the full input of available spectral bands (i.e., PAN_MS_V) to assess training data expansion impacts on classification accuracy. Supplementary Table S2 provides details for the 155 model configurations we considered in this analysis, with their corresponding F-scores.

2.4.2.3 Feature Importance

Machine learning models with computable feature importance metrics, such as random forest, are useful for helping to identify the most important inputs, removing unimportant inputs, and reducing the feature space to balance computational efficiency with model performance. In this work, rather than reducing feature space, we primarily used feature importance to compare important inputs between different model configurations.

Both the Gini importance and permutation importance are implemented as feature importance metrics in scikit-learn. The Gini importance (mean decrease in impurity [MDI]) is a measure of how often a random pixel is incorrectly classified when given a random label according to the feature class label distribution. This makes the Gini importance strongly dependent on the class label distribution. Permutation importance (mean decrease in accuracy [MDA]) is the decrease in model score when a single feature class label is randomly shuffled and the linkage between data and label is deliberately broken. Through this shuffling, important features are identified as those that generate more error when intentionally mislabeled (Altmann et al., 2010). Despite the increased processing time required, we chose to analyze permutation importance (MDA) for its improved accuracy. Feature importance values were used to evaluate which inputs were consistently important for building models that could accurately classify land cover in the presence or absence of other input layers.

2.4.2.4 Training Data

Polygons outlining clusters of pixels for each feature class were manually delineated to use during model training for each unique image acquisition date (Table 2.1). We attempted to maintain uniform spatial distribution of polygons (Berhane et al., 2019) (e.g., Supplementary Figure S1) and attempted to minimize classification bias by balancing feature classes (Millard &

Richardson, 2015) with ~65,000 pixels (~0.1 km2 at 1.2 m GSD) per feature class. This was more challenging for snow-covered images with little exposed ground, surface water, and/or few clouds. Polygons were combined to create 3–6 different priority class labels: illuminated snow, shaded snow, vegetation, exposed ground, surface water, and cloud cover. In some cases, the number of feature classes varied with date; for example, the Grand Mesa model (26 March 2019; M17 in Supplementary Table S3) was trained with fewer feature classes than the North Cascades model (20 May 2015; M7 in Supplementary Table S3), as land and frozen water surfaces were heavily snow-covered. When exposed ground was visible at the Grand Mesa site, training polygons were focused on shale and sandstone units, as the basalt cap was covered with snow or vegetation.

2.4.2.5 Accuracy Assessment

Model development involved a train/val/test split of 70% training, 15% validation, and 15% testing (Koenig & Gueguen, 2016) for each unique image acquisition date, implemented in scikit-learn. We used 10-fold cross-validation to assess overall model performance (i.e., accuracy) and stability via *Precision*, *Recall*, and *F*-scores (Equations 2.1–2.3). Weighted macro *F*-scores were calculated to account for any training label imbalances by "weighting" label metrics based on the number of true instances (i.e., support). We did not use Cohen's Kappa, a traditional accuracy metric reported for land cover classifications, as other studies have indicated that it unnecessarily accounts for correct predictions due to chance (Elmes et al., 2020; Foody, 2020; Pontius Jr & Millones, 2011; Stehman & Foody, 2019).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{2.1}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{2.2}$$

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2.3}$$

2.4.2.6 Model Transfer Tests and Generalization

We attempted to assess model generalization using our pre-trained models for new images, with the goal of reducing duplicate training effort. After training, testing, and validating models for a single image (i.e., single-scene models), we conducted model transfer tests using a "base

model" with acquisition date that offered the best feature class distribution for each site. We used the M7 model (20 May 2015 image data stack) for the North Cascades site and the M17 model (26 March 2019 image data stack) for the Grand Mesa site base model. These tests were used to assess a single-scene model's ability to generalize to images from the same physiographic region and images outside the physiographic region in which the model was trained. We also conducted similar model transfer tests using the multiple-scene model (M101) to assess generalization to new images, though all images were from the same physiographic region based on the inputs for the multiple-scene model.

We expected physiography to affect both initial snow grain conditions and subsequent redistribution, metamorphism, and snowmelt processes, all of which impact the reflectance of snow. Each image in Table 2.1 was assigned to either the North Cascades region or the Grand Mesa region. Acquisition dates were thus grouped together as follows: 24 April–27 May for the North Cascades site and 1 February–3 April for the Grand Mesa site. Images acquired as in-track stereopairs inherently shared the same day of year and physiography, near-identical solar illumination, but different viewing geometry than the M7 and M17 models. The M7 and M17 models were trained on stereo image data stacks for the smaller off-nadir view angle. We define the "first" stereo image in each stereopair as the image with the smaller off-nadir view angle and define the "second" as the image with the larger off-nadir view angle.

These qualitative categorizations were used to inform interpretation of overall model generalization performance. Supplementary Table S3 shows the full set of model transfer experiments and corresponding F-scores.

2.4.2.7 Snow Cover Products and Comparison with Other Snow Cover Datasets

To assess the quality of coarser-resolution fractional snow cover products, we downsampled our VHR land cover classifications to prepare fractional snow-covered area products (WV fSCA). First, we extracted binary snow cover from the random forest classification maps by assigning a value of 1 to pixels classified as snow or shaded snow and a value of 0 to all other non-snow classes (Figure 2.3). We then reprojected and downsampled our VHR snow cover maps using an averaging algorithm to match the publicly accessible, "viewable" fractional snow-

covered area products from MODIS (MODSCAG fSCA, ~500 m GSD) (Painter et al., 2009) and canopy-adjusted products from Landsat (Collection 1 fSCA, ~30 m) (D. Selkowitz et al., 2017; U.S. Geological Survey, Earth Resources Observation And Science Center, 2018).

We created binary cloud masks using the VHR land cover classification and resampled these masks to produce WV cloud cover percentage products to match the coarser fSCA products. We excluded pixels with resampled cloud cover percentage values of 50% or higher. We also used the revised cloud mask products (REVCM) from Landsat and cloud flag values from MODSCAG and omitted these cloud-flagged pixels from analysis.

Finally, we subtracted the resampled, cloud-masked WV fSCA product from corresponding coarser resolution products to produce difference maps, and then calculated per-pixel statistics for each difference map and aggregated statistics for all difference maps at both sites (Supplementary Table S4). All products were reprojected to a local Universal Transverse Mercator (UTM) zone for visualization (Figure 2.3).

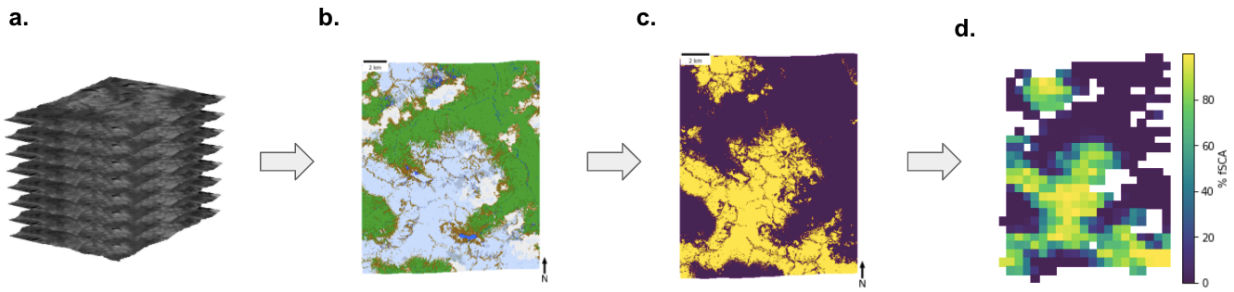

Figure 2.3. WorldView fractional snow-covered area (WV fSCA) workflow: (a) input data stack of WV images, (b) model-derived land cover classification, (c) binary snow map, and (d) WV fSCA map, downsampled to match the coarser resolution fSCA products for analysis.

We used the MODSCAG fSCA products from the same day as the corresponding WorldView image acquisition. The local time of observation for MODIS was ~10:30 AM +/− 5 min (Hall et al., 2019). We used the viewable MODSCAG fSCA products for comparisons with our resampled WV fSCA products, as they identify snow cover that is unobstructed from the sensor's view. The more recent "on-the-ground" STC-MODSCAG fSCA products (Rittger et al., 2020) include additional corrections based on forest canopy to estimate occluded snow cover, but these products were not publicly available at the time of this analysis.

We used the Landsat Collection 1 fSCA products with the shortest possible temporal offset from the corresponding WorldView image acquisition (~2–5 days, Supplementary Table S4). The limited temporal offsets should not significantly impact our fSCA comparisons. There were no new precipitation events and no large wind events that could have resulted in substantial redistribution between Landsat and WorldView collections. The local time of observation for Landsat 7/8 was ~10:00 AM local +/− 15 min (Arvidson et al., 2001). The Landsat Collection 1 fSCA products have all undergone canopy correction post-processing and are thus "on-the-ground" datasets. Viewable fSCA have since been made available for Collection 2, but these products were not released for our sites and acquisition dates at the time of this analysis.

We expected the canopy-corrected Landsat fSCA products to display greater differences over dense vegetation than the viewable MODSCAG fSCA or WV fSCA, although we recognized that the magnitude would vary based on the actual view angles for each of these sensors. To avoid potential issues related to view angle, we focused our comparisons on "open" areas, defined by a fractional vegetation cover product (WV fVeg) for subsequent aggregation. As with the WV fSCA products, we used a binary vegetation/non-vegetation classification approach followed by average downsampling to match coarser resolution fSCA product grids. Open areas were designated as pixels where WV fVeg < 25% and densely vegetated areas as pixels where WV fVeg ≥ 25%, broadly following the canopy-cover-density classes used by McGrath et al. (McGrath et al., 2019).

We also compared per-pixel fSCA values for completely snow-free (0% fSCA) or completely snow-covered (100% fSCA) pixels to assess snow detection performance. This approach was adopted to minimize the impact of mixed pixels for fSCA comparison at the coarse sensor resolution, with the presence of only the snow endmember or complete absence of the snow endmember. Accurately detecting snow absence (i.e., completely snow-free pixels) is integral to calculating snow disappearance dates and constraining the surface albedo evolution for energy balance and hydrologic models (Nolin et al., 2021). Compared to intermediate fSCA ($0\% < x < 100\%$ fSCA) values, we expected good agreement between the three fSCA products for completely snow-free and completely snow-covered pixels.

2.5 RESULTS

2.5.1 Single-Scene Model Performance

The 10-fold cross-validated F-scores using the validation sets from the training data discussed in Section 3.2.4 show that the single-scene models generated accurate classifications with all model configurations, attaining F-scores of 84% or higher (Supplementary Table S2). The models trained with 8-band MS inputs (e.g., M7–M10 in Supplementary Table S2) displayed the highest accuracy (F-scores > 99%), with minimal class confusion and both precision and recall scores nearing 100%. Figure 2.4 shows the classification results for two of the base 8-band MS single-scene models (M7 and M17 in Supplementary Table S2) for each study site.

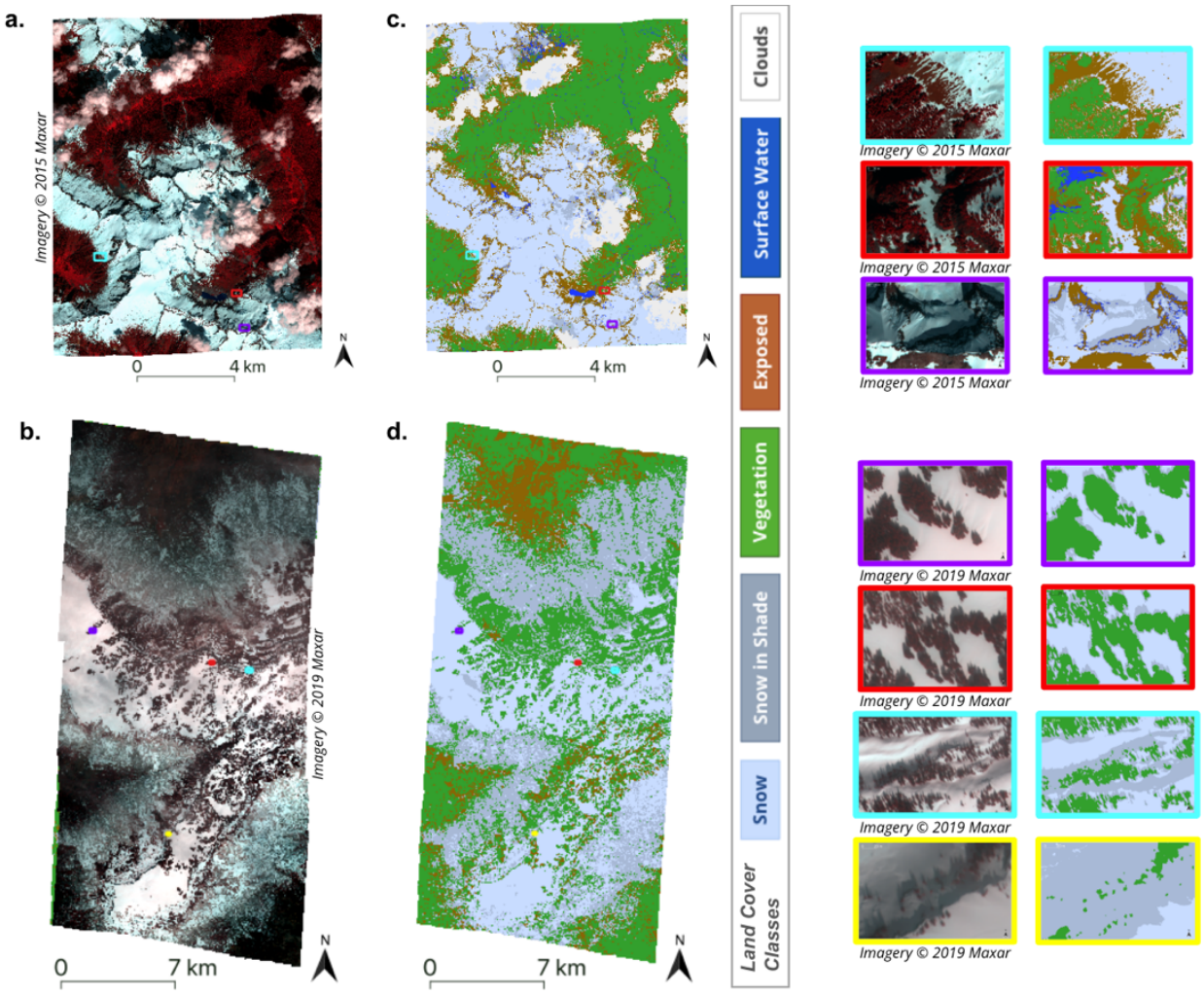

Figure 2.4. (a) Color-infrared WorldView-3 (WV-3) images acquired on 20 May 2015 over the North Cascades site and (b) 26 March 2019 over the Grand Mesa site. Corresponding land cover classification maps from random forest models (c) M7 and (d) M17 (see Supplementary Table S2). Zoomed insets show detail for corresponding outline colors in (a)–(d).

In general, the snow, shaded snow, and vegetation classes were correctly classified (e.g., Figure 2.4). Exposed surfaces, cloud, and water classes were the most common feature classes for which the pixels were incorrectly classified (i.e., errors of commission). These misclassified pixels were located most often at the boundaries of feature classes (e.g., pixels between snow and exposed ground misclassified as cloud cover) and in dark, shaded areas with low reflectance values (which were misclassified as surface water).

The limited band models (e.g., M1–M6 in Supplementary Table S2) consistently produced lower classification accuracies than the models that included all eight MS bands (e.g., M7–M10 in Supplementary Table S2). The models trained with a single spectral band and normalized indices (i.e., PAN_VW, coastal_V, and coastal_VW, such as in M1–M3 in Supplementary Table S2) performed comparably with the models trained on data stacks without normalized indices (e.g., M4, M5 in Supplementary Table S2).

2.5.2 Single-Scene Feature Importance

Feature importance tests for the varying single-scene model data stack inputs indicated that the coastal blue and NDVI inputs were important for both study sites (Figure 2.5). When restricting the model training stacks to these two inputs, the classifications were less accurate than the 8-band MS models, but still performed quite well (F-scores: 97.7% vs. 99.8%). The models with five or fewer input layers relied more heavily on individual inputs than the models with more than five inputs.

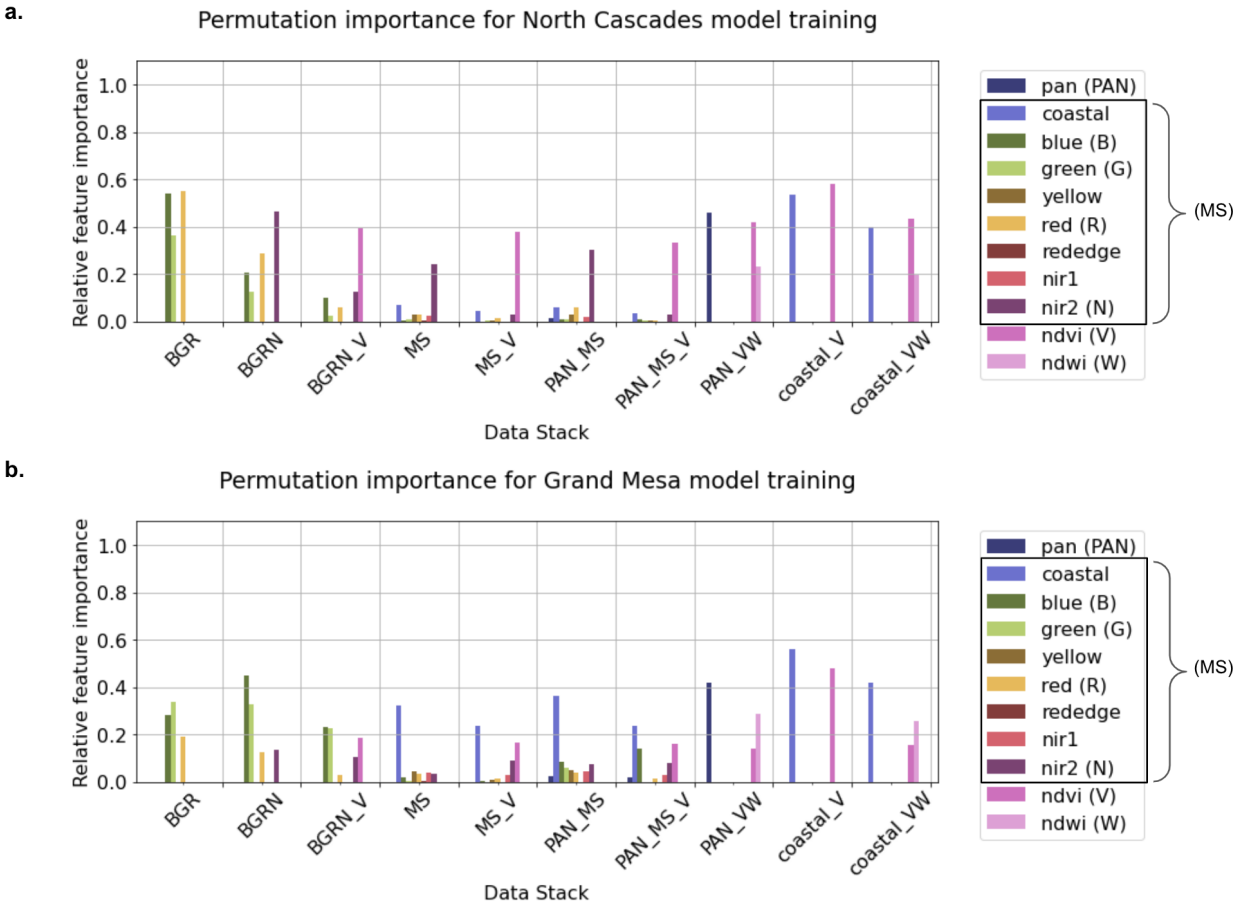

Figure 2.5. Permutation importance (mean decrease in accuracy [MDA]) for single-scene models with variable input data layer configurations (Table 2.2) using images acquired on (a) 20 May 2015 at the North Cascades site and on (b) 26 March 2019 at the Grand Mesa site. NDVI is consistently important (as is the NIR2 band, especially in the absence of NDVI), and the coastal blue band is important compared to other multispectral (MS) bands.

We observed different feature importance proportions for the two sites. The RGB models for the North Cascades site showed that blue and red inputs were more important than the green input, while the RGB models for the Grand Mesa site showed the opposite. What is additionally noteworthy is the apparent importance of the coastal blue input in the Grand Mesa models for several different data layer configurations.

2.5.3 Model Transfer and Generalization

Single-scene models for the North Cascades and Grand Mesa sites (M7 and M17, respectively, Supplementary Table S2) were capable of limited generalization. The best model

performance was observed when classifying a different image acquired for the same site during a similar time of year (i.e., the second image from an in-track stereopair for a model trained on the first image). As anticipated, the classification accuracy decreased when models were transferred and applied to images from different locations (Figure 2.6).

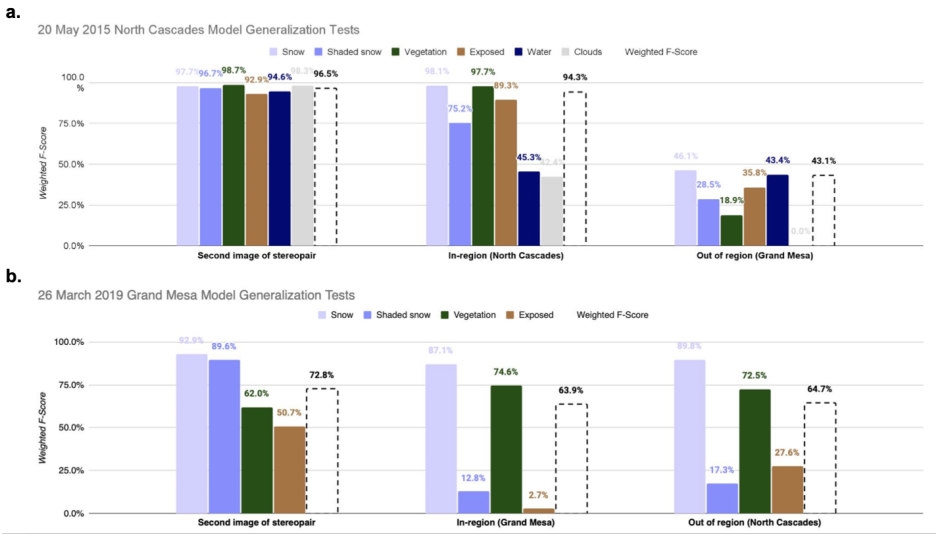

Figure 2.6. Aggregated results of the model transfer experiments presented in Supplementary Table S3 (the second image of an in-track stereopair, in-region images from the same site as the model, and out-of-region images from a different site than the model) for the: (a) 20 May 2015 North Cascades base model (M7) and (b) 26 March 2019 Colorado Grand Mesa base model (M17). Weighted macro average F-scores are shown for individual feature classes. The expected similarity of images to model training data decreases nonlinearly from left to right.

The models applied to the second image of an in-track stereopair (e.g., E1, E8, E15, and E22 in Supplementary Table S3) were the most accurate (F-scores: 96.5% in the North Cascades and 72.8% in Grand Mesa for M7 and M17 models, respectively) while the models applied to the in-region images (e.g., E2, E3 in Supplementary Table S3) had the second highest accuracy (F-scores: 94.3% in the North Cascades and 63.9% in Grand Mesa). Figure 2.6 shows the F-scores by feature class, ranging between 50.7–98.7% for the second image of an in-track stereopair, and much more widely (F-scores: 0–98.1%) for the other tests. The Grand Mesa model showed a good out-of-region performance for illuminated snow (average F-score: 89.7%), while the North Cascades model did not perform as well (average F-score: 46.1%).

The multiple-scene model (M101 in Supplementary Table S2), built from WV-3 images over both the North Cascades and Grand Mesa sites, performed better than any of the 8-band MS single-scene models for out-of-region transfer (Supplementary Figure S2, Supplementary Table S3). However, single-scene models produced more accurate classifications than multiple-scene models when employed in-region (Supplementary Figure S2), that is, for the second image of a stereopair or an image acquired on a different date for the same site. Based on the higher in-region transfer accuracy and single-scene stack accuracy presented in Section 4.1, we used the 8-band MS single-scene models (e.g., M7, M17, M27, M128, and M138 in Supplementary Table S2) for subsequent snow classification and fSCA comparisons.

2.5.4 Snow Classification Comparisons

2.5.4.1 Qualitative Assessment of fSCA Difference

When compared to the downsampled WV fSCA products (Figure 2.7), both the Landsat and MODSCAG products showed higher fSCA values (e.g., Figure 2.7, middle and bottom rows in blue) near areas classified as vegetation. The MODSCAG fSCA products showed differences in fSCA estimates over large areas classified as snow (Figure 2.7). The largest fSCA differences were observed for the Grand Mesa site, where MODSCAG had higher fSCA values both on and off the mesa, and lower fSCA values on adjacent slopes.

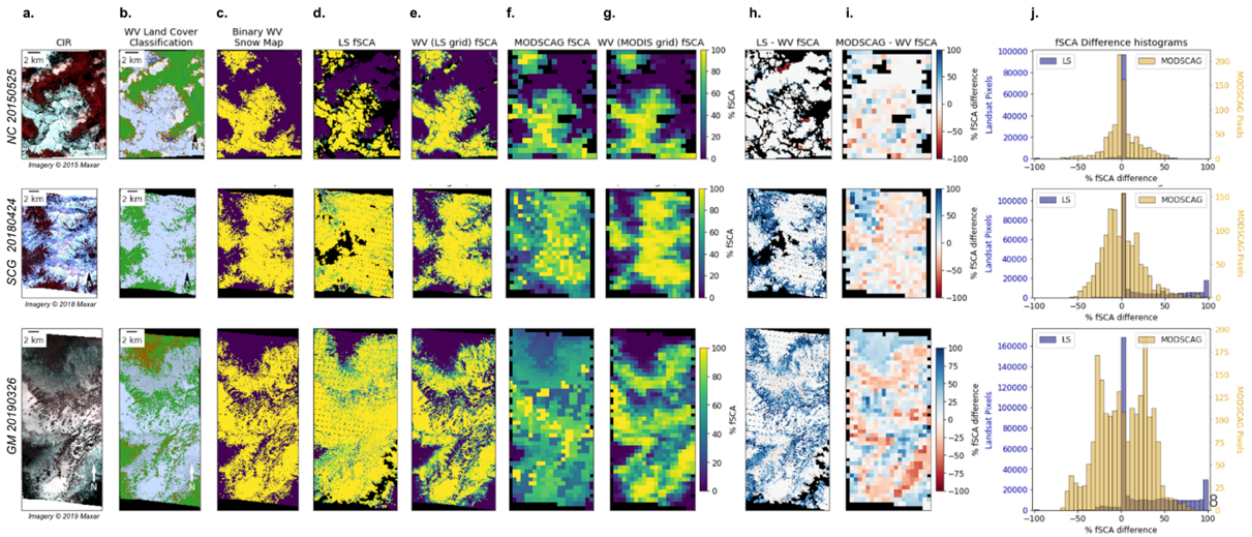

Figure 2.7. Selected snow classification comparison results for the North Cascades site (top row: 20 May 2015, middle row: 24 April 2018) and Grand Mesa site (bottom row: 26 March 2019). (a) WorldView

color-infrared context image, (b) land cover classification map and (c) binary snow cover map, (d) Landsat fractional snow-covered area (fSCA), (e) WV fSCA for Landsat grid, (f) MODSCAG fSCA, (g) WV fSCA for MODIS grid. Difference maps of fSCA values for (h) WV fSCA subtracted from Landsat fSCA in Landsat grid, and (i) WV fSCA subtracted from MODSCAG fSCA in MODIS grid with (j) corresponding histograms. Missing data shown in black for all panels.

Due to the nature of the available VHR in-track stereo collections, most of the images in this study had mean off-nadir viewing angles larger than 20° (Table 2.1), which can result in occlusions near trees and high-relief terrain. Comparisons of the per-pixel fSCA values with only the smallest mean off-nadir viewing angle image (CATID: 104001000CB3D400, off-nadir viewing angle: 6.9°) showed smaller fSCA differences (Supplementary Figure S3). Supplementary Figure S3 shows individual difference maps and histograms for all the comparisons in Supplementary Table S4.

2.5.4.2 Quantitative Assessment of Aggregated fSCA Difference

The aggregation of all fSCA difference products for both sites and all time periods (Supplementary Figure S3) showed good agreement between the coarse resolution fSCA products and the downsampled WV fSCA products (Figure 2.8). In the aggregate, the median difference was 0% for both Landsat fSCA and MODSCAG fSCA, but both products had slightly positive mean fSCA differences (Landsat mean of 14% and MODSCAG mean of 7%; Supplementary Table S5, Figure 2.8).

Further analysis over open the areas defined using the WV fractional vegetation cover (WV fVeg < 25%, see Section 2.4.2.7) showed no median difference (0%) between Landsat and WV fSCA and a small negative median difference (−2%) between MODSCAG and WV fSCA (Figure 2.8, Supplementary Table S5). Over dense vegetation (WV fVeg ≥ 25%), both Landsat and MODSCAG fSCA (Figure 2.8, Supplementary Table S5) showed higher median difference values (Landsat median of +25% and MODSCAG median of +5%). While the measures of spread were relatively stable for MODSCAG fSCA (~20–25%), the standard deviation and interquartile range (IQR) of the fSCA difference values was much lower for the Landsat fSCA estimates over open areas compared to densely vegetated areas (SD—19% vs. 35% and IQR—4% vs. 62%, respectively).

The analysis of the completely snow-covered pixels (100% fSCA) showed that both Landsat and MODSCAG fSCA products identified more snow-covered pixels than WV fSCA (Supplementary Table S6). Out of the 26,315 valid fSCA values for the common MODSCAG grid, MODSCAG identified 3418 pixels with 100% fSCA, while WV identified 257 pixels with 100% fSCA. In other words, MODSCAG overestimated the number of snow-covered pixels by a factor of 13.3. Out of the 4,515,817 valid fSCA values for the common Landsat grid, LS identified 2,090,153 pixels with 100% fSCA while WV identified 927,935 pixels—LS overestimated the number of completely snow-covered pixels by a factor of 2.3.

Analysis of the completely snow-free pixels (0% fSCA) showed that MODSCAG identified 2.7 times as many pixels as WV (MODSCAG—5717; WV—2139). Conversely, the Landsat fSCA products identified nearly the same number of pixels as WV (~0.8×; LS—1,165,256; WV—1,416,735), slightly underestimating the number of snow-free pixels.

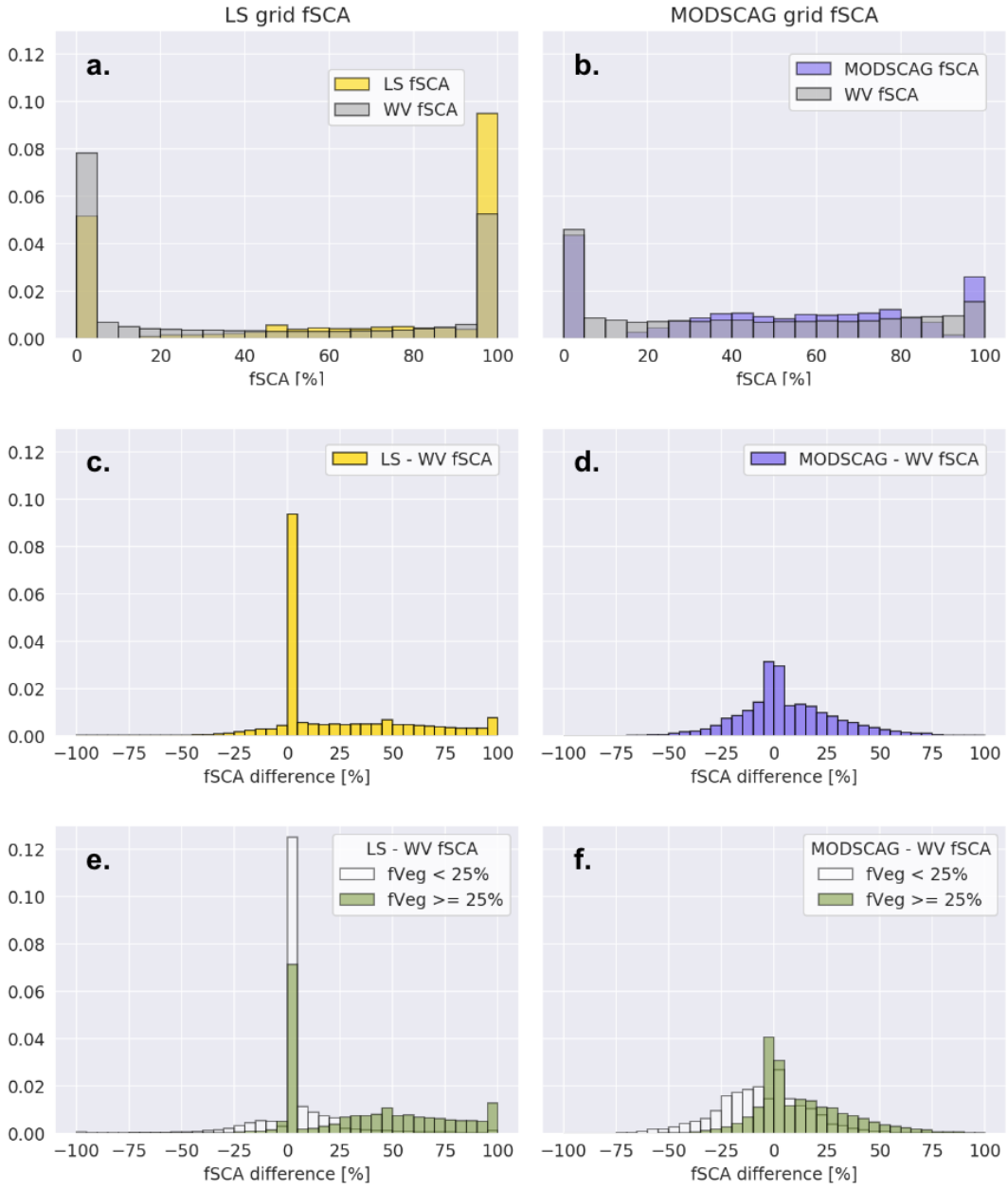

Figure 2.8. Aggregated statistics for all fractional snow-covered area (fSCA) comparisons (Supplementary Table S4, Supplementary Figure S3). Top row shows histograms of fSCA for (a) Landsat grid, (b) MODIS grid. Middle row shows histograms of per-pixel fSCA difference for (c) WV fSCA subtracted from Landsat fSCA in Landsat grid, and (d) WV fSCA subtracted from MODSCAG fSCA in MODIS grid. Bottom row shows corresponding histograms of fSCA difference values when separated by vegetation density (WV fVeg) for (e) Landsat and (f) MODSCAG. Bin size is 5% fSCA for all panels and all histograms are normalized by bin height so that the integral sums to 1.

2.6 DISCUSSION

VHR land cover classification maps with short repeat time intervals can enable quantitative analyses of rapidly changing landscapes. These detailed maps can be used to delineate feature classes of interest (e.g., snow, vegetation, and exposed surfaces), track their evolution, and evaluate/improve the accuracy of coarser land cover and snow cover products. Beyond these tasks, accurate single-scene land cover classifications offer valuable, dense labels for training deep learning models (Sun et al., 2017).

2.6.1 Single-Scene Models

Our results show that the single-scene random forest model configurations produced accurate snow and land cover classification maps using WV images for both study sites. Despite the higher resolution and fewer mixed pixels compared to the coarser resolution products, the qualitative assessment of the WV classifications indicated that the misclassification of mixed pixels (e.g., pixels along two feature class boundaries) remains a challenge. The feature importance tests highlighted the importance of the NDVI input across the single-scene models trained at each of the sites. At both sites, the limited band models produced classifications with accuracies comparable to those produced by the 8-band models. This suggests that despite the limitations in multispectral coverage, sensors with high radiometric quality and fine spatial resolution may still produce broad land cover classifications at accuracies that compete with more extensive spectral coverage.

Despite attempts to define small, well-distributed training data polygons, the semivariogram tests for each feature class indicated that the reflectance values and single-scene accuracy assessments were affected by spatial autocorrelation in the training data. The common practice of using a random cross-validation approach for our single-scene accuracy assessments resulted in inflated accuracy scores of >99%, as has been documented elsewhere (Ploton et al., 2020). To gauge the impact of this spatial autocorrelation, we created a separate set of test polygons spaced several hundred to several thousand meters from the initial polygons. Using the original models to classify these test polygons (Supplementary Figure S4), new accuracy assessments indicated

F-scores of ~94% for the North Cascades model (M7) and ~82% for the Grand Mesa model (M17), which are more representative of the models' performance.

2.6.2 Model Transfer and Generalization

In the model transfer tests, single-scene models performed well on similar images (i.e., similar physiography and illumination conditions) but did not generalize well to out-of-region images. The differing performance between the models (Section 0) can be attributed to the differences in the acquisition dates, the number of feature classes, and terrain-induced spectral variability within the feature classes. The topography of Grand Mesa and its influence on snow distribution may have also contributed to the differences in spectral variability of the training data and resulting model transfer. The consistently strong performance of the Grand Mesa model for classifying illuminated snow (Figure 2.6, F-scores > 87%) could have resulted from capturing a wider range of spectral variability for the illuminated snow class over open, windblown surfaces on the mesa. Its consistently poor performance in classifying both in-region and out-of-region shaded snow (Figure 2.6; F-scores < 20%) could have arisen from a smaller range in spectral variability due to the relative scarcity of shaded snow in the Grand Mesa images.

Overall, the transfer tests using the multiple-scene model showed better performance across for both sites than the single-scene models transferred to out-of-region images (Supplementary Figure S2). However, the single-scene models still outperformed the multiple-scene models when deployed in the same physiographic region as the model training site (Supplementary Figure S2). Both single-scene models (M7 and M17) had strong in-region performance, outperforming the multiple-scene model (M101), and possessed differing feature importance metrics, suggesting that regionally specific models may be more attainable than a global model (Hermosilla et al., 2022) when seeking strong generalization performance and accurate classification.

Our feature importance analysis (Section 2.4.2.3) also showed differences between sites, especially for the coastal blue band, the NIR2 band, and the NDVI inputs. The apparent importance of the coastal blue band across the model configurations could be partially related to the susceptibility of this shorter wavelength band (397–454 nm) to downwelling scattering and

increased reflected path radiance, especially from neighboring snow pixels, which are not accounted for in the TOA reflectance values.

Although it should have minor impacts on the variation of reflectance within the feature classes, atmospheric corrections to obtain surface reflectance (rather than top-of-atmosphere reflectance) may help reduce some variability in the absolute feature class reflectance values between images. Additionally, corrections for the topographic, view, and illumination effects are important in areas of rugged terrain (Lamare et al., 2020; Qiu et al., 2019) and may help to reduce reflectance variability within feature classes for improved model generalization and land cover classification. Using a hierarchical series of binary classifiers (Ham et al., 2005) rather than a single classifier for multiple feature classes could also improve model generalization by simplifying the classification tasks.

2.6.3 Snow Classification Comparisons

The VHR snow cover (SCA and fSCA) observations currently offer the finest spatial resolution products available for our study sites and may serve to complement coarser fSCA products with a better temporal resolution, spatial coverage, and historical archives. Our aggregate analyses showed good agreement between both the Landsat and MODSCAG fSCAs when compared to the WV fSCA over open areas. The Landsat fSCA had a particularly small IQR while both coarser resolution products had near-zero median differences. Larger off-nadir view angles arising from the in-track stereo collection strategy impacted WV viewable snow cover near trees and in areas of high relief, just as with oblique perspectives near the edges of the MODIS swath and corners/edges of the Landsat images (Rittger et al., 2020; Xin et al., 2012). This means that despite good overall agreement in the viewable snow cover, both the MODSCAG fSCA and WV fSCA underestimated the true amount of snow cover in the areas of dense vegetation due to occlusions. These findings highlight the significance of sensor view angle impacts for forest-snow analysis (Pestana et al., 2019) as well as the importance of adjusting for canopy cover (Rittger et al., 2020) to accurately estimate snow cover using MODIS. Future efforts to minimize off-nadir acquisition geometry and implement canopy corrections for VHR images could improve overall snow-cover-mapping accuracy.

The assessment of completely snow-free (0%) and completely snow-covered (100%) fSCA values showed that both coarser resolution products overestimated the number of completely snow-covered pixels (LS: ~2×; MODSCAG: ~13×), suggesting that despite the near-zero median per-pixel differences, the coarser products evaluated here tended to overestimate the total fractional snow cover for the full scene. While the MODSCAG products also overestimated the completely snow-free pixels to a lesser extent (~2.7×), the Landsat fSCA products more accurately identified these pixels and only slightly underestimated the number of snow-free pixels (~0.8× as many pixels as WV fSCA). Our initial analysis suggests that the Landsat fSCA may better characterize both completely snow-covered pixels and completely snow-free pixels than the MODSCAG fSCA, though a more detailed consideration of viewable vs. canopy-corrected products for additional sites/times is warranted. Fusion products leveraging the spatial resolution of satellites such as Landsat and Sentinel and the temporal coverage of sensors such as MODIS and VIIRS [78,90,91] may offer further improvements. These approaches could reduce snow-cloud discrimination errors, provide finer resolution observations on shorter timescales, and generate a longer and denser time series for evaluation. While not evaluated here, the VIIRS binary snow cover products posted at 375 m (Hall et al., 2019) and the spectral unmixing fractional snow cover products posted at 1 km (Rittger et al., 2021) may provide continuity for daily global snow observations as MODIS is decommissioned.

2.6.4 The Need for Fine-Scale Snow Cover

Fine-scale remote sensing observations are needed to accurately monitor changes in snow cover at critical locations and times. Along with the spatial boundaries of different land cover classes (e.g., near forest edges and within forest gaps), observing seasonal boundaries is also important. In late summer, sparse snow cover in the form of snow patches and drifts (Macander et al., 2015; Parr et al., 2020) support plant communities, maintain alpine meadows (Billings & Bliss, 1959), and can be used as indicators of climate change (Watson et al., 1994). In addition to shaping plant species diversity and distribution, snow patches can determine the timing and quantity of hydrologic and nutrient inputs (Björk & Molau, 2007; Marshall et al., 2019; Zong et al., 2022).

These patches can have outsized impacts, but can be challenging to observe with spaceborne sensors, which can result in inaccurate hydrologic modeling outputs. Budd Creek, an ephemeral stream in California's Sierra Nevada range, surrounded by granite peaks, exemplifies this issue (J. Lundquist, personal communication). Draining a northeastern cirque, late season streamflow in Budd Creek is sustained by snow-filled fissures and adjoining snow drifts. Small patches of snow are not easily detectable by 500 m MODSCAG fSCA products (Hao et al., 2019; Rittger et al., 2021), as they likely occupy small percentages of any given pixel. Though The Landsat fSCA products offer an improvement by showing later snow disappearance dates, these products are also unable to detect snow later in the season as Budd Creek continues flowing.

The fine-scale snow mapping approaches presented in this work could provide the observations necessary for detecting these small, ecosystem-sustaining snow patches. As the climate continues to warm, these outliers of late snow disappearance may grow in ecohydrological importance, controlling biotic range shifts and buffering temperature fluctuations to become the climate refugia of the future (Dobrowski, 2011; Ford et al., 2013; Lundquist & Flint, 2006).

2.6.5 Limitations and Considerations

One of the greatest strengths of WorldView-2 and WorldView-3 images—fine resolution—also presents one of the largest challenges for model transfer. The spatial variability of reflectance values for most surfaces is inherently linked to the spatial (and radiometric) resolution of the image. Though small pixels result in less spectral mixing from fewer feature classes in each pixel, there is also a wider range of reflectance values within each feature class. In other words, where unmixing methods are needed for other images, the WorldView-2 and WorldView-3 images can capture detailed surface properties—whether the snow is clean or dirty, whether the exposed surfaces are rock or soil, whether the trees are coniferous or deciduous, etc.

When aggregating these features with broad labels (i.e., snow, vegetation, and exposed), each feature class then contains more diverse values than what is observed and observable with coarse-resolution sensors. This makes a wide-ranging coverage of feature class representation an

important consideration when generating training data for machine learning models. A two-stage approach would likely improve model transfer, employing an initial evaluation stage to focus classification efforts in generating large datasets that more adequately represent intra-class heterogeneity. Generating synthetic datasets based on these datasets by systematically perturbing values for each band (i.e., data augmentation) may also improve generalization.

While the relatively high accuracy of these simple random forest model configurations for a handful of land cover classes is encouraging, validation sets need to be made more rigorous for the widely varying reflectance within feature classes. One approach is to assign labels pixel by pixel rather than by polygons. While more efficient, polygon-based delineation potentially captures more uniform spectral reflectance within a polygon than a comparable number of randomly selected, individually labeled pixels. A preferable compromise may be to use a patch of pixels with a standard size to randomly select batches of single feature classes to label. Furthermore, stratifying the training and test sets via block cross-validation is recommended to eschew biased non-spatial cross-validation in the accuracy assessments (Roberts et al., 2017).

The wide range and variability of observable reflectance values in our VHR images suggests that feature class expansion (more and narrower classes) could improve model accuracy, though potentially at the expense of model generalization. In other words, the number of feature classes needs to increase to manage larger ranges in the individual pixel reflectance values arising from a finer pixel resolution. In the case of accurately mapping small snow patches and snow boundaries, enumerating more targeted feature classes is preferable as there will be a larger proportion of boundary pixels with more distinct spectral signatures than those found in spatially contiguous areas of open snow. Extra care and attention in the training data curation is needed for areas with a rapidly changing snow presence (e.g., wind-scoured landscapes, ablation near the snowline during spring, etc.).

Traditional approaches rely on expert knowledge or unsupervised methods, which present challenges to including more training data. While the size of each VHR image (10^8–10^9 pixels per band for 1.2 m WorldView-3 MS products) precludes manually labeling each pixel, unsupervised methods such as spectral clustering can result in dozens of clusters within a single

feature class that still require manual aggregation and labeling (Bair et al., 2020). Though there are existing datasets that could be used to derive land cover feature class labels, such as the 30 m National Land Cover Database (Dewitz, 2021), these products are too coarse, outdated, or insufficiently labeled (i.e., they do not capture or represent seasonal snow or clouds as thematic classes) for our purposes. Potential solutions to the "training data bottleneck" include crowd-sourcing methods (Deng et al., 2009), attempts to adapt pre-existing labeled datasets (Sumbul et al., 2019), and weak labeling approaches (Zhou, 2018). The deep learning approaches developed by Cannistra et al. (Cannistra et al., 2021) and John et al. (John et al., 2022) produce binary snow cover maps from 3 m PlanetScope imagery using convolutional neural networks trained on thresholded aerial lidar snow-depth products. As mentioned earlier, the training requirements for neural networks greatly exceed those for random forest models—Cannistra et al. (Cannistra et al., 2021) trained their models on 370 million pixels compared to our ~65,000 samples per land cover class (~400,000 training pixels per model). Regardless of the machine learning approach applied, sufficiently representative and accurate training data labels for land cover beyond the built environment remains the primary challenge for accurate VHR optical land cover classifications.

A lesser confounder for model classification and transfer has been the magnitude and distribution of saturated pixels in the input data stack layers. In every combination of input layers, saturated regions in some bands have resulted in misclassification of the regions with extremely bright snow as clouds and vice versa. Dai and Howat (Dai & Howat, 2018) documented these saturation "striping" issues in the Level-1B WorldView images that can propagate to land-cover classification outputs. Though a few of our images were affected by these issues (Supplementary Figure S5), we observed limited impacts, such as classifying areas of relatively uniform snow features in one strip as 'shaded snow' and in another strip as 'illuminated snow'.

To develop robust pixel-based models capable of accurately classifying multiple land-cover classes in VHR images from different seasons, locations, and illumination conditions, we recommend exploring:

1. Improved preprocessing routines to obtain reliable absolute-surface-reflectance values for inter-image comparison;
2. A larger library of training data (i.e., more labeled images to cover the range of reflectance values within each class);
3. Simplified classification via hierarchical binary models;
4. Regionally specific models over global models.

These improved models can then be deployed for operational seasonal-snow monitoring, which in turn will offer a growing library of training data from which to adaptively improve the models. Adopting approaches from the field of computer vision for remote-sensing science transcend employing the latest machine learning techniques and algorithms. As we strive to imbue domain relevance and physical meaning into model assessment metrics, we must also generate benchmark datasets, training sets, and create baseline models upon which we can iterate and improve as a community.

2.6.6 Operational Potential of WorldView-2 and WorldView-3 Snow Cover Products

Operational snow cover products need to be timely, accurate, reliable, and accessible (Carroll et al., 2001; Hall et al., 2002). We demonstrated that the accurate pixel-based classification of snow in WorldView-2 and WorldView-3 images is possible without SWIR bands, but our current models are limited in their generalizability. Figure 2.9 shows the cloud-free archive of WV-2 and WV-3 images collected between 2009 and 2020 for the Western U.S, which can potentially provide such a sample of seasonal snow cover across the region.

It is important to remember that most VHR satellites are tasked—collecting images for predetermined points and areas—and do not constantly acquire regular images over systematic paths and rows such as Landsat or Sentinel-2. With strategic tasking for areas of interest at critical times of year, VHR images have the potential to provide fine measures of snow cover. Furthermore, the in-track stereo collections analyzed here offer two viewing perspectives to observe clouds and other feature classes, which can be combined to provide a measure of uncertainty characterization for the resulting classification products. Finally, beyond snow cover,

the in-track stereo collections can also offer precise snow depth measurements (Deschamps-Berger et al., 2020; McGrath et al., 2019), enabling many additional applications.

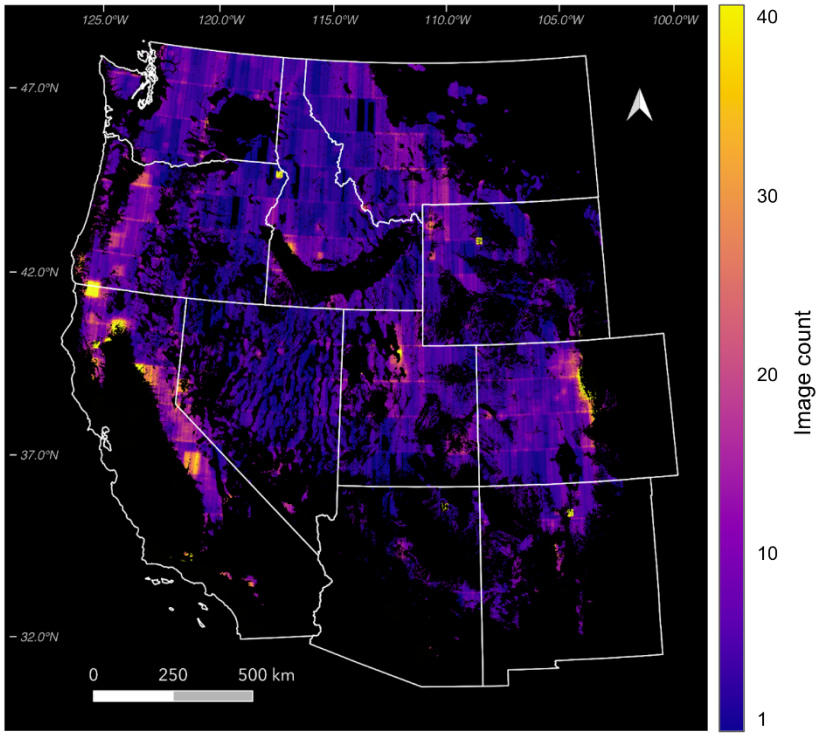

Figure 2.9. Heatmap showing coverage of multispectral WorldView-2 and WorldView-3 images in the Maxar archive with less than 25% cloud cover over the mountainous Western United States between 1 Oct and 1 Jun for the period spanning 2009 to 2020. Seasonal mountain snow mask adapted from (Wrzesien et al., 2019) and non-mountainous areas masked with black overlay.

2.7 CONCLUSIONS

Very-high-resolution images from WorldView-2 and WorldView-3 provide new opportunities and approaches for the detailed satellite mapping of seasonal snow cover in mountainous terrain, with many spectral bands and native image GSDs of 0.31, 1.2 m, and 3.7 m for PAN, MS, and SWIR bands, respectively. However, these detailed images involve large data volumes that prohibit manual analysis and flexible machine-learning approaches are required for large VHR satellite image archives.

We developed a suite of pixel-based classification models to better understand the potential for deriving land cover maps and snow cover products from WV images. We show that there is considerable potential for using random forest models to create classification maps (for a subset

of priority land cover classes) that can accurately map snow cover to complement and evaluate coarser resolution products. Even when limited to a few spectral band inputs, these RF models offer a high accuracy for snow, suggesting that images collected at fine spatial and radiometric resolutions could overcome limited spectral coverage, provided that the spectral bands are strategically located.

We observed the best model generalization performance for images from the same site and a similar time of year, with the highest accuracy in the illuminated snow class and the vegetation class for all the transfer tests, especially at the Grand Mesa site. Variability in feature class reflectance likely contributed to the poor performance with the attempted out-of-region model transfer, and this intra-class variability remains a challenge to developing a single, highly accurate, and robust multi-class land cover classification model for VHR images. Such a model will likely require larger libraries of training data, simplified feature classes, and potentially regional and seasonal specificity.

Comparisons of the completely snow-free and completely snow-covered WV fSCA pixels showed that both the Landsat and MODSCAG fSCA products identified many more completely snow-covered pixels when compared to the downsampled WV fSCA products (2× and 13×) but that the Landsat fSCA products more closely estimated the number of completely snow-free pixels than MODSCAG fSCA (~0.8× vs. 2.7×). Aggregate comparisons of all WV fractional snow cover products with coarser resolution fSCA products showed good agreement over open areas. The differences in the fSCA over dense vegetation between the three products can be partially attributed to differences in the viewing geometry and canopy correction approaches. Future snow cover analyses using VHR images should prioritize smaller off-nadir view angles and canopy correction to minimize such issues. Regardless, the growing archives of VHR satellite images offer potential for global seasonal snow observation, measurement, and monitoring efforts as both standalone products and when combined with complementary coarser resolution observations.

2.8 APPENDIX A

The following supporting information can be downloaded at: **https://www.mdpi.com/article/10.3390/rs14174227/s1**, Figure S1: Example of training data distribution and pixel counts per feature class; Figure S2: Multiple-scene model (M101) performance; Figure S3: Individual fSCA difference maps and histograms; Figure S4: North Cascades example of training and test polygon distribution and impacts; Figure S5: Saturation "striping" example; Table S1: Commercial very-high-resolution (VHR) sensors and products; Table S2: Study model configurations; Table S3: Model generalization/transfer test experiments; Table S4: Temporal offset between image collections for fSCA comparison; Table S5: Aggregated per-pixel fractional snow-covered area difference product statistics; Table S6: Completely snow-free and completely snow-covered fSCA pixel comparison.

Chapter 3. SIX CONSECUTIVE SEASONS OF HIGH-RESOLUTION MOUNTAIN SNOW DEPTH MAPS FROM SATELLITE STEREO IMAGERY

3.1 ABSTRACT

Fine-scale seasonal snow depth observations can improve estimates of snow water equivalent at critical times of year. Airborne lidar is the current gold standard for snow depth measurement, but it involves high costs and relatively limited coverage. Using very-high-resolution satellite stereo images from WorldView-2, WorldView-3, and Pléiades-HR 1A/1B, we produced a six-year time series (2017–2022) of spatially continuous DEMs for an 874 km2 study area over Grand Mesa, Colorado. We generated high-resolution stereo snow depth maps that capture intra- and interannual variability and span multiple anomalous years (58–158% of median peak SNOTEL snow depth). Comparisons with near-contemporaneous airborne lidar snow depth measurements showed good agreement, with median offset of -0.13 m, precision of 0.19 m and accuracy of 0.31 m. Our results suggest that satellite stereo can provide snow depth observations with the spatiotemporal coverage needed to improve operational forecast models and inform adaptive management strategies.

3.2 INTRODUCTION

Seasonal snow is a vital component of the climate system and the global hydrological cycle. The amount of water in seasonal snowpack can be estimated from snow depth measurements and modeled snow density. In mountainous, snow-dominant watersheds, these estimates are needed for streamflow forecast modeling and reservoir management. Knowledge of initial hydrologic conditions (i.e., snow depth at the beginning of the melt season) drives forecasting skill (Anghileri et al., 2016), so timely and spatially distributed observations of snowpack state are needed for accurate forecasting and management, especially during anomalous years (Modi et al., 2022) and in a changing climate (e.g., Gordon et al., 2022; Marshall et al., 2019).

Stereo photogrammetry with very-high-resolution (VHR, <1–2 m resolution) optical satellite images provides digital elevation models (DEMs) that can precisely measure snow surface

elevation. These "snow-on" DEMs offer fine-scale snow depth mapping when differenced with a snow-free DEM. Several recent studies demonstrated this potential using Pléiades-HR stereo images (0.5 m ground sample distance [GSD]) to produce meter-scale snow depth maps with 0.45–0.90 m accuracy and 0.35–0.80 m precision (Deschamps-Berger et al., 2020; Eberhard et al., 2021; Marti et al., 2016; Shaw, Gascoin, et al., 2020). Maxar's WorldView-3 and WorldView-2 satellites (WV) have 0.31 m and 0.46 m GSD, respectively, and state-of-the-art attitude control systems, which can offer improved accuracy for both DEM and derived snow depth maps. McGrath et al. (2019) showed that a WV-3 snow depth product had an accuracy of 0.24 m (RMSE) compared to ground-penetrating radar (GPR) observations.

Previous satellite stereo snow depth studies were limited to a single snow-on observation, except for Shaw et al. (2020), who documented snow depth over two non-consecutive drought years in the Andes. While these studies demonstrate that satellite stereo snow depth mapping can offer high-resolution "snapshot" observations, further testing and validation is needed to assess its repeatability for operational potential. In this work, we prepared and evaluated a 6-year snow depth time series from VHR stereo images acquired between 2017 and 2022 to study the detailed intra- and interannual spatiotemporal variability of mountain snow depth over a large area.

3.3 Study Site

The Grand Mesa, in Western Colorado, is the largest high-elevation, flat-topped mountain in the world. It ranges in elevation from 3000 to 3400 meters above sea level, with isolated stands of trees that increase in density from west to east (Figure 3.1). The mesa surface has a predominantly southwestern aspect and limited surface slopes (median: 3˚. These characteristics make Grand Mesa an ideal site to study the influence of vegetation on snow depth, which led to its prioritization for multiple NASA SnowEx campaigns (Kim et al., 2017). For the available 18-year record (water years 2006–2023), the median of peak seasonal snow depth measurements was 1.51 m at the Mesa Lakes SNOTEL station (Figure 1). During our six-year study period between 2017 and 2022, peak snow depth was 0.86 m (58% of median) in water year (WY) 2018, 2.36 m (158%) in WY 2019 and within ±0.22 m (±15%) of median for all other water years.

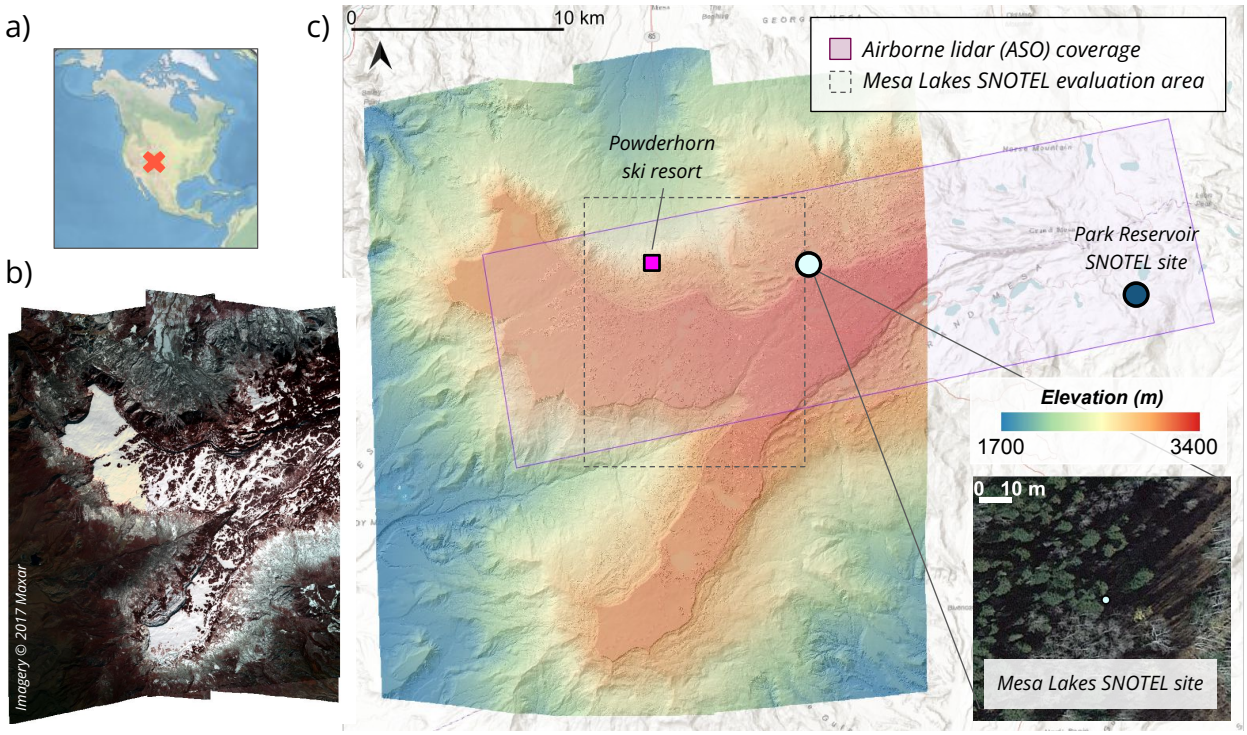

Figure 3.1 Grand Mesa, Colorado study site. a) Context map. b) Color infrared orthoimage mosaic of four WorldView-3 images acquired on 1 February 2017. c) Corresponding stereo DEM composite, covering 858 km^2. Inset shows Google satellite basemap for the Mesa Lakes Snow Telemetry (SNOTEL) site.

3.4 DATA

3.4.1 Commercial very-high-resolution satellite data

We used ten in-track stereo pairs collected by WorldView-2 and WorldView-3 between 1 February 2017 and 2 March 2022 (Table B 1). Two multi-view stereo collections, each with two overlapping in-track stereo pairs, coincided with the SnowEx campaign in February 2017. All other collections were acquired within two weeks of 1 April (commonly assumed timing for peak SWE) each year. We also processed two Pléiades-HR collections – one Pléiades-HR 1A (PHR1A) stereo pair from 31 March 2020 and one Pléiades-HR 1B (PHR1B) stereo pair from 31 March 2021. The WorldView collections covered 710–1200 km^2 while the Pléiades-HR collections covered 217–400 km^2 (

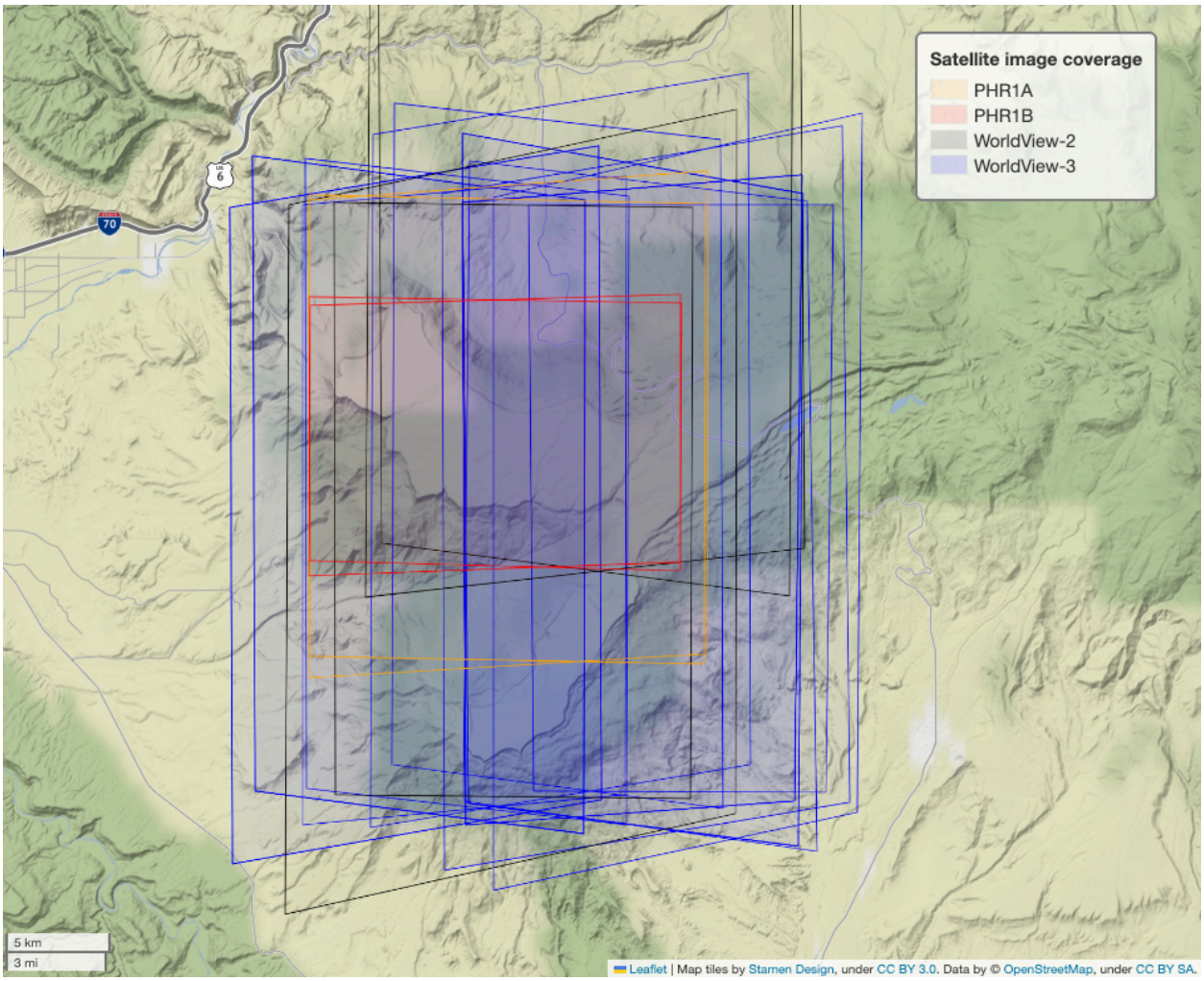

Figure B 1).

3.4.2 Land cover maps

Land cover maps (1.2–2.0 m GSD) were prepared from the WV multispectral images following the methodology outlined in Hu & Shean (2022) and used during DEM evaluation, co-registration, and filtering. We also used fractional snow-covered area (fSCA) products from the daily gridded MODIS Snow Covered Area and Grain size (MODSCAG) dataset (Painter et al., 2009).

3.5 REFERENCE AND VALIDATION DATASETS

3.5.1 Airborne lidar data

We used the Mesa and Delta County airborne laser scanning (ALS) data from the U.S. Geological Survey 3D Elevation Program (3DEP) to generate snow-free reference DEMs. Flights over Mesa County occurred 13-14 July 2016, while flights over Delta County occurred between 19 February 2016 and 31 October 2016.

We used two 3 m airborne lidar snow depth datasets collected by the Airborne Snow Observatory (ASO; Painter et al., 2016) on 8 February 2017 and 25 February 2017 (Painter et al., 2018) for comparison with the stereo snow depth products. The ASO flights covered 340 km^2 (Figure 3.1c).

3.5.2 In situ validation data

During the 2017 Grand Mesa SnowEx campaign, field teams measured snow depth using snow pits (Elder et al., 2018), probe transects (Brucker et al., 2018), ultrasonic arrays (Jennings et al., 2018), and GPR (Webb et al., 2019) . We used daily averages of the SnowEx ultrasonic measurements collected with a 15-minute sampling interval. All other in situ measurements were collected 6–9 days after the 1 February 2017 WV-3 stereo collection and 1–3 days before the 26 February 2017 WV-3 stereo collection (Table B 2). The longer time offsets (47 to 64 days) between the SnowEx 2020 campaigns (27 January to 13 February 2020) and the 31 March 2020 WV stereo collection precluded direct comparison.

We also analyzed daily snow depth measurements from ultrasonic sensors at the Mesa Lakes (622) and Park Reservoir (682) SNOTEL stations (Figure 3.1c) for the full study period.

3.6 METHODS

3.6.1 Stereo processing and DEM generation

3.6.1.1 Snow-on stereo DEM generation

We used the NASA Ames Stereo Pipeline (ASP; Beyer et al., 2018; D. E. Shean et al., 2016) to produce DEMs for all possible stereo pairs in Table B 1 following the methodology outlined in Bhushan & Shean (2021), with output DEM posting of 3 m to match the ASO lidar snow depth products. In-track stereo images were acquired approximately 50–75 seconds apart along the same orbit. We also processed all possible combinations of same-day cross-track pairs for each multi-view collection (Section 3.4.1, Table B 1). For each of these dates, we produced an initial per-pixel median composite, used the `dem_align.py` utility in the `demcoreg` package (D. Shean et al., 2021) to align each DEM to the initial composite, and then used the ASP `dem_mosaic` utility to produce a refined per-pixel median composite.

3.6.1.2 Snow-free reference DEM generation

We used the Point Data Abstraction Library (PDAL Contributors, 2022) to process the ALS data. We reprojected the snow-free ALS point clouds to a UTM 12N coordinate system (EPSG:32612) with heights above the WGS84 Ellipsoid. We filtered outliers and removed points classified as low noise (class 7), high noise (class 18) or water (class 20). We used the ASP `point2dem` utility to produce a lidar digital surface model (DSM) from first (or only) returns and a digital terrain model (DTM) from ground returns (class 2) using the weighted average of point elevation values within a 1-meter search radius of each 1-meter output grid cell.

3.6.2 DEM co-registration

We used the iterative co-registration approach described by (Nuth & Kääb, 2011) and implemented in `dem_align.py` in the `demcoreg` package (D. Shean et al., 2021) to align the snow-on stereo DEMs with the snow-free reference DTM.

We developed and tested the following co-registration approaches: "all-surface," "stable-surface," and "two-stage." The all-surface approach used all surfaces (including vegetation and

snow cover) common to both the snow-on and snow-free reference for horizontal and vertical alignment. The stable-surface approach limited the alignment to exposed, snow-free asphalt pixels over paved and plowed roads, identified in the WV land cover products. The two-stage approach we applied the horizontal translation from the all-surface approach, and then removed the median vertical offset computed over stable surfaces.

3.6.3 Snow depth derivation

We generated elevation difference maps by subtracting the snow-free reference DTM from the co-registered snow-on stereo DEMs. To distinguish snow-free from snow-covered areas, we prepared binary snow and vegetation masks from the contemporaneous multispectral land cover maps (section 3.4.2), and binary snow masks from daily MODSCAG fSCA products using a 20% threshold (2x detectable fSCA limit of 10%; Painter et al., 2009) when multispectral land cover maps were unavailable. We used a range filter to remove elevation difference values less than -0.3 m or greater than 10 m, similar to the filtering approaches in (Grünewald et al., 2013) and (Deschamps-Berger et al., 2020). We then used a 2D Gaussian filter (3x3 pixel, 1 sigma) to smooth any residual noise in the snow depth products. In total, we prepared eight snow depth maps from the stereo DEMs available between 2017 and 2022 (Table B 1).

3.6.4 Evaluation

We subtracted the reference ASO snow depth from the WV snow depth values at all 3 m pixels within the common 210 km^2 intersection area for two periods: "early February" (WV 1 February 2017 vs. ASO 8 February 2017) and "late February" (WV 26 February 2017 vs. ASO 25 February 2017). We also sampled the stereo and ASO snow depth rasters at the locations of all in situ measurements (Table B 2) using bilinear interpolation and subtracted the reference in situ snow depth values.

For each observation i of N total observations, we computed the difference (ΔSD_i) between the measured snow depth ($\widehat{SD_i}$) and the reference snow depth (SD_i). We report the median ($med(\Delta SD)$, offset), normalized median absolute deviation (NMAD, precision, Eq. 3.1), and

root mean squared error (RMSE, accuracy, Eq. 3.2), of the observed stereo snow depth residuals for each validation dataset.

$$NMAD = 1.4826 \times med(|\Delta SD_i - med_{\Delta SD}|) \quad (3.1)$$

$$RMSE = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} \widehat{SD}_i - SD_i\right)^2} \quad (3.2)$$

We also performed a detailed analysis for the "Mesa Lakes SNOTEL evaluation area" (MLSEA) – a representative 136 km^2 area on the mesa surface within ±100 m elevation of the Mesa Lakes SNOTEL site (Figure 3.1c). For each stereo snow depth product, we computed the median and spread (defined by the 16th to 84th percentile) over a common valid data region (~1.81x10^6 pixels) in the MLSEA and compared with the corresponding daily SNOTEL snow depth observations.

In addition to the above snow depth evaluation metrics, we computed the median and NMAD of residuals over exposed stable terrain (paved roads, Figure B 2), where snow depth should be 0 m (e.g., Hugonnet et al., 2022; Pelto et al., 2019). We are unable to report similar metrics for the ASO snow depth products, as all values over exposed road surfaces are manually set to 0 m (Deschamps-Berger et al., 2020).

To study measurement repeatability and interannual snow depth variability, we prepared a per-pixel median snow depth composite and per-pixel standard deviation map using the six stereo snow depth rasters closest to peak SWE timing in each year.

3.7 RESULTS

3.7.1 Snow depth spatial and temporal variability

We observe greater snow depth toward the east side of the mesa, where forest density is higher, compared to the exposed west side of the mesa (Figure 3.2). The stereo products provide snow depth retrievals in canopy gaps as small as 400 m^2 (less than half of a 30-m Landsat-8/9

pixel). In general, we observe greater snow depth within canopy gaps than in adjacent open areas. We also observe greater snow depth on steeper slopes within a few kilometers beyond the mesa edge, and thinner snow depth at lower elevations surrounding the mesa.

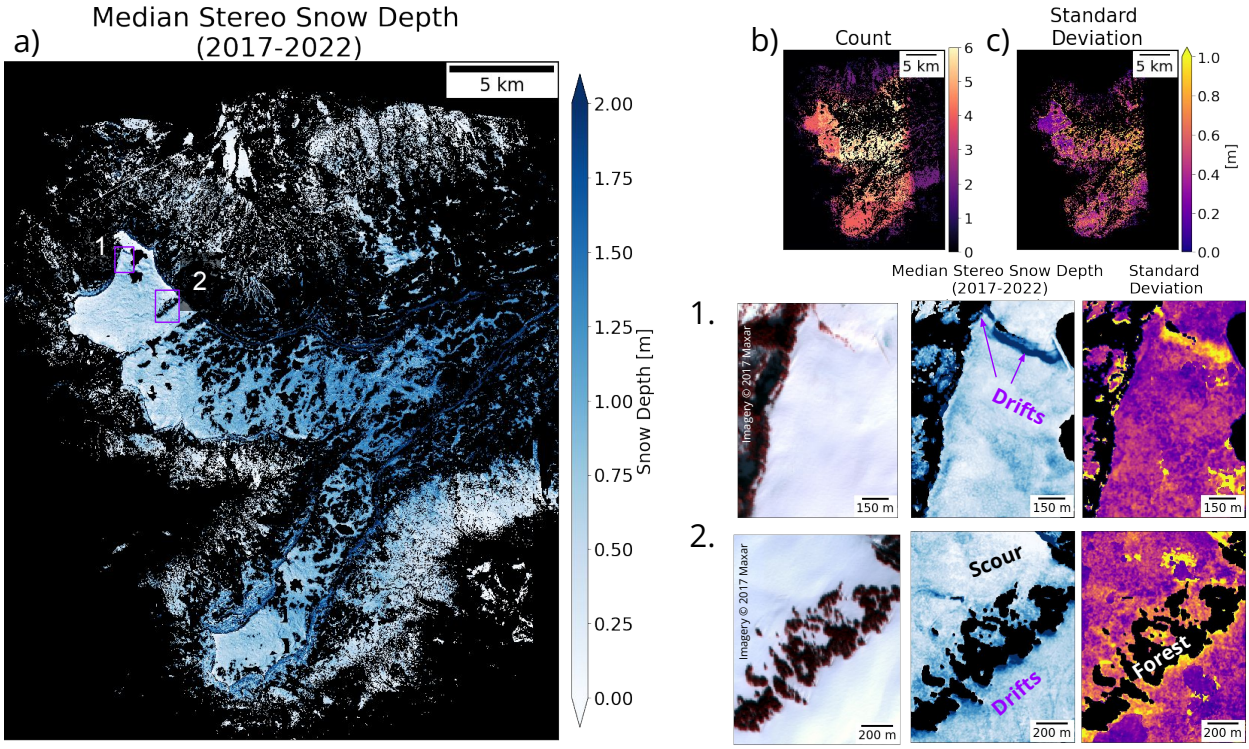

Figure 3.2 a) Per-pixel median composite b) count, and c) standard deviation for the six-year time series of peak snow depth products from satellite stereo DEMs. Insets 1 and 2 show detail of spatiotemporal variability for drifts near the mesa edge and a forest stand.

The six-year per-pixel median snow depth composite reveals areas of recurring snow depth patterns including deep drifts in local depressions, near dams, near the edges of forest stands, and below steep scarps (Figure 3.2). The thin snow near mesa edges and other areas of relatively low median snow depth (e.g., dam embankments) had low interannual standard deviation, while areas of higher median snow depth tended to have greater variability.

The stereo time series provides intra- and interannual measurements of snow depth (Figure 3.3). Repeat measurements also capture the characteristic spatial distribution of snow depth for the study site, which can be used to assess and identify issues with individual measurements. For

example, we identified systematic artifacts in the 2021 PHR1B products (Figure 3.3b, Figure B 3), which we excluded from analysis in sections 3.7.2 and 3.7.2.2.

In general, we observed a consistent spatial distribution of relative snow depth magnitude over the six-year period. However, the time series also reveals deviations from this spatial pattern, especially during extreme snow years (e.g., 2018 and 2019), which contributes to the observed interannual variability shown in Figure 3.2. For example, in the high snow year of 2019, areas with relatively high characteristic snow depth had proportionally more snow than areas with relatively low snow depth.

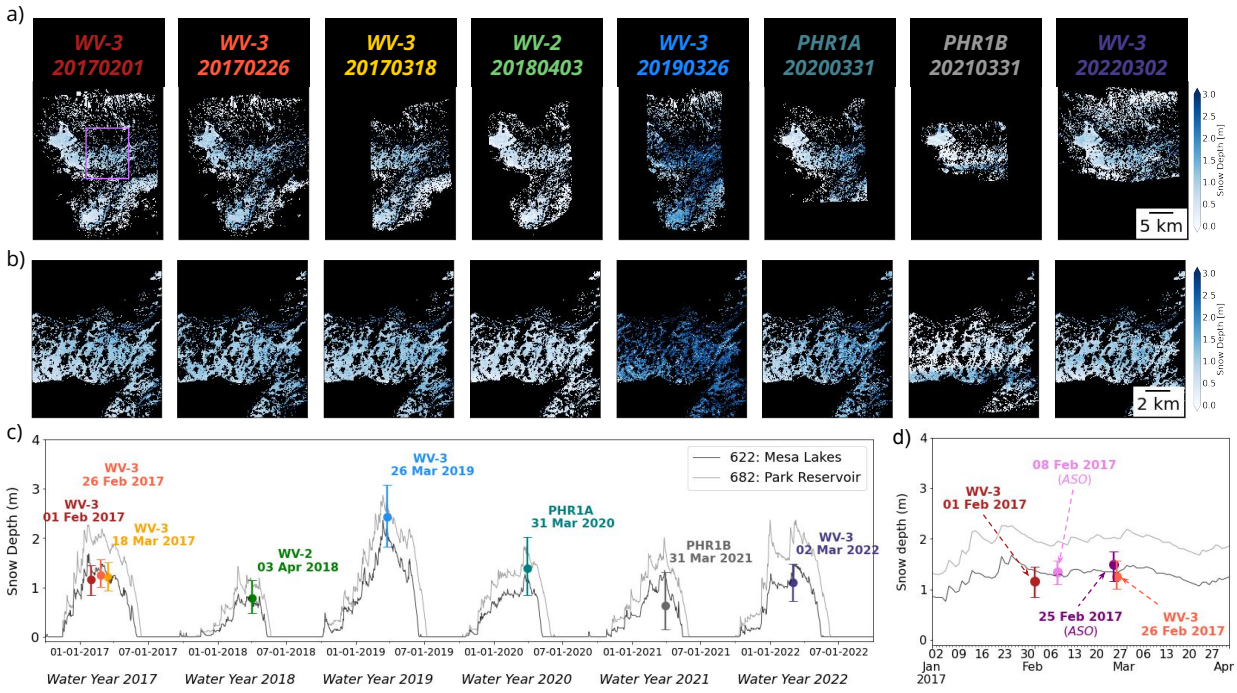

Figure 3.3 a) Time series of intra- and interannual stereo snow depth products generated from WorldView-2 (WV-2), WorldView-3 (WV-3), and Pléiades-HR (PHR1A, PHR1B) images. b) Detail of the Mesa Lakes SNOTEL evaluation area. c) Time series showing median (circles) and 16–84\% spread (bars) of stereo snow depth values within the area shown in b. Solid lines show SNOTEL daily snow depth. d) Detail of February 2017 stereo and ASO observations with solid lines showing SNOTEL daily snow depth as in c.

3.7.2 Stereo snow depth evaluation

3.7.2.1 Exposed, snow-free stable surfaces

We evaluated measurement repeatability by examining elevation difference values over stable surfaces in the MLSEA (Figure B 2). The February 2017 multi-view stereo products had the best precision (lowest NMAD values), while the 2019 stereo product had the largest vertical offset and worst precision (highest NMAD values), potentially due to some combination of stereo acquisition geometry and local snow conditions near plowed roads. Overall, the median and NMAD of elevation differences over stable surfaces (n = 5,899–7,120 pixels) ranged from 0.05–0.22 m and 0.17–0.38 m, respectively, for the 7 stereo products with limited artifacts (Figure B 2). Thus, we observe decimeter-scale precision over multiple years, providing confidence in measurement repeatability in both low and high snow years.

3.7.2.2 SNOTEL comparison

The median peak annual stereo snow depth in the MLSEA generally agreed with snow depth measurements from nearby SNOTEL stations (Figure 3.3c). We observed better agreement with the Mesa Lakes station (within ±15%, RMSE = 0.16 m) than the Park Reservoir station (within ±43, RMSE = 0.60 m).

The median of stereo snow depth values within the MLSEA was lowest in WY 2018 (0.8 m; 51% of the 18-year median for peak Mesa Lakes SNOTEL snow depth values) and highest in WY 2019 (2.4 m; 162%).

3.7.2.3 Airborne lidar snow depth comparison

In general, we observe good agreement between the stereo and ASO snow depth products (Figure 3.4). The snow depth difference values for the early February comparison had a Pearson's correlation coefficient (r) of 0.76 with median offset of -0.13 m (Figure 3.4g, NMAD of 0.19 m, and RMSE of 0.31 m (Figure 3.4).

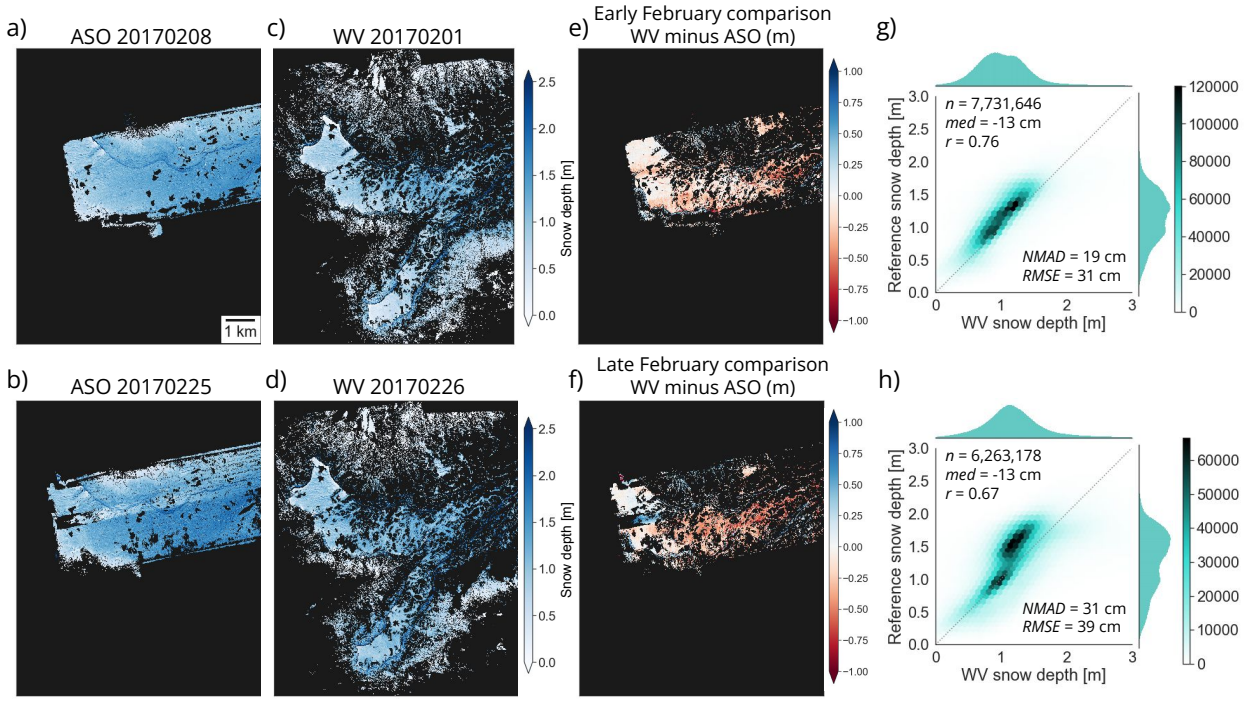

Figure 3.4 Snow depth products from (a-b) the Airborne Snow Observatory (ASO) and (c-d) WorldView-3 stereo, with (e-f) corresponding difference maps (stereo minus ASO) and (g-h) hexbin plots. See Figure B 4 for ASO vs. ASO comparison showing residual flightline swath misalignment artifacts in the late February ASO snow depth product.

Difference values for the late February comparison were less correlated (r = 0.67), with two distinct clusters in the hexbin plot (Figure 3.4h). The difference map reveals systematic offsets along the ~1 km wide ASO swaths (Figure 3.4f) due to residual relative swath alignment issues in the 25 February 2017 ASO snow depth product. A difference map between the 8 February and 25 February ASO snow depth products shows these artifacts, with apparent snow depth changes of -0.45 to +0.25 m (Figure B 4). The SNOTEL stations recorded limited (<0.10 m) change in snow depth over this period (Figure 3.3d).

3.7.2.4 In situ snow depth comparison

The satellite stereo snow depth observations showed good agreement with available in situ snow depth measurements (Figure 5), with decimeter-scale offset (-0.08 to -0.33 m), precision (NMAD: 0.13 to 0.46 m), and accuracy (RMSE: 0.23 to 0.48 m). The corresponding ASO snow

depth measurements had slightly smaller offsets (-0.08 to 0.09 m), better precision (NMAD: 0.09 to 0.20 m), and slightly better accuracy (RMSE: 0.14 to 0.46 m).

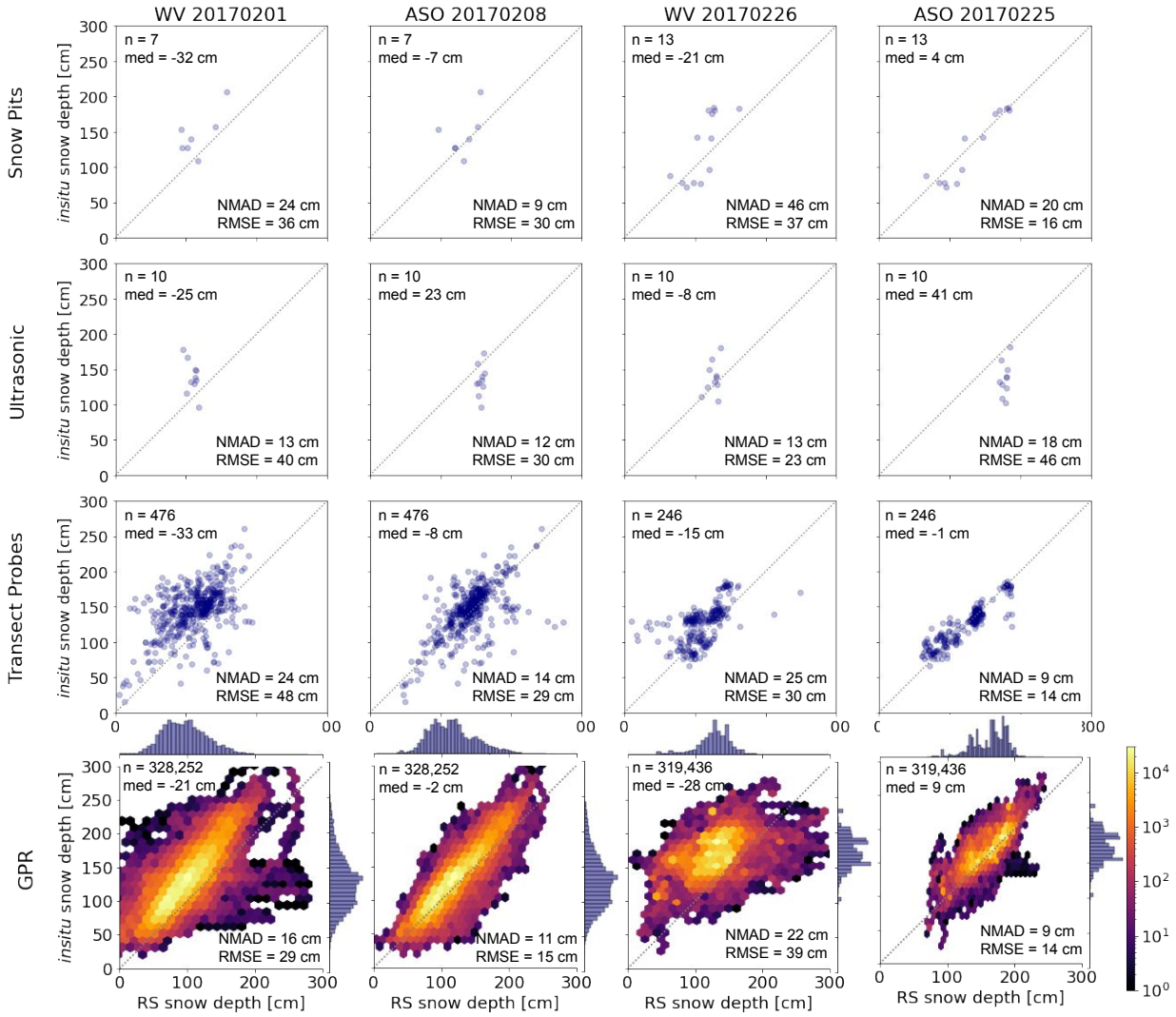

Figure 3.5. Comparison of sampled WV stereo and ASO snow depth values with in situ snow depth values for the early and late February 2017 campaigns, with dates of in situ data collection in Table S3.

3.8 DISCUSSION

3.8.1 Spatiotemporal variability

The Grand Mesa site and SnowEx campaigns provided an important opportunity to evaluate the satellite stereo snow depth time series. The snow depth time series revealed repeat scour on

exposed mesa edges and larger drifts on leeward forest edges along the northern arm of the mesa. The interannual snow depth variability in the MLSEA was generally higher in areas with greater snow depth (e.g., drifts). Furthermore, in extreme low and high snow years, observed snow depth in these areas appeared thinner and thicker, respectively, than would be expected from uniformly scaling the six-year per-pixel median snow depth magnitude. These results demonstrate the value of repeat stereo snow depth observations to characterize and understand complex spatiotemporal variability.

3.8.2 Stereo snow depth evaluation

Evaluations of WV stereo snow depth with SnowEx in situ measurements generally showed good agreement (section 3.7.2.4, Figure 3.5). The larger offsets observed for the early February probe depth comparisons were an exception, which may be explained by a combination of overprobing (Sturm & Holmgren, 2018), bias due to undersampling of thin snow (Currier et al., 2019) and probe geolocation uncertainty (+/- 2.5 m for magnaprobe; Sturm & Holmgren, 2018 and +/-5 m for manual probes (Currier et al., 2019).

Differences between the early and late February comparisons (Figure 3.5) may stem from larger temporal offsets between datasets in the early February comparison (6-9 days vs. 1–3 days in late February, Table B 2), as well as the respective sampling locations. We attribute the apparent stereo snow depth offset relative to the early February probe transect measurements and late February GPR measurements (Figure 3.5) to in situ sampling prioritization near forested areas, as stereo DEMs can have increased error and/or artifacts near forest edges. For example, the accuracy of the early February stereo snow depth in small canopy gaps (400-1000 m^2) appeared to be worse (RMSE: 0.38 m) than in open areas (RMSE: 0.28 m), compared to corresponding ASO snow depth measurements.

The -0.13 m offset of WV stereo snow depth relative to ASO snow depth (section 3.7.2.3, Figure 3.4) could be attributed to a number of causes, including different measurement geometry and temporal offsets between acquisitions. Here, we focus on causes that can potentially be corrected with improved processing, including residual co-registration error and systematic artifacts observed in both products.

3.8.3 DEM co-registration

Successful DEM co-registration is crucial to produce accurate snow depth estimates, as DEM geolocation error will introduce systematic slope- and aspect-dependent vertical errors (Nuth & Kääb, 2011). While our evaluation suggests that our two-stage co-registration approach worked well in most cases, we observed negative elevation difference values in wind-scoured areas on the northern mesa arm for some DEMs (e.g., 1 February 2017), where the snow-on stereo DEM was lower than the snow-free reference DTM.

Co-registration could be improved by posting stereo DEMs at 1 m, rather than matching the 3 m ASO posting. Coarser postings can smooth true surface roughness, and lead to less precise mapping of stable surface boundaries (e.g., separating plowed road surfaces from high snowbanks along shoulders). For scenes with mostly snow-covered and/or vegetated surfaces, the limited spatial distribution and small area of exposed, stable surfaces can also reduce the accuracy of vertical offset corrections during co-registration. For Grand Mesa, most of the exposed, stable surfaces were plowed roads on north-facing slopes on the north side of the mesa. Larger snowbanks and associated differences in the plowed road extent in 2019 could have affected our two-stage co-registration approach, which would explain the larger observed bias and reduced precision for the 2019 product (section 3.7.2.1, Figure B 2). While manual selection of a unique set of stable surfaces for each stereo DEM would likely improve both the accuracy and precision of the resulting stereo snow depth maps, our study involves a production-ready workflow geared towards operational stereo snow depth mapping. One simple solution to mitigate these issues would be to expand future stereo acquisition footprints to include more stable, snow-free surfaces at lower elevations. Alternatively, areas with thin snow, such as wind-scoured surfaces near the mesa edge, could serve as additional control surfaces for smaller DEMs lacking exposed snow-free surfaces.

3.8.4 Sensor-specific errors

Beyond geolocation and sampling considerations, sensor-specific acquisition and processing errors can affect snow depth measurement accuracy. For example, the observed residual

systematic offsets between flightlines in ASO snow depth products (Figure 3.4b, S4) are much larger than published error estimates (RMSE 0.08 m; Painter et al., 2016).

Unmodeled attitude error ("jitter") in linescan images can result in stereo DEM artifacts with ~0.5–1.0 pixel (±0.2–0.3 m for WV-3) magnitude (Figure B 3). Known issues with the PHR1B satellite during our study period resulted in 0.6–1.4 m artifacts in the 31 March 2021 collection (Figure B 3), which is consistent with other published values (0.6 m; Deschamps-Berger et al., 2020).

Many previous studies attempt to correct these systematic artifacts in the DEM difference products (e.g., fitting higher-order polynomials to residuals and removing). This approach requires an adequate distribution of samples over stable terrain to avoid overfitting, which can remove real signals of snow depth variability. Due to the limited distribution of stable terrain for our study site and DEM footprints, and the comparable magnitude of these artifacts and observed snow depth (1–2 m), we did not use these approaches to correct our snow depth products. Future work may benefit from a new tool (ASP `jitter_solve`) to correct these attitude errors in each of the original linescan camera models (rather than the derived elevation difference maps), which should improve future stereo snow depth precision.

3.8.5 Comparison with previous studies

The multi-view WorldView-3 stereo snow depth maps in this study offer higher accuracy and precision (RMSE: 0.31–0.39 m, NMAD: 0.19–0.31 m) than stereo snow depth maps from previous studies using two-image or triplet Pléiades-HR stereo (RMSE: 0.8–1.2 m; NMAD: 0.65–0.69 m), when compared with ASO measurements.

The comparison between our 1 February 2017 WV snow depth map and the 8-10 February 2017 GPR measurements (Figure 3.5) shows similar results to those reported by McGrath et al. (2019), with similar accuracy (RMSE: 0.29 m vs. 0.24 m in McGrath et al., 2019), higher precision (NMAD: 0.16 m vs. 0.24 m), and larger median offset (-0.21 m vs. -0.03 m). We attribute residual differences in these comparisons to differences in stereo DEM posting (3 m

here vs. 8 m) and methodology (e.g., vegetation mask preparation, co-registration approach, raster sampling strategy).

3.8.6 Towards operational snow depth from space

Our results show that VHR satellite stereo can complement existing operational airborne and field efforts to measure surface elevation and map snow depth. Satellite stereo can also offer greater spatial and temporal coverage compared to current airborne approaches, with acceptable accuracy and precision for many applications, especially for deeper snow. Several challenges remain, however, and accurate, operational snow depth retrievals require timely snow-on stereo image acquisition, access to high-quality snow-free digital terrain models, and robust co-registration approaches.

One of the major limitations of satellite stereo is the inability to directly measure snow depth beneath dense vegetation. Stereo-lidar fusion, however, offers a promising solution. We envision approaches that rely on initial ALS observations to capture characteristic snow depth distribution over both open and vegetated areas for a given basin. Statistical models trained on these ALS observations can then leverage repeat, multi-view stereo collections over open areas and canopy gaps to generate spatially continuous, basin-wide predictions of snow depth, with corresponding SWE estimates from empirical density models (e.g., Hill et al., 2019).

The commercial VHR satellite image archives span the past ~15 years, offering a valuable sample of historical snow depth observations across the Western U.S. and the world. There are no direct data acquisition and delivery costs for federal use under existing licensing agreements, and all processing software is open source with permissible licenses. The “on-demand” stereo tasking capabilities offer rapid data collection, and our processing workflows can produce DEMs and snow depth measurements within 24-48 hours of image collection. With sufficient planning and prioritization, systematic satellite stereo tasking campaigns can provide a representative sample of regional snow depth observations needed to meet operational demands.

3.9 Conclusions

We produced, evaluated, and analyzed six consecutive seasons of very-high-resolution satellite stereo snow depth maps over Grand Mesa, Colorado. Our stereo snow depth time series captures characteristic spatial patterns, depth-dependent interannual variability, and the evolution of small-scale features such as drifts and wind-scoured surfaces. We observed good agreement between stereo snow depth measurements and in situ snow depth measurements from SnowEx campaigns, in situ SNOTEL records, and airborne lidar snow depth measurements. Our results offer improved accuracy, precision, and spatial coverage compared to previous evaluations of VHR stereo snow depth products. Multi-view stereo collections offer broader coverage, improved precision, and better snow depth retrievals near forest edges and within canopy gaps. Repeat measurements demonstrate consistent decimeter-scale accuracy and precision, even in years with anomalously high or low snow depth. Collectively, our findings highlight the potential of repeat high-resolution satellite stereo snow depth observations, which can complement current approaches and provide the spatiotemporal coverage needed for operational water resource management in remote, data-scarce regions.

3.10 Open Research

3.10.1 Data

The Maxar WorldView Level-1B stereo images used to generate the DEMs for this study are available under the NGA EOCL license. The derived stereo DEM, snow depth, and lidar DSM/DTM products are available from the NSIDC SnowEx data archive (Hu et al., 2023b). The VHR land cover maps (Hu & Shean, 2023) used for co-registration and snow depth filtering are also available from NSIDC. The Mesa County lidar data (U.S. Geological Survey, 2017b) are available from the USGS 3D Elevation Program (3DEP) and the Delta County lidar data are available from the Colorado Hazard Mapping Lidar Portal (Colorado Hazard Mapping Portal, 2016). The ASO 3 m snow depth data (Painter et al., 2018) are available from NSIDC. The in situ SnowEx 2017 campaign data are available from NSIDC: snow probe data (Brucker et al., 2018), snow pit data (Elder et al., 2018), ultrasonic array data (Jennings et al., 2018), and GPR snow depth data (Webb et al., 2019).

3.10.2 Software

The NASA Ames Stereo Pipeline (ASP v3.2.0; Alexandrov et al., 2023) used for stereo processing is available on GitHub. We processed all airborne laser scanning data with the Point Data Abstraction Library (PDAL Contributors, 2022). The `demcoreg` (v1.1.0; D. Shean et al., 2021) package used for co-registration and the `get_modscag.py` utility (D. Shean, 2018) used to access and pre-process MODSCAG fSCA data are available on GitHub.

3.11 Author Contributions

Conceptualization, J.M.H. and D.S.; Methodology, J.M.H., S.B. and D.S.; Software, J.M.H.; Validation, J.M.H.; Formal Analysis, J.M.H.; Investigation, J.M.H.; Resources, D.S.; Data Curation, J.M.H. and D.S.; Writing—Original Draft Preparation, J.M.H.; Writing—Review and Editing, J.M.H., D.S., S.B.; Visualization, J.M.H.; Supervision, D.S.; Project Administration, D.S.; Funding Acquisition, D.S. All authors have read and agreed to the published version of the manuscript.

3.12 Acknowledgements

This research was funded by NASA awards 80NSSC18K1405 and 80NSSC22K0682 to the University of Washington. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. We sincerely thank Oleg Alexandrov for critical ASP software development and guidance in stereo processing. We thank Chris Hiemstra and Eli Deeb for assistance with tasking and delivery of VHR imagery. We thank the NASA SnowEx team and all volunteers who participated in the SnowEx 2017 field campaign. We thank two anonymous reviewers, Jessica Lundquist, and Ellyn Enderlin for providing helpful feedback on this manuscript. We acknowledge the Coast Salish and Ute peoples, upon whose traditional lands and territories the bulk of this work was conducted (Native Land Digital, 2022). The authors declare that there are no conflicts of interest, real or perceived.

3.13 APPENDIX B

This appendix includes figures and tables that supplement the main article. This includes satellite stereo image footprints, information regarding images and stereo pair geometry, stereo and co-registration precision, evaluation results, and documentation of artifacts in aerial lidar and satellite stereo snow depth products.

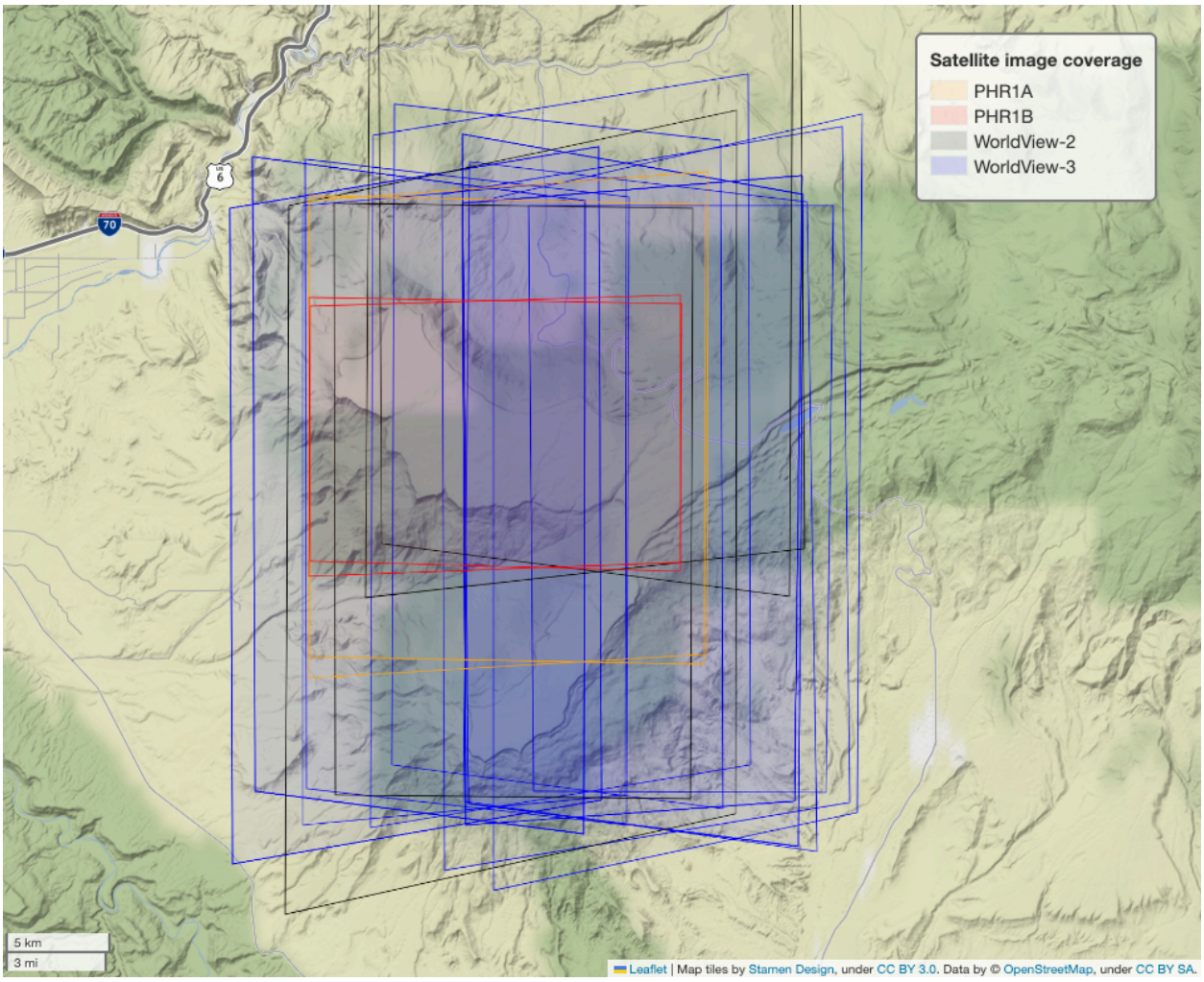

Figure B 1. Footprint coverage of the WorldView-2, WorldView-3, and Pléiades-HR (PHR1A, PHR1B) panchromatic stereo images used in this study.

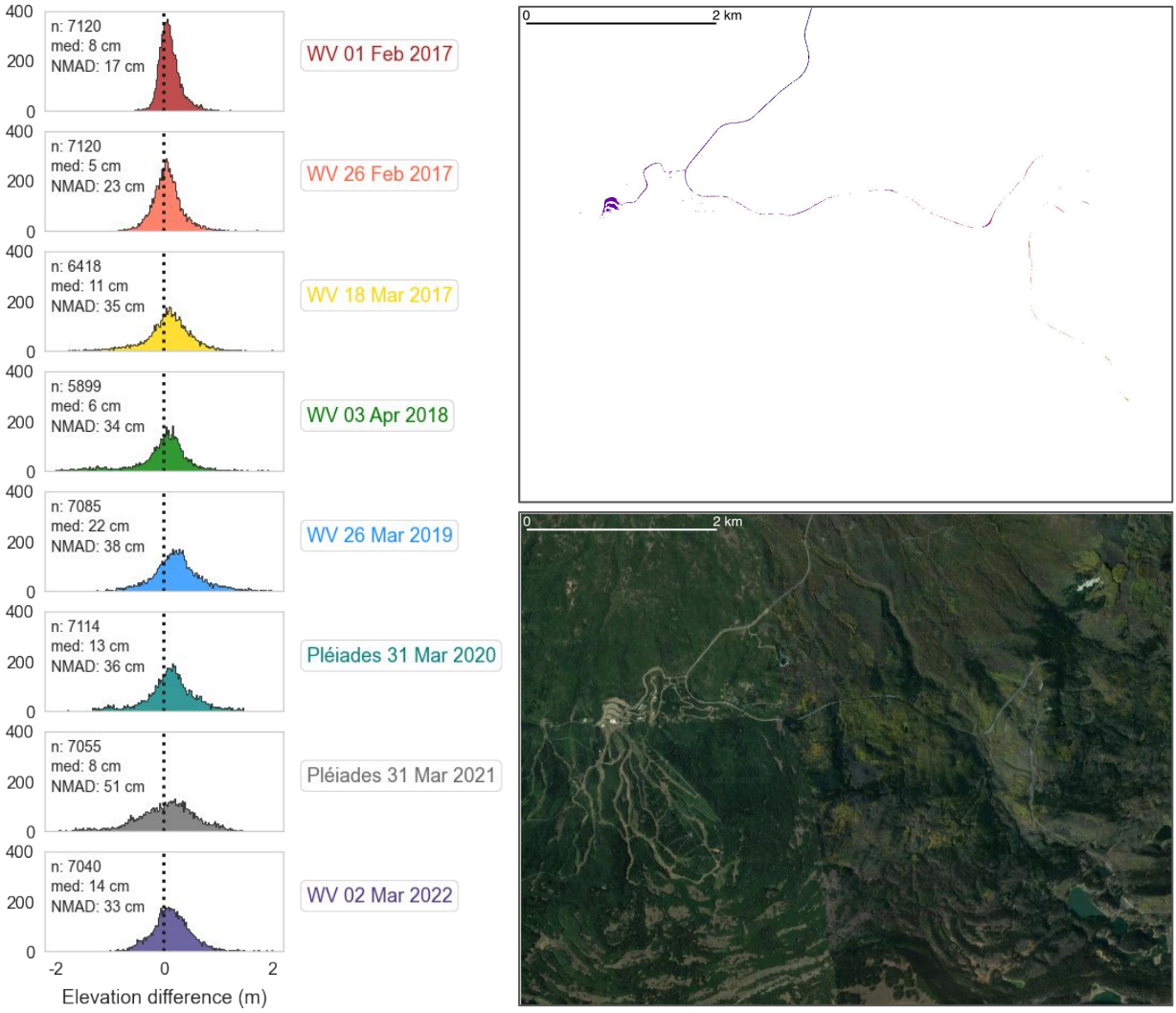

Figure B 2. Histograms of elevation difference values (snow-on stereo DEM minus snow-free 3DEP DTM) over stable surfaces (purple) within the Mesa Lakes SNOTEL evaluation area.

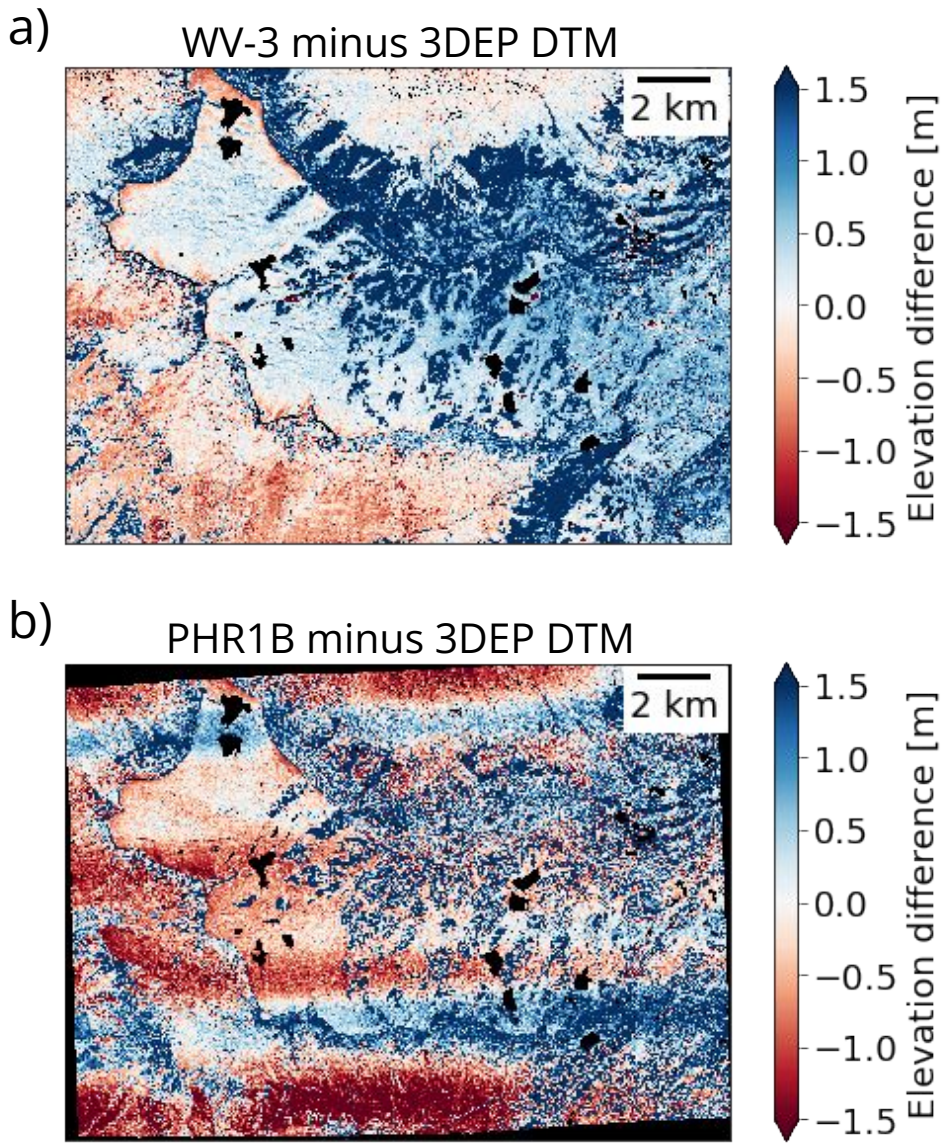

Figure B 3. Vertical elevation error due to unmodeled linescan sensor attitude variations (”jitter”) for WorldView-3 (top) and Pleiades-HR 1B (bottom). (a) 1 February 2017 WV-3 elevation difference map (snow-on DEM minus 3DEP snow-free DTM), and (b) 31 March 2021 Pléiades-HR 1B elevation difference map (snow-on DEM minus 3DEP snow-free DTM). Note large magnitude of these errors in PHR1B DEM, and variations with relief that preclude a simple ”cross-track median” correction.

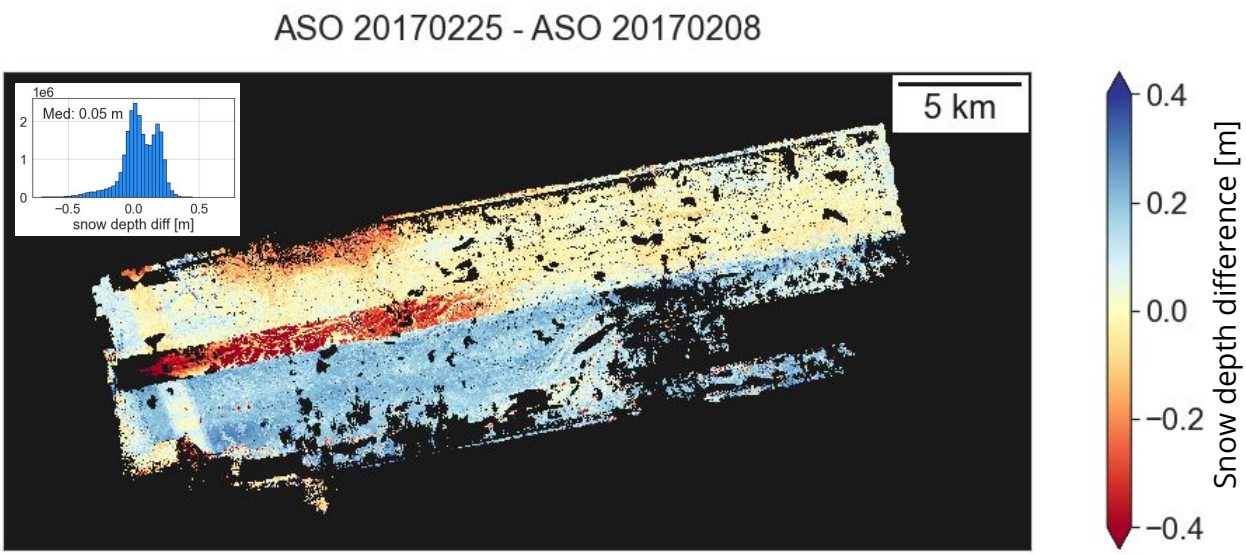

Figure B 4. Difference map of ASO snow depth products (ASO 25 February 2017 snow depth minus ASO 8 February 2017 snow depth) showing swath misalignment artifacts with WSW- ENE orientation. The RMS of difference values is 0.52 m, which is mostly error due to these artifacts, though we expect a real change (~0.10-0.15 m) in snow depth during this period based on SNOTEL time series (Figure 3d). These artifacts limit the WV stereo snow depth evaluation and interpretation of real snow depth change for Grand Mesa using ASO data during this period. It should be possible to improve the swath registration of the original ALS point cloud data (e.g., DeLong et al., 2022) and to release a new version of these ASO snow depth products for Grand Mesa.

Table B 1. Metadata for all VHR stereo image pairs used in this study, including ground sample distance (GSD) of the more nadir image in the pair, relevant acquisition geometry parameters (convergence angle [CA], base-to-height ratio [BH], asymmetry angle [AA], bisector elevation angle [BIE]), intersection area (Area), and pair acquisition type (in-track, "along" or cross-track, "cross").

Date	Catalog IDs	GSD [m]	CA [°]	BH	AA [°]	BIE [°]	Int Area [km^2]	Type
20170201	10400100276B9500 10400100286A3900	0.36	9.1	0.16	20.0	60.4	187	Cross
20170201	1040010026C28A00 1040010028192C00	0.37	9.3	0.16	20.0	60.0	164	Cross
20170201	1040010026C28A00 10400100286A3900	0.36	30.8	0.55	0.0	68.0	167	Cross
20170201	1040010026C28A00 10400100276B9500	0.37	39.9	0.73	4.5	66.7	456	Along
20170201	1040010028192C00 10400100286A3900	0.36	40.1	0.73	4.7	68.0	466	Along
20170201	10400100276B9500 1040010028192C00	0.40	49.2	0.91	0.2	68.0	195	Cross
20170226	1040010029504500 1040010029817600	0.38	58.6	1.12	4.2	72.8	188	Cross
20170226	1040010029504500 1040010029BF5800	0.33	28.2	0.5	11.1	70.1	127	Cross
20170226	1040010029817600 104001002A091600	0.35	15.4	0.27	25.8	59.8	182	Cross
20170226	1040010029BF5800 104001002A091600	0.33	15.0	0.26	10.6	70.5	124	Cross
20170226	1040010029504500 104001002A091600	0.35	43.2	0.79	3.5	72.5	439	Along
20170226	1040010029817600 1040010029BF5800	0.33	30.4	0.54	18.4	65.9	394	Along
20170318	104001002991D700 104001002BA5A500	0.38	35.1	0.63	7.2	61.3	493	Along
20180403	103001007AODBD00 103001007B395800	0.48	37.2	0.67	17.7	69.4	474	Along
20190326	1040010048434C00 104001004918EB00	0.37	38.5	0.7	5.5	65.2	484	Along
20200331	PHR1A_P_202003311814196 PHR1A_P_202003311815114	0.73	28.0	0.5	3.6	81.3	405	Along
20210331	PHR1B_P_202103311808189 PHR1B_P_202103311807290	0.73	27.0	0.48	3.4	84.2	217	Along
20220302	10300100CF123900 10300100CF2AAF00	0.56	36.3	0.66	4.1	65.3	1,206	Along
20220302	10300100CF696C00 10300100CF83A300	0.55	35.2	0.64	2.8	66.6	1,179	Along

Table B 2. Dates of in situ data collection compared to remote sensing collection for early and late February 2017 campaigns.

	WV 1 February 2017	ASO 8 February 2017	WV 26 February 2017	ASO 25 February 2017
snow pit	*7-8 February 2017*	*7-8 February 2017*	*24-25 February 2017*	*24-25 February 2017*
ultrasonic	*1 February 2017*	*8 February 2017*	*26 February 2017*	*25 February 2017*
transect probe	*7-8 February 2017*	*7-8 February 2017*	*25 February 2017*	*25 February 2017*
gpr	*8-10 February 2017*	*8-10 February 2017*	*23-25 February 2017*	*23-25 February 2017*

Chapter 4. Evaluating Satellite Stereo Snow Depth Mapping of a Shallow Subarctic Boreal Forest Snowpack in Fairbanks, AK

4.1 Abstract

Satellite stereo photogrammetry is a promising technique for seasonal snow depth observation. However, few studies have evaluated the limitations of such an approach over a shallow snowpack. We produced and assessed 36 satellite stereo snow depth maps in interior Alaska's boreal forest from WY 2017–2022. Comparisons with airborne lidar snow depth and SNOTEL snow depth records show underestimates of open area snow depth by ~0.2–0.3 m and differences in retrieved distributions. Underestimates are in part attributed to challenges involved in co-registration and underlying snow-free reference datasets. Our evaluation of cross-track stereo pairs shows varied quality, but the highest precision DSMs and snow depth maps resulted from pairs with > 25˚ convergence angles and < 80˚ bisector elevation angles. Our preliminary assessment of Pléiades Neo products revealed significant processing artifacts, impeding the retrieval of snow depth. ArcticDEM DSM products showed promise but require further evaluation. Despite challenges with an unstable ground surface, our findings of consistent precision and resolved details highlight the potential of extracting valuable information from both in-track and cross-track stereo products in shallow snowpack environments.

4.2 Introduction

Boreal forest snow cover is highly reflective and an effective insulator, providing a critical control on thermal surface exchange between the ground and atmosphere (Moore, 1987). Up to 80% of boreal forests may be underlain by permafrost (Carpino et al., 2018), which serves as an important pool for global carbon stocks (Tarnocai et al., 2009). As deeper winter snow insulates the ground from cooler air temperatures, it modulates permafrost thaw and active layer depth, affecting the rate and timing of greenhouse gas release into the atmosphere (Jan & Painter, 2020; Natali et al., 2011; Sannel et al., 2016). The sparse nature of in situ monitoring networks and available data in the subarctic (Carpino et al., 2018) has made remote sensing a necessity to observe physical processes in these vast, remote spaces. Satellite sensors such as MODIS and

Landsat have been indispensable for earth observation, but finer spatial resolution is necessary to directly observe small-scale heterogeneity in snow accumulation, redistribution, and ablation processes.

Very-high-resolution (VHR) satellite stereo photogrammetry has been used to quantify surface elevation and meter-scale change signals for forest canopies (e.g., Lagomasino et al., 2015; Montesano et al., 2017; Persson & Perko, 2016), to map bathymetry (Hodúl et al., 2018), for building height estimation (e.g., Krauß et al., 2019), and to estimate mass balance and flow velocity of glaciers (e.g., Bhushan et al., in review). Seasonal snow studies have assessed VHR satellite stereo digital surface models (DSMs) in the retrieval of late winter, high-elevation alpine snowpack depth, with average depths > 0.8 m (Deschamps-Berger et al., 2020; Eberhard et al., 2021; Hu et al., 2023a; Shaw, Deschamps-Berger, et al., 2020). Though seasonal snow covers 39% of the terrestrial globe and two-thirds of these regions have average winter depths exceeding 0.5 m (Sturm & Liston, 2021), limited work has been completed to assess the limits of elevation change detection over shallow snowpacks (0.5 m or less) common to boreal forest regions.

Maxar possesses an extensive archive of monoscopic (single capture) WorldView-1, WorldView-2, and WorldView-3 images over Fairbanks, Alaska. This dataset enables the generation of cross-track stereo pairs (images collected from different orbits) with diverse stereo geometries, sensor combinations, and temporal offsets. While in-track stereo images are typically acquired within minutes of one another, cross-track stereo pairs comprised of monoscopic acquisitions (hereafter referred to as cross-track stereo pairs) can have large temporal offsets, resulting in dissimilar ground appearances and reduced correlation success needed in elevation reconstruction. Moreover, unlike in-track stereo pairs – which have pre-defined stereo acquisition geometries – cross-track stereo pairs can have variable stereo viewing geometries. Recent studies over vegetated, snow-free terrain have assiduously studied the theory and simulated impacts of varying viewing geometries, illumination conditions, and ground sample distance (GSD) on vertical and horizontal geolocation error of stereo DSMs (Jeong & Kim, 2016; Yin et al., 2023). However, no studies have yet assessed viewing geometry impacts on accuracy for cross-track stereo DSMs of snow-covered scenes.

In this work, our primary objective was to evaluate the feasibility of satellite stereo snow depth retrieval in a subarctic boreal forest environment. Following the methodology established in Hu et al. (2023a), we examined satellite stereo surface elevation and snow depth products for a shallow snowpack in Fairbanks, Alaska. We assessed the accuracy and precision of in-track satellite stereo surface elevation and snow depth using near-contemporaneous aerial and ground-based validation datasets collected as a part of the SnowEx 2023 campaign. Additionally, we evaluated the potential of new VHR satellite stereo constellations (i.e., Pléiades Neo) and publicly available VHR stereo DEMs (i.e., ArcticDEM) for measuring snow depth. Finally, we investigated the potential of cross-track satellite stereo pairs formed from monoscopic imagery for elevation reconstruction and snow depth retrieval.

4.3 Study Site

Our study area is centered on Creamer's Field Migratory Waterfowl Refuge, near Fairbanks, in interior Alaska (Figure 4.1) and was one of the main boreal forest (taiga) sites for the SnowEx 2023 campaign. Creamer's Field consists of bog flats, with underlying near-surface permafrost covering the majority of the refuge (Douglas et al., 2021). Elevation within this site ranges from ~140 to 347 m above the WGS84 Ellipsoid. Vegetation at the site is made up of shrubs (38.9%), coniferous forest (35%), fields (15.4%), deciduous forest (6.3%), and herbaceous bog (2.4%) (*Creamer's Field Migratory Waterfowl Refuge Resource Inventory*, n.d.). For this study, we focus on three evaluation areas to characterize snow depth: woody wetlands, Creamer's Field (the field closest to Creamer's dairy building) and a field to the northeast of the preceding areas. The woody wetlands evaluation area is defined by the National Land Cover Database classification and contains areas with > 20% forest or shrubland vegetation, including a mix of the aforementioned vegetation composition, and where the ground "is periodically saturated with or covered with water" (Dewitz, 2021) .

Two SNOTEL stations are located near or within the study area: Fairbanks F.O. (1174) is located southwest of the study area, and the Creamer's Field (1302) is located within the study area. Both sites are in semi-forested areas, with elevations of 144 m and 145 m, respectively (heights above WGS84 Ellipsoid). While mean annual snowfall for the area is ~1.6 m (Arguez et al., 2012), median peak SNOTEL snow depth between WY 2012–WY 2023 is 0.64 m.

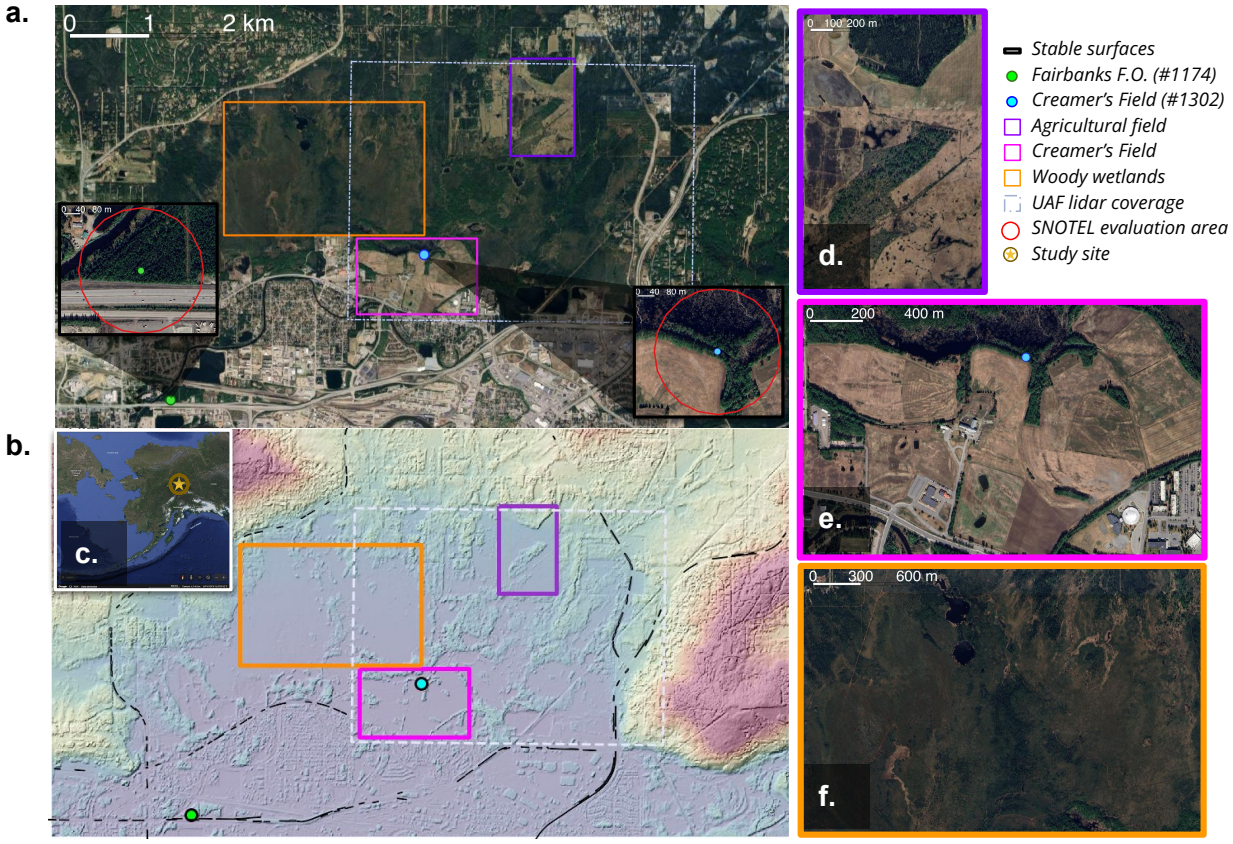

Figure 4.1 Study site in Fairbanks, Alaska. (a) True color Google Basemap image of study area showing evaluation areas and context for SNOTEL sites. (b) WorldView composite DSM elevation map of site showing stable surfaces. (c) Context map. (d) Northern field evaluation area. (e) Creamer's field evaluation area. (f) Woody wetland evaluation area.

4.4 DATA

4.4.1 Commercial very-high-resolution images and data products

4.4.1.1 Satellite stereo images

We used twelve L1B (Level 1B) images from six archived in-track stereo pairs collected by WorldView-1 (WV-1, one pair), WorldView-2 (WV-2, two pairs), and WorldView-3 (WV-3, three pairs) between March 2018 and October 2022 (Table C. 1). Two collections coincided with spatially extensive datasets from the 2023 NASA SnowEx airborne and ground campaign.

Candidate cross-track stereo pairs had convergence angles between 6˚–60˚, a maximum of four days temporal offset between mono image collections, and less than or equal to 30% cloud

cover (e.g., D. Shean, 2017). Between March 2017 and March 2023, we prepared 13 cross-track, same-sensor stereopairs collected by WV-1 (two pairs), WV-2 (nine pairs), and WV-3 (two pairs), and 17 cross-track, cross-sensor pairings (WV-2 and WV-3). Though images collected in March 2023 coincided with the SnowEx campaign, validation data for this time period has not yet been released.

While the original footprints of collections covered 172 to 630 km^2, the area centered on the study site and evaluated in this work varied between 44 to 58 km^2. One in-track stereo pair, collected by Pléiades Neo (0.30 m GSD) on 13 March 2023, covered 28 km^2.

4.4.1.2 ArcticDEM products

The University of Minnesota's Polar Geospatial Center provides WV-1, WV-2, and WV-3 (and select GeoEye) stereo DSM strip and mosaic products at 2 m posting for the Arctic (ArcticDEM) between 2007 and 2022. All ArcticDEM products are processed using the Surface Extraction from TIN-based Searchspace Minimization (SETSM) production workflow (Noh & Howat, 2015), which refines the rational polynomial coefficient (RPC) sensor model to relate image coordinates with ground coordinates, rather than the available rigorous linescan sensor model. The s2s041 version of the strip DSM data (Porter et al., 2022) is used in this work.

By comparing our in-track stereo DSMs with the corresponding in-track ArcticDEM strip DSMs prepared from the same Level-1B images, we directly evaluate the impacts of stereo and post-processing workflows on DSM quality and accuracy. At the time of writing, no ArcticDEM snow-on products were available after 28 October 2021 over our study site, so we were unable to evaluate DEMs constructed from the 2022 in-track WV-3 stereopairs. Comparisons were completed using WV-1 and WV-2 DSMs from four snow-on dates (22 March 2018, 28 February 2019, 15 March 2021, and 18 April 2021, noted with asterisks in (Table C. 1).

4.4.2 Reference datasets

The USGS 3D Elevation Program (3DEP) acquired airborne lidar data over Fairbanks between 13 June – 1 July 2017. We processed the 3DEP lidar point cloud data following Hu, et al. (2023), reprojecting point clouds to a UTM 6N coordinate system (EPSG:32606) with heights

above the WGS84 ellipsoid to produce a digital surface model (DSM) and digital terrain model (DTM) for the study area. We then interpolated values to fill data gaps in resulting DEMs using the `gdal_fillnodata.py` utility with a distance of 100 pixels, a 3x3 averaging filter, and three smoothing iterations.

Snow depth maps, DTMs and canopy height models (CHMs) of smaller extent were also produced from UAF lidar data (GSD 0.5 m) collected on 24 October 2022 and 11 March 2022 for the SnowEx 2023 campaign. In spite of a dusting of light snow on the ground < 0.1 m during the 24 October 2022 flight, we combined the CHM and DTM products to construct a DSM to directly compare with a stereo DSM from 28 October 2022. We also used daily records of snow depth from the Fairbanks F.O. (1 Jan 2011–21 July 2020) and Creamer's Field SNOTEL sites (13 August 2021–present).

4.4.3 Land cover datasets

By subtracting the 3DEP DTM from the DSM, we constructed a CHM at 1 m resolution and used this product to mask vegetation. Linear swath artifacts of magnitudes between 0.05 and 0.1 m were observed in this CHM product. We used the 2016 NLCD Alaska (30 m) land cover dataset as a coarser mask for vegetation (classes 41–43) and developed classes (classes 21–24) to identify and mask non-snow pixels.

4.5 METHODS

4.5.1 Stereo processing

Stereo snow depth processing closely mirrors that of Hu, et al. (2023), but includes WorldView-1 images, cross-track stereo pairs generated from mono images, and posting stereo DSMs at a finer resolution (1 m). The WorldView L1B images are mosaics generated from several adjacent CCD-acquired Level 0 products (Shean et al., 2016) and can contain residual artifacts (i.e., sub-pixel offsets) visible as vertical stripes, arising from CCD alignment error. Though these artifacts are less pronounced in WV-3 images, they significantly affect WV-1 and WV-2 images and their derivatives. Maxar now corrects sub-pixel offsets due to CCD geometry alignment error in WV-2 products (for orders placed after June 2022), but these artifacts remain

uncorrected in WV-1 images. We used the ASP utility, `wv_correct`, to apply empirical CCD geometry corrections to all WV-1 images during pre-processing.

4.5.2 Co-registration

We used the `dem_align.py` utility to perform two-stage co-registration, similar to that of Hu et al. (2023a). The first stage employs iterative alignment using all surfaces in the reference 3DEP DSM and stereo DSM (including snow and vegetation) to solve for the horizontal offset that minimizes elevation differences over the full distribution of terrain slope and aspect values. The second stage removes the median vertical offset between the stereo DSM and the reference DTM over a limited subset of stable surfaces (i.e., plowed roads).

4.5.3 Snow depth derivation

Co-registered elevation difference maps were processed to filter out vegetation using a binary mask (pixels with values >= 0.6 m are masked) derived from the 3DEP CHM. Elevation difference values beyond -0.3–10 m range were also omitted from the snow depth map. Filtered maps were then smoothed using a Gaussian kernel (kernel = 3x3, sigma = 1) and once more filtered with the binarized vegetation mask.

4.5.4 Evaluation of DEM and snow depth rasters

We evaluated ASP-generated and co-registered WorldView and Pléiades Neo DEMs by computing elevation difference maps using a reference 3DEP DTM. To assess ArcticDEM DSMs, we computed elevation difference maps using corresponding ASP-generated DSMs. For all comparisons, we computed the median (*med*) and normalized median absolute deviation (NMAD, Equation 4.1) for elevation residuals over stable and snow-covered surfaces, each represented by (Δx_i), for each observation *i*, of *N* total observations. We used NMAD in elevation residuals over stable surfaces as a proxy of snow depth measurement precision. Using UAF lidar snow depth from 11 March 2022 as a reference, we computed snow depth residuals by subtracting the reference snow depth (SD_i) from measured snow depth ($\widehat{SD_i}$) for each observation *i*. Along with NMAD for each snow depth dataset, we report the median (i.e., offset) and root mean squared error (RMSE, Equation 4.2) of observed snow depth residuals.

$$NMAD = 1.4826 \times med(|\Delta x_i - med_{\Delta x}|) \quad (4.1)$$

$$RMSE = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} \widehat{SD}_i - SD_i\right)^2} \quad (4.2)$$

To assess spatial autocorrelation in snow depth, we computed a semivariogram for the 11 March 2022 UAF lidar snow depth map and identified a reference length scale of 200 m for the study site. Using a radius of this reference length to enclose a circle around each SNOTEL site, we calculated the median and 16th to 84th percentile spread of stereo snow depth in this area and compared with daily SNOTEL records (Figure 4.1 Study site in Fairbanks, Alaska. (a) True color Google Basemap image of study area showing evaluation areas and context for SNOTEL sites. (b) WorldView composite DSM elevation map of site showing stable surfaces. (c) Context map. (d) Northern field evaluation area. (e) Creamer's field evaluation area. (f) Woody wetland evaluation area.).

4.5.5 Stereo geometry theory

In satellite stereo photogrammetry, the position and orientation of the satellite sensors relative to the area of interest and the corresponding stereo image that comprise the stereo pair strongly influences vertical and horizontal error in resulting DEMs. We focus on three stereo angles that are used to describe viewing geometry: convergence angle (CA), bisector elevation angle (BIE), and asymmetry angle (AA). The convergence angle is formed from the left and right rays and is expected to have the largest impact on vertical accuracy – small convergence angles can result in large vertical error (e.g., Jeong & Kim, 2014). The bisector elevation angle is formed between the bisector and the ground plane and indicates the degree of obliquity of acquired images – small BIE (further from and less than 90˚) will increase horizontal error, and potentially decrease vertical error (Jeong & Kim, 2014; Figure 4.2). The asymmetry angle is a measure of scale difference between each stereo image – large AA (further from and larger than 0˚) will increase horizontal error but may decrease vertical error (Figure 4.2). We calculated these angles for each stereo pair (both in-track and cross-track pair types) to evaluate optimal viewing geometries for stereo elevation reconstruction over snow at our study site.

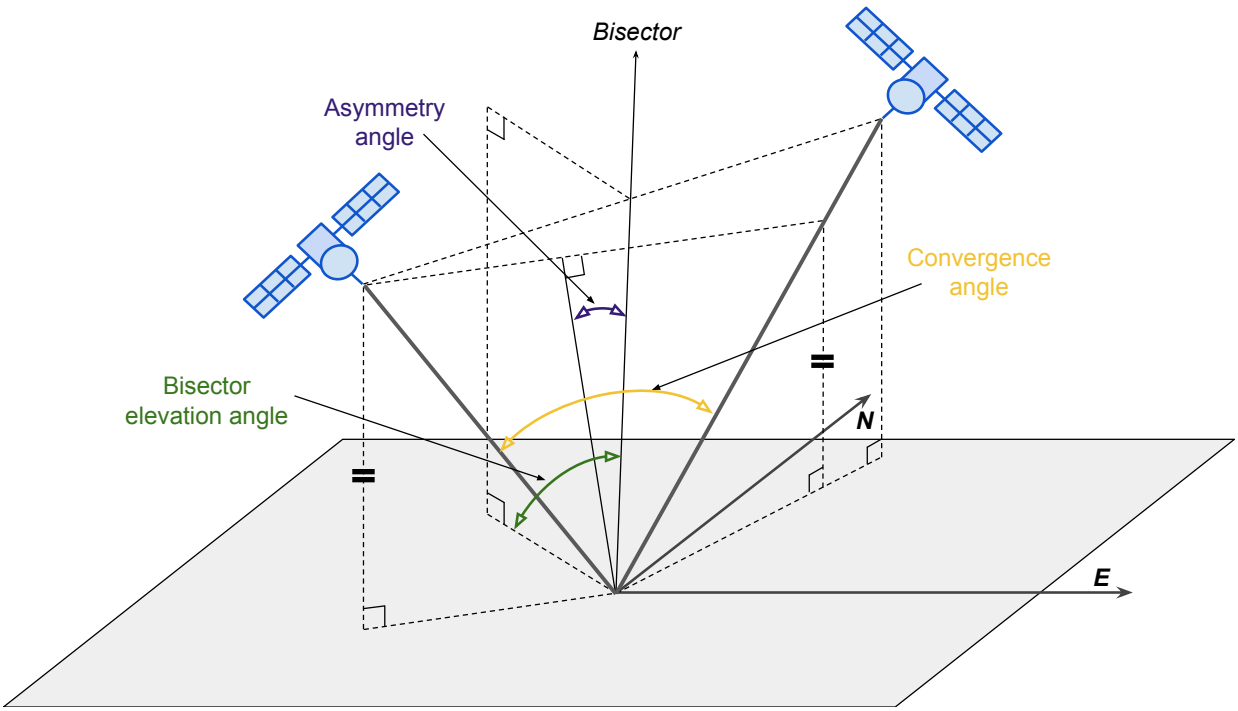

Figure 4.2 Stereo geometry diagram illustrating convergence angle, bisector elevation angle, and asymmetry angle (e.g., Jeong & Kim, 2016).

4.6 RESULTS

4.6.1 Evaluation of DEMs

4.6.1.1 Comparison of WorldView ASP DSM with UAF lidar DSM

For the October 2022 comparison (28 Oct stereo vs. 24 Oct UAF lidar), a small dusting of snow < 0.1 m is visible in orthoimages. Since expected measurement uncertainty is 0.2–0.3 m, we evaluated these DSMs directly rather than the snow depth products, which can integrate additional error from processing (Figure 4.3a). Stereo elevation values are similar to UAF lidar values, with a median offset of -0.12 m (stereo minus lidar) that is consistent over open fields and woody wetlands alike (Figure 4.3b). Elevation difference values have the largest NMAD over woody wetlands where there is more vegetation of variable height (woody wetlands: 0.37 m vs. open fields: 0.13 to 0.18 m). Elevation differences between the stereo DSM and UAF DSM over stable surfaces have a right-tailed distribution, with NMAD of 0.31 m.

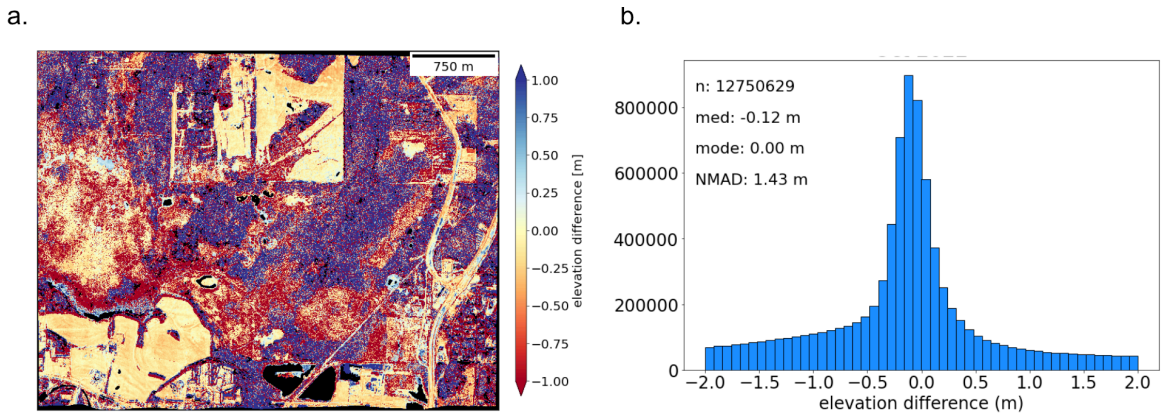

Figure 4.3 (a) Snow-free DSM elevation difference map between co-registered WV-3 stereo and UAF lidar DSMs (28 Oct 2022 stereo minus 24 Oct 2022 lidar) and (b) corresponding histogram.

4.6.1.2 Comparison of ASP DSMs with ArcticDEM DSMs

The snow-on ArcticDEM DSMs were generally smoother than ASP stereo DSMs, with offsets of -0.03 to 0.41 m (ASP minus ArcticDEM). Though measurement precision over stable surfaces were similar (ASP NMAD: 0.36–0.44 m vs. ArcticDEM NMAD: 0.28–0.43 m), maps of elevation difference (ASP – ArcticDEM) showed longwave vertical undulations and compounding of residual CCD artifacts (~0.3 m; Figure 4.4a).

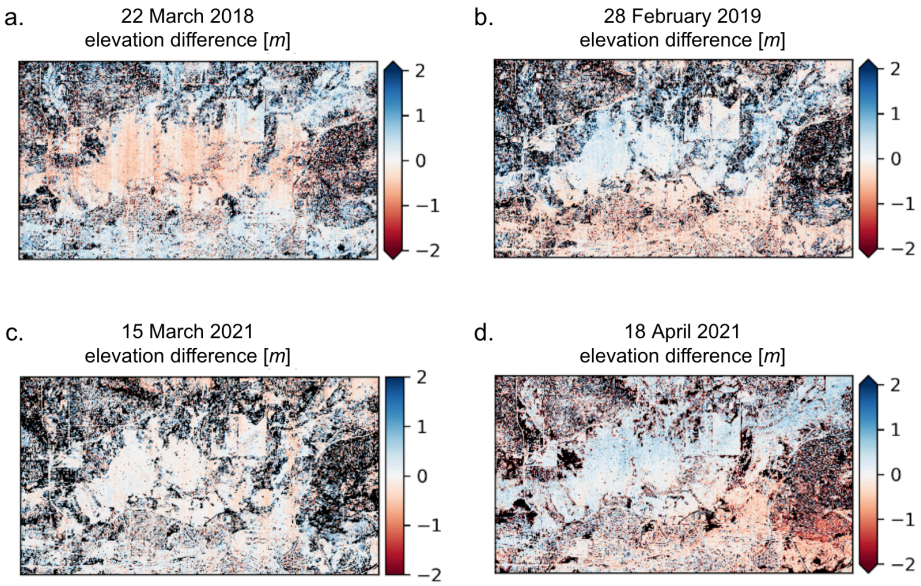

Figure 4.4 Elevation difference of co-registered ArcticDEM DSM with ASP DSM (ASP – ArcticDEM) for (a) 22 March 2018, (b) 28 February 2019, (c) 15 March 2021, and (d) 18 April 2021.

4.6.1.3 Evaluation of Pléiades Neo DSM

A near-planar offset, increasing from the southern edge of the image (~ -1 m) to the northern edge (~2 m) is visible in elevation difference products using the Pléiades Neo DEM from 13 March 2023 (Figure C. 4a). Repeating striations with wavelength 150–200 m and amplitude 0.5 m are also visible (Figure C. 4b). To correct for these artifacts, we explored applying a polynomial surface fit, followed by a counterclockwise rotation (-0.77˚) and a row-wise median subtraction to remove high-frequency striations (Figure C. 4c). Although corrections reduced the tilt and high frequency artifact magnitudes, remnant artifacts prevented reliable recovery of snow depth signals (Figure C. 4d).

4.6.2 Evaluation of in-track stereo snow depth

4.6.2.1 Interannual evaluation of ASP snow depth

Over the study site, median stereo snow depth ranged from a low of 0.48 m to a high of 0.70 m (Figure 4.5a, Table C. 2), with snow depth NMAD ranging between 0.26–0.46 m for the six snow-on in-track stereopairs. Snow depth was consistently higher over woody wetlands by ~0.2 m (median 0.54–0.86 m; Figure 4.5b), compared to over open fields (median: 0.17–0.69 m; Figure 4.5c and d), but with similar NMAD (0.15–0.29 m) across all three evaluation areas. Elevation difference values over stable surfaces (Figure 4.1b) had NMAD ranging between 0.33–0.44 m. See Table C. 2 for individual metrics.

Small magnitude CCD artifacts (vertical striping; 0.2–0.4 m) are visible in elevation difference products from the 28 Feb 2019 WV-2 product (Figure 4.5c, Figure C. 1). Longwave undulations are visible in the 18 April 2021 WV-1 product (Figure C. 1b) of ~0.2 m amplitude and wavelength of 1.5–1.8 km. These artifacts suggest observed differences between ArcticDEM and ASP DSMs are in part stemming from residual CCD misalignment in WV-2 (Maxar-corrected) products and unmodeled attitude error in WV-1 products (ASP DSMs).

4.6.2.2 Comparison with SNOTEL snow depth records

Four of the six in-track stereo snow depth collections have corresponding SNOTEL records across the Fairbanks F.O. and Creamer’s Field SNOTEL sites. Over these four dates, SNOTEL

sites recorded 0.38–0.76 m of snow. Median stereo snow depth within a 200 m radius of SNOTEL sites (Figure 4.5e) generally underestimated daily SNOTEL records by 0.27–0.37 m, except for 28 Feb 2019 where median snow depth was 0.25 m (65%) higher than the recorded SNOTEL snow depth (WV-2 snow depth 0.63 m vs. SNOTEL snow depth 0.38 m). Though SNOTEL records were unavailable for WY 2021, stereo snow depth measurements within 200 m of the Creamer's Field SNOTEL site for both 2021 dates had similar spread and general magnitude of snow depth values as the following year (Figure 4.5e).

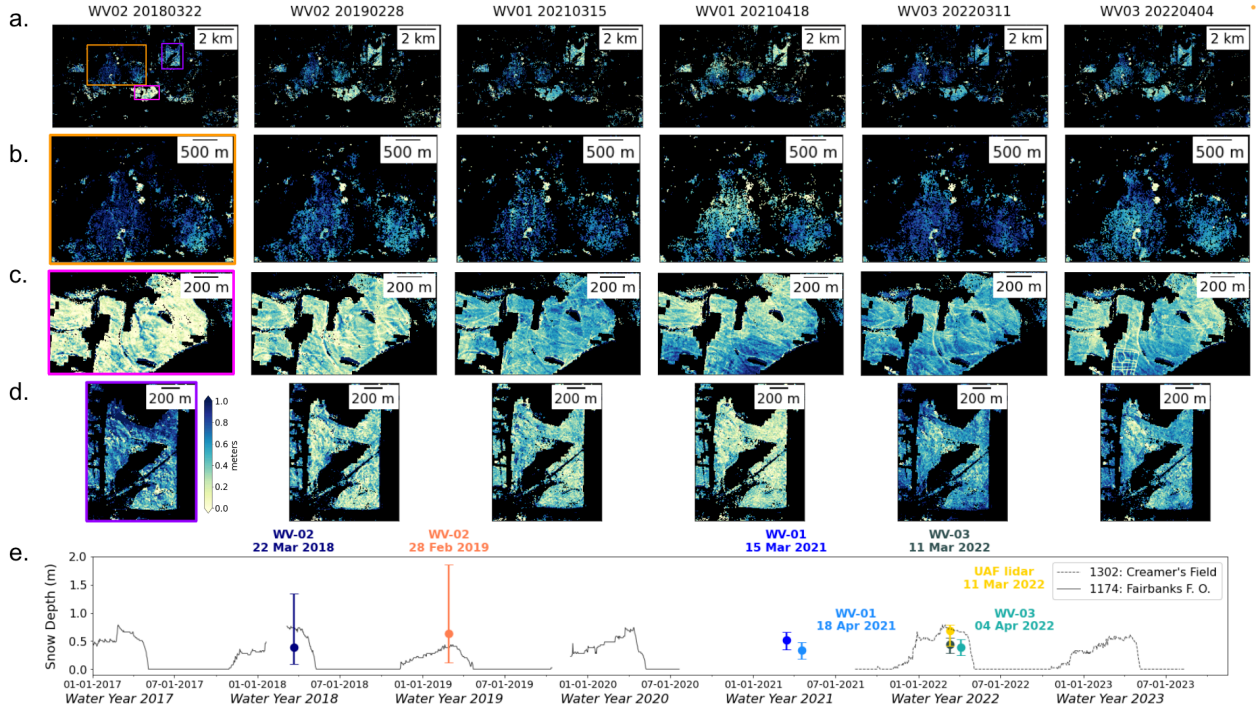

Figure 4.5 In-track stereo snow depth over (a) Creamer's Field/Farmer's Loop study site in Fairbanks, AK. (b) Stereo snow depth over woody wetlands (orange box); (c) snow depth over Creamer's Field proper (pink box), and (d) snow depth over an agricultural field (purple box). (e) Time series of daily SNOTEL snow depth, with stereo and UAF lidar snow depth within 200 m of the respective SNOTEL sites. Circles represent distribution median values and bars show 16–84% spread values.

4.6.2.3 Comparison with SnowEx UAF lidar snow depth

Snow depth maps from UAF lidar measurements on 11 March 2022 showed a higher and more uniform distribution (median 0.82 m) than WorldView-3 stereo measurements (median 0.65 m). Using UAF lidar as a reference, differences between distributions amounted to an offset

of -0.17 m and RMSE of 0.51 m (Figure 4.6). When coarsening both datasets, offsets were similar (-0.15 to -0.17 m), but RMSE decreased from 0.50 m at 1 m posting (i.e., grid cell size) to 0.45 m (10 m posting), to 0.36 m (30 m posting), and finally, RMSE of 0.26 m at 100 m posting. Stereo snow depth also had a larger NMAD over snow than UAF lidar (0.24 m vs. 0.07 m, respectively).

Differences in snow depths between datasets were smallest over the woody wetlands (offset -0.04 m, RMSE 0.19 m; Figure 4.6d), where vegetation is ~0.2–1.6 m high, and largest over open fields (offset -0.24 to 0.25 m, RMSE 0.41 m and 0.53 m; Figure 4.6e and f), where vegetation is typically < 0.1 m. Elevation differences over stable surfaces for stereo products had an NMAD of 0.43 m (Figure C. 2). Compared to SNOTEL records, median stereo snow depth within 200 m of the site was much lower than UAF lidar (-0.30 m vs. -0.06 m, respectively).

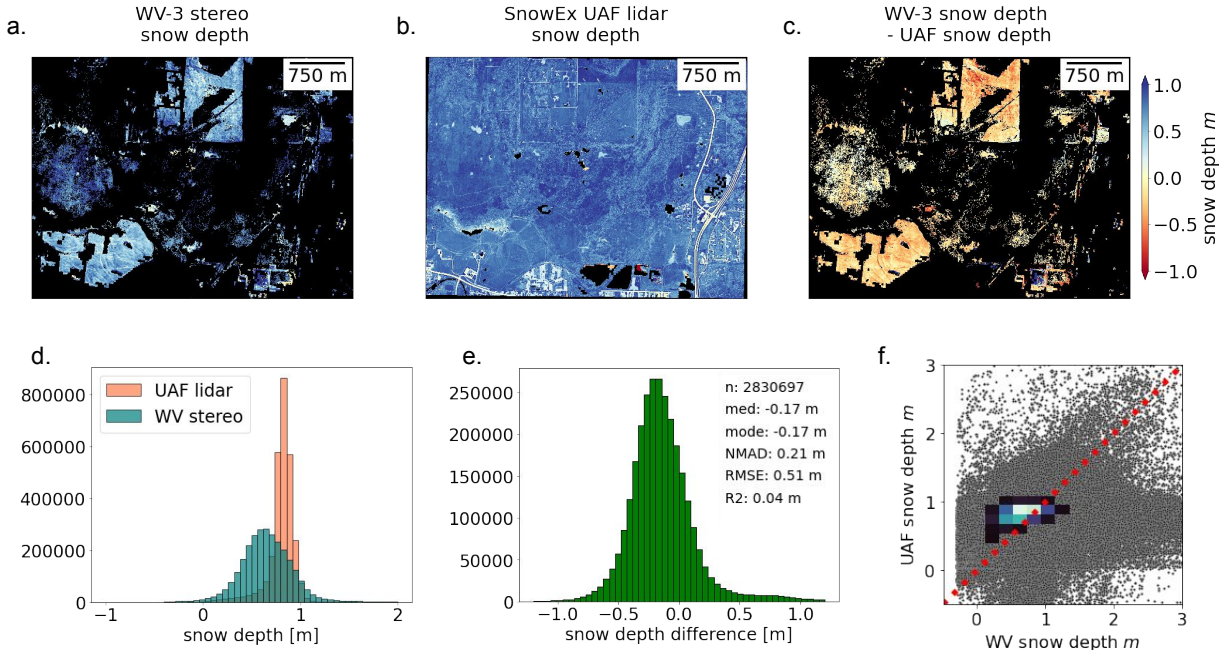

Figure 4.6 Snow depth maps of (a) SnowEx UAF lidar, (b) WV-3 stereo, (c) WV-3 minus UAF snow depth difference for 11 March 2022. Histograms of (d) snow depth distributions for pixels common to both snow depth products and (e) per-pixel snow depth differences (WV-3 minus UAF). (f) Scatterplot of WV-3 stereo snow depth against UAF snow depth

4.6.3 Evaluation of cross-track stereo snow depth

Snow depth retrieval from cross-track stereo pairs was highly variable. Some pairings produced noisy, speckled maps, while others appeared to reconstruct surfaces with more fidelity (Figure 4.7a). In general, we observed worse measurement precision over stable surfaces for cross-track stereo products (NMAD; WY 2017: 0.32–0.51 m; WY 2020: 0.45–0.89 m; WY 2023: 0.40–2.73 m) than in-track stereo products. Compared to SNOTEL records, cross-track snow depth had widely-ranging median offsets (-0.21 to 1.93 m), large NMAD values (0.71 m to 2.99 m) and meter-scale RMSE (1.43–3.75 m). See Figure 4.7 or Table C. 2 for more detail.

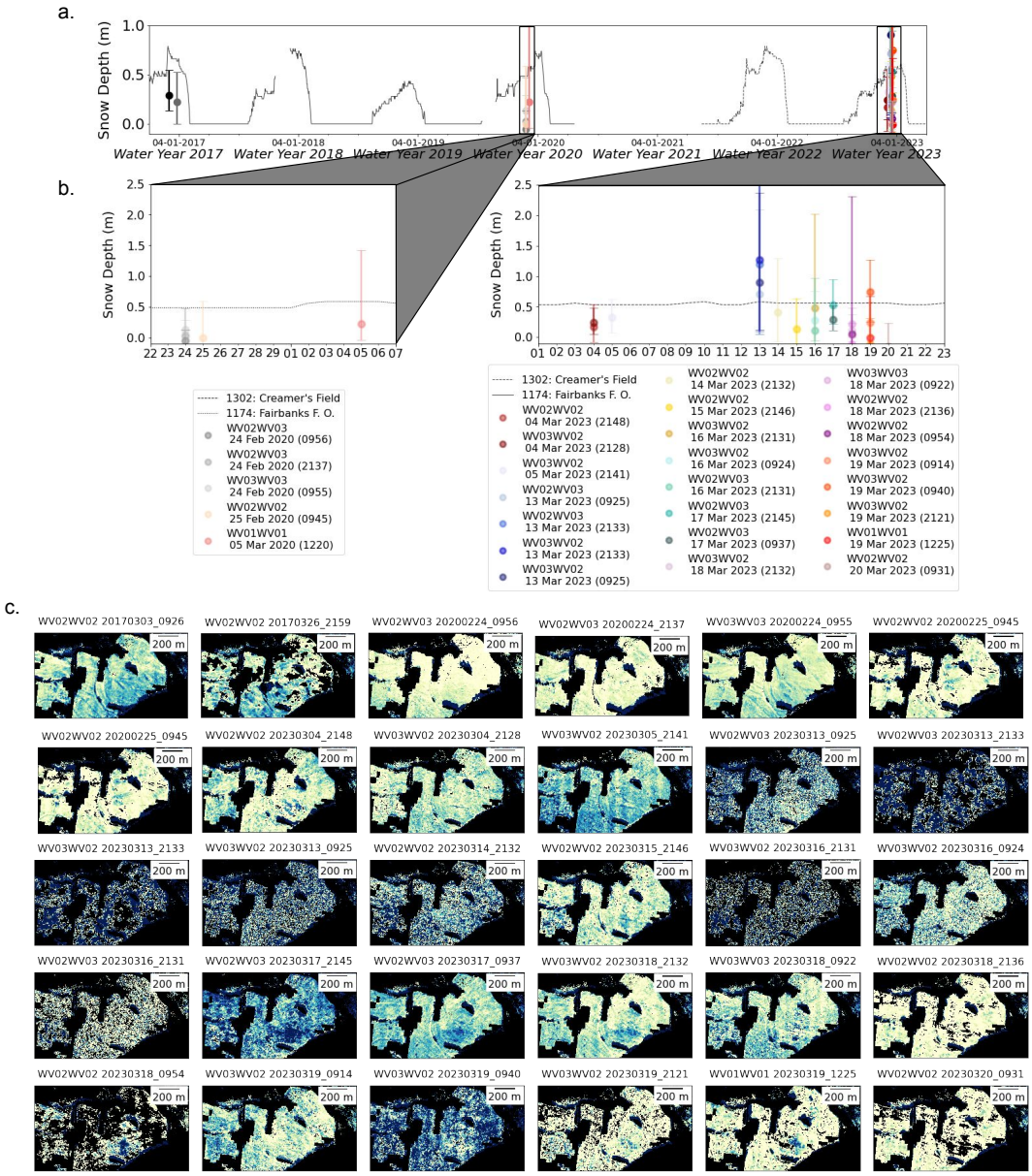

Figure 4.7 (a) Time series of daily SNOTEL snow depth and cross-track stereo snow depth measurements from SNOTEL evaluation area for WY 2017, WY 2020, and WY 2023. Circles represent distribution median values and bars show 16–84% spread values. (b) insets of time series for values from 22 Feb–7 March 2020 and March 2023. (c) Creamer's Field snow depth maps for all processed cross-track stereo pairs. All titles indicate the sensor name for both images and a timestamp (YYYYMMDD HHMM UTC).

For all evaluated metrics, in-track stereo pairs produced DSMs and snow depth maps with lower NMAD and a smaller range of snow depth differences from SNOTEL records (Figure 4.8), a trend that also held over each evaluation area.

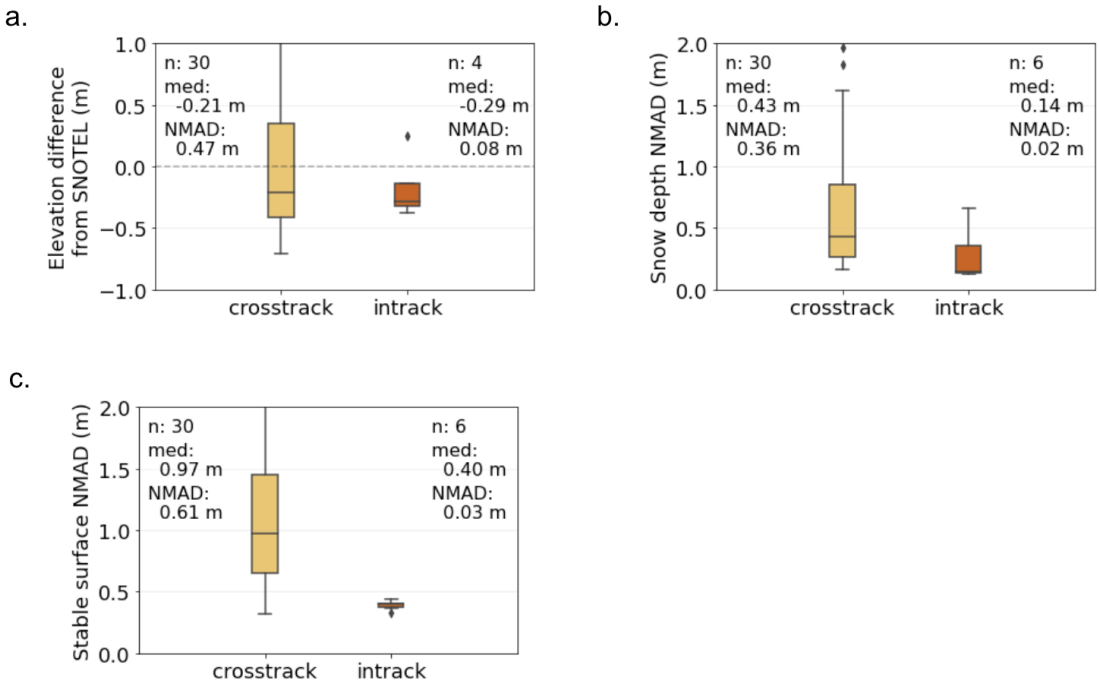

Figure 4.8 Cross-track and in-track stereo pair values for: (a) median snow depth difference from SNOTEL record for values within a 200 m radius of SNOTEL sites, (b) spread in snow depth NMAD values for the entire study site, and (c) spread of elevation difference values over stable surfaces.

4.6.4 Evaluation of stereo geometry

When examining all stereo pair viewing geometries evaluated in this study, worse precision was generally correlated with narrower convergence angles (< 20˚; Figure 4.9a) and with more weakly correlated with larger bisector elevation angles (> 80˚; Figure 4.9b), though many exceptions to both exist. Asymmetry angles did not appear to be strongly correlated with measurement precision (Figure 4.9c).

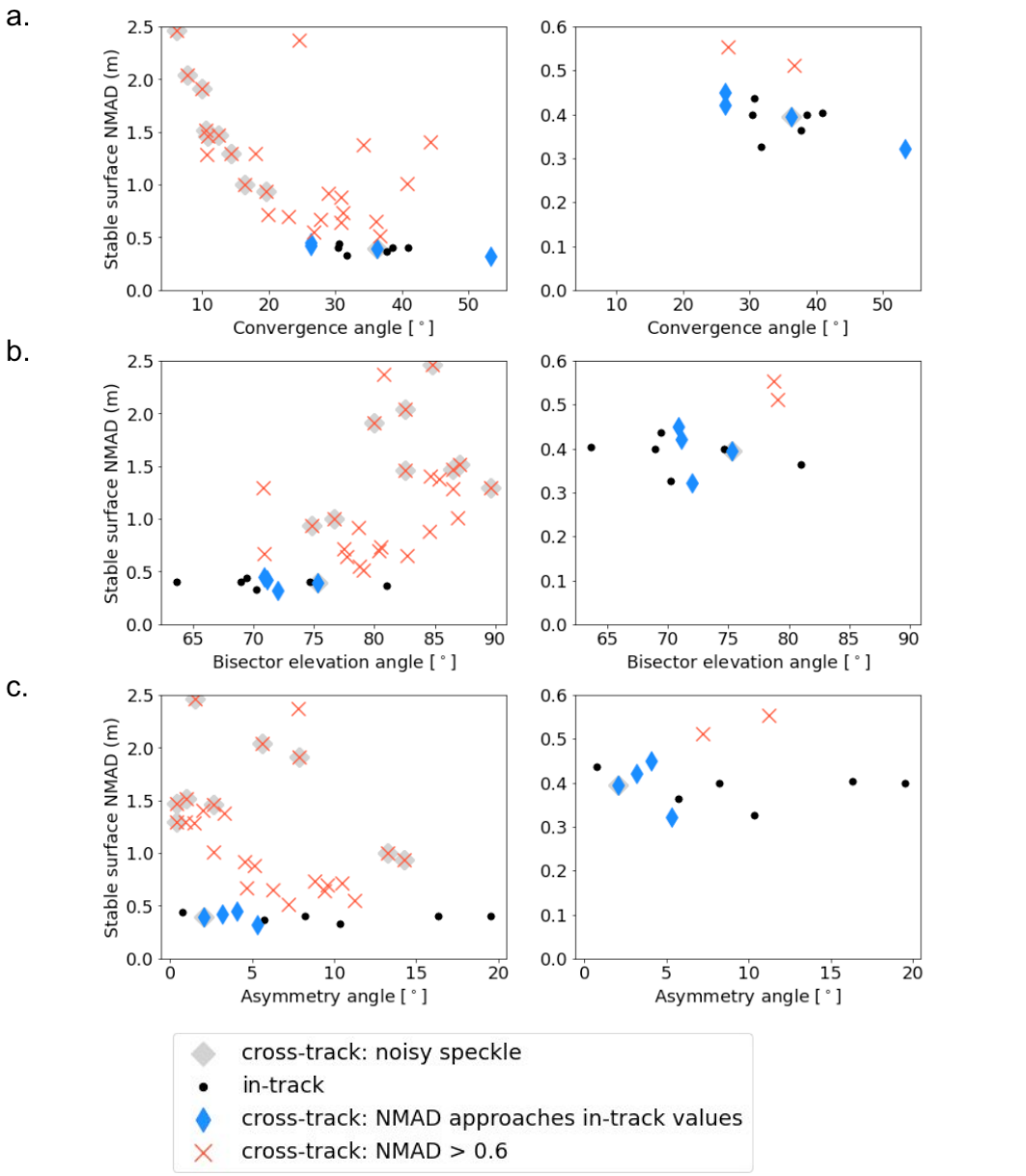

Figure 4.9 Evaluation of stereo geometry against NMAD of elevation difference values over stable surfaces for: (a) asymmetry angle , (b) convergence angle, and (c) bisector elevation angle, with zooms to the right of corresponding plots.

4.7 DISCUSSION

4.7.1 Elevation reconstruction in a permafrost environment

4.7.1.1 Evaluation of satellite stereo DSMs

The observed -0.1 m median offset (stereo minus lidar) over all evaluation areas in the DSM comparison indicates the magnitude of systematic difference we may expect when using stereo estimates of surface elevation at this site. The NMAD of elevation difference values over stable surfaces was also smaller in the DSM comparison than for the stereo snow depth map (0.30 m vs. 0.43 m), indicating lower precision in the stereo snow depth map. In both cases, precision is on the worse end of previously observed stereo snow depth measurement precision values. This variability in elevation difference values over stable surfaces may impact the accuracy of co-registration efforts (see Section 4.7.4).

Our analyses of other in-track stereo collections (i.e., Pléiades Neo) and products (i.e., ArcticDEM) illustrated potential effects of image and stereo processing. The Pléiades Neo DSM we evaluated from 13 March 2023 contained significant errors. Even though post-processing corrections were able to reduce artifacts, an accurate snow depth signal was not recoverable. Comparisons with ArcticDEM DSM products showed that variability in stereo processing can result in per-pixel elevation differences of up to ~0.5 m, a combination of processing errors from each product. Nevertheless, small median offsets (~0.1 m) and similar NMAD values suggest that ArcticDEM products may be useful as they are in regions and time periods with higher snow depth signals (i.e., > 1 m). With a sample size of four, more comprehensive evaluation is needed. Once additional products are released for WY 2022, our framework can easily be applied to evaluate additional in-track ArcticDEM DSMs.

4.7.1.2 Variability in underlying lidar DTMs

Minimizing error in snow-free DTMs is integral to producing accurate elevation difference products. The four available snow-free airborne lidar DTMs over our study site ranged between June 2017 and October 2023. In a brief comparison of these DTMs, we found per-pixel standard deviation up to ~0.5 m (median 0.05 m; 75th percentile 0.08 m; 95th percentile 0.18 m; 99th

percentile 0.44 m; Figure 4.10) and up to ~0.2 m variability over stable surfaces. While some amount of variability can be attributed to misaligned flightlines (up to 0.1 m), real signals of landscape dynamism (e.g., ground subsidence, road paving) in between collections likely make up a more substantive portion of this variability. This underlying variability underscores the importance of collecting high-quality snow-free references, taking care in the selection of co-registration surfaces, and more frequent snow-free collections in similarly dynamic environments.

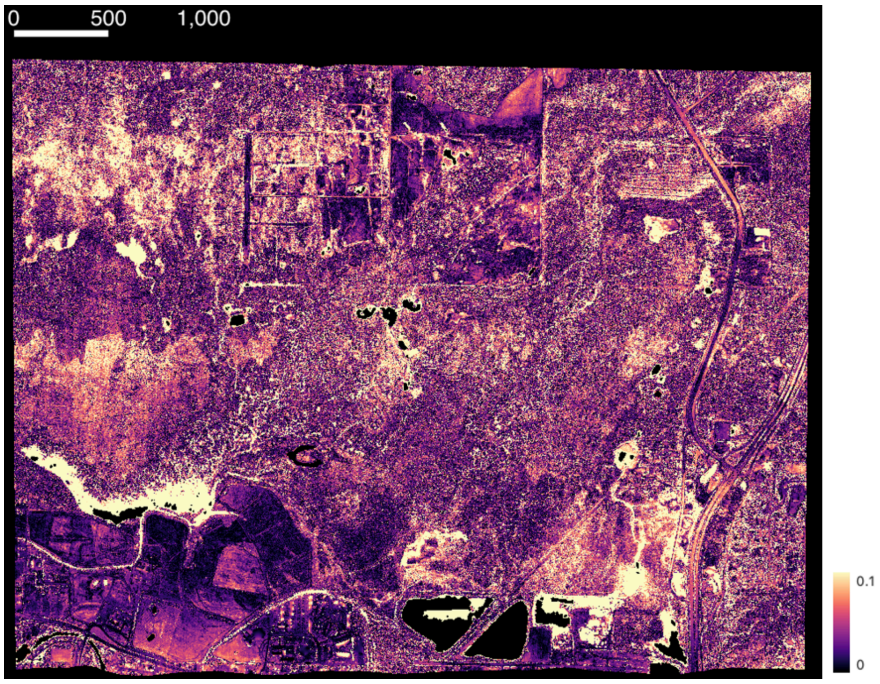

Figure 4.10 Standard deviation of co-registered snow-free airborne lidar DTMs over study site.

4.7.2 Stereo snow depth retrieval

The boreal forest snowpack evaluated in this study had a median depth of 0.63 m, and similar variability in snow depth across evaluation areas for a given collection. We did not observe substantial drifts and most of the variability in open fields appeared to arise from underlying topography (furrows, berms), snowmobile tracks, and proximity to taller vegetation. Our findings of higher elevation difference NMAD values over stable surfaces for in-track pairs (0.26–0.46 m) were slightly higher than in previous work (0.17–0.38 m; Hu et al., 2023a) as well as for NMAD values over snow in this work.

4.7.2.1 Comparison with airborne lidar

In our evaluation of stereo snow depth with UAF airborne lidar snow depth, we observed offsets of -0.3 m over open fields and smaller offsets over woody wetlands (-0.04 m; RMSE 0.19 m). We attribute the smaller offsets observed over woody wetland areas in part to conflation of vegetation height with snow depth, rather than improved stereo accuracy. The tight distribution of shrubs likely led to retrieving an elevation estimate of mixed shrub and snow surfaces rather than snow alone.

In spite of noise and residual error, stereo snow depth mapping can recover surface elevation details mixed with noise. From the airborne lidar comparison, we found that noise at this signal level makes for lower accuracy measurements. As our resampling tests indicated, spatial aggregation of measurements can help to improve the snow depth signal. When taken with our observations of consistent precision and magnitude of retrieved snow depth from two observations in WY 2021, this suggests that in the absence of other measurements, a snow depth map with lower accuracy can still be useful.

4.7.2.2 Comparison with SNOTEL snow depth records

Stereo snow depth comparisons with SNOTEL measurements show a striking difference in the distributions between the two disparate sites – for all in-track pairs, snow depth NMAD was similar to UAF lidar NMAD at Creamer's Field (0.12–0.22 m vs. 0.16 m, respectively) and much higher at the Fairbanks F. O. site (0.38–0.66 m). Higher variability in snow depth at this site can be attributed to the dense forest, multi-lane highway (i.e., Johansen Expressway), and the Noyes Slough located within 200 m of the Fairbanks F.O. site, all of which increase local topographic heterogeneity and variability, especially when compared to the field and forest at the Creamer's Field site.

Despite differences in snow depth variability between sites, for the majority of the 34 stereo collections in which SNOTEL records were available, median stereo snow depth was offset from both SNOTEL sites by ~ -0.3 m. As discussed in the spatially extensive comparison with UAF airborne lidar snow depth, this offset is likely a combination of differences between snow-free DTMs and measurement and processing error.

The one in-track exception to this observed offset was snow depth from 28 Feb 2019, with a positive median offset of 0.25 m (65% higher than SNOTEL). Stable surfaces in the study area, similar to many road systems, are predominately arranged in a curvilinear grid. The residual artifacts which manifest in this stereo DSM are roughly aligned with these stable surfaces (CCD artifacts as vertical strips that are oriented North to South, jitter artifacts are oriented East to West). Because these regions are assumed to have zero elevation change, the co-location and magnitude of residual artifacts likely led to a positively biased vertical co-registration of the stereo DSM, resulting in a positively biased stereo snow depth map for the 2019 observation.

4.7.3 Stereo pairings, geometries, and measurement precision

Overall, in-track stereo pairs generated elevation products with higher precision over snow and stable surfaces than cross-track pair products. Nevertheless, certain cross-track pairs approached the precision and quality of in-track pair products with NMAD values over stable surfaces and snow < 0.5 m (WV-3 pair from 24 Feb 2020, WV-2 pair from 26 March 2017, and a cross-sensor WV-2/WV-3 pair from 5 March 2023). Snow depth maps from cross-track stereo pairs that approached in-track quality did not visibly contain the remnant artifacts (CCD error in WV-1 or WV-2 products) in in-track snow depth maps.

Similar to the stereo geometry for in-track pairs, these cross-track pairs had BIE between 70–80˚ and variable asymmetry angle. However, cross-track pairs had a larger range of convergence angles (26–53˚ compared to in-track 30–38˚). The cross-track stereo pairs with the lowest precision over stable surfaces had BIE angles > 80˚, across a range of convergence angles and asymmetry angles. These findings agree with literature suggesting that large BIE angles can increase vertical error while asymmetry angles appear to have little impact. Taken together, these findings also expand the “optimal” range of convergence angles for stereo snow depth retrieval in areas with relatively little topographic occlusion.

4.7.4 Considerations, remaining challenges, and future work

Variability over stable surfaces is especially relevant for this workflow. Predominantly positive values over stable surfaces will produce a negative vertical shift and result in moving

snow-on DSMs lower, and vice versa. These shifts can result in an artificially low snow-covered surface and underestimate elevation change. A morphological erosion step (i.e., where polygon centroid pixels are preserved and edge pixels are ”eroded”) may reduce non-ground capture and blunder error; however, careful selection and evaluation of stable surfaces is needed to assess physical changes.

Precise co-registration of snow-free and snow-covered DEMs can be as important as inputs to derive accurate elevation change products. Ground sample distance and vertical heterogeneity in the form of slope and aspect variability can also affect the effectiveness of the co-registration approach used here (Nuth & Kääb, 2011). The finer 1 m posting we use in this work can yield more detail in vertical heterogeneity to improve precision of horizontal alignment. In our stable surface vertical alignment approach, delineated regions are assumed to be stable through time. Non-ground captures (e.g., motor vehicles, lampposts, snowbanks along road shoulders, etc.), blunders (e.g., vegetation) and actual physical changes (i.e., ground subsidence) in delineated stable surfaces in either stereo DSM or reference DTMs can impact vertical co-registration and introduce error into resultant snow depth maps.

Despite Maxar’s updates to reduce CCD geometry offsets in WV-2 collections, vertical artifacts visible as bands and stripes remain (e.g., Figure C. 1a; ~0.2–0.4 m) and appear to have larger magnitudes than those observed in WV-1 pairs corrected with ASP’s `wv_correct` utility (~0.1 m). Pointing error, the difference between the stereo modeled position and true position for a given point, accounts for the vertical error (~0.3 m magnitude) visible in the WV-1 stereo DSM from 18 April 2021 (Figure 4.3a). These remaining artifacts can constitute a substantial portion of error in retrieved signals, although they are not always present in stereo snow depth products. Future empirical corrections at the image level to correct for CCD misalignment and unmodeled attitude error will likely improve elevation models, DEM co-registration, and resulting stereo snow depth products.

This work has direct implications for mapping shallow snow in prairie snow climates, which make up one quarter of global terrestrial seasonal snow coverage. Prairie environments have little forest cover and significant wind redistribution, generating highly variable snow

accumulation, where large drifts co-exist with exposed bare ground needed for co-registration. Plentiful exposed surfaces that are stable through time may make prairie snow a more favorable environment for applying this stereo snow depth retrieval workflow.

4.8 CONCLUSIONS

In this study, we produced and evaluated 36 satellite stereo snow depth maps over boreal forest in interior Alaska from WY 2017 and WY 2022. When compared to snow depth from contemporaneous airborne lidar, stereo snow depth maps underestimated values by ~0.3 m over open areas, attributable to a combination of errors in surface elevation reconstruction, co-registration, and snow-free reference. Analyses with SNOTEL records also show median offsets with stereo snow depth measurements of ~0.2–0.3 m. While snow depth maps generated from cross-track stereo pairs were much more variable in overall quality and accuracy, DSMs and snow depth maps with higher precision were produced with constructed pairs with convergence angles > 25˚ and bisector elevation angles < 80˚. Our initial examination of Pléiades Neo DSMs reveals large processing artifacts, which require significant correction before elevation products can be used. Finally, the ArcticDEM snow depth products assessed here indicate some facility for snow depth retrieval, but more comprehensive evaluations are needed. Overall, our findings suggest that satellite stereo snow depth retrieval faces significant challenges in areas with unstable ground. However, consistency in precision, resolved detail, and internal coherence suggests that there is still valuable information to be extracted from both in-track and cross-track stereo products in shallow snowpack environments.

4.9 APPENDIX C.

Table C. 1 Metadata for WorldView-1 (WV-1), WorldView-2 (WV-2) and WorldView-3 (WV-3) images analyzed in this study. Ground sample distance (GSD) of the more nadir image in the pair, relevant acquisition geometry parameters (convergence angle [CA], base-to-height ratio [BH], asymmetry angle [AA], bisector elevation angle [BIE]), intersection area (Area), and notes on ground conditions and pair acquisition type (in-track or cross-track) also shown. Time difference (TLCTIME diff), center time of stereo pair, and individual stereo image collection datetimes (id1_dt, id2_dt) also included. Image pairs used in ArcticDEM comparisons are noted with an asterisk.

Condition	***Pair Catalog IDs***	***GSD ortho [m]***	***CA [°]***	***BH***	***AA [°]***	***BIE [°]***	***Int Area [km2]***	***TLCTIME diff***	***Center time***	***id1_dt***	***id2_dt***
snow-on, cross-track	WV02WV02_20170303_103001006337D600_1030010067955400	0.49	36.7	0.7	7.2	79.0	260	0 days 23:22:54.252800	2017-03-03 9:26:30	2017-03-03 21:07:57	2017-03-02 21:45:02
snow-on, cross-track	WV02WV02_20170326_1030010067873000_1030010068894600	0.56	53.3	1.0	5.3	72.0	630	0 days 00:01:46.696200	2017-03-26 21:59:01	2017-03-26 21:59:54	2017-03-26 21:58:08
*snow-on, in-track	WV02WV02_20180322_1030010078D20500_103001007A25FF00	0.54	30.6	0.6	0.7	69.5	273	0 days 00:00:59.599400	2018-03-22 22:05:57	2018-03-22 22:06:27	2018-03-22 22:05:27
*snow-on, in-track	WV02WV02_20190228_103001008C1B3D00_103001008F35F800	0.50	38.5	0.7	8.2	74.6	255	0 days 00:01:14.349000	2019-02-28 21:55:04	2019-02-28 21:54:27	2019-02-28 21:55:42
snow-on, cross-track	WV03WV03_20200224_104001005877A000_1040010056CD1A00	0.34	26.3	0.5	4.1	70.9	221	1 days 00:14:39.598800	2020-02-24 9:55:26	2020-02-23 21:48:06	2020-02-24 22:02:46
snow-on, cross-track	WV02WV03_20200224_10300100A1B03B00_104001005877A000	0.34	27.8	0.5	4.7	70.8	223	1 days 00:15:49.304000	2020-02-24 9:56:01	2020-02-24 22:03:55	2020-02-23 21:48:06
snow-on, cross-track	WV02WV03_20200224_10300100A310AB00_104001005877A000	0.34	30.9	0.6	5.2	84.6	214	1 days 23:39:09.130000	2020-02-24 21:37:41	2020-02-25 21:27:15	2020-02-23 21:48:06
snow-on, cross-track	WV02WV03_20200225_10300100A310AB00_1040010056CD1A00	0.36	30.4	0.5	8.7	78.3	229	0 days 23:24:29.531200	2020-02-25 9:45:00	2020-02-25 21:27:15	2020-02-24 22:02:46
snow-on, cross-track	WV02WV02_20200225_10300100A1B03B00_10300100A310AB00	0.48	30.8	0.6	9.4	77.7	249	0 days 23:23:19.826000	2020-02-25 9:45:35	2020-02-24 22:03:55	2020-02-25 21:27:15
snow-on, cross-track	WV01WV01_20200305_1020010095672E00_1020010095B67A00	0.52	36.2	0.7	6.3	82.7	340	0 days 23:37:39.798333	2020-03-05 12:20:28	2020-03-06 0:09:18	2020-03-05 0:31:38
*snow-on, in-track	WV01WV01_20210315_10200100A53A3300_10200100A6A81800	0.54	37.6	0.7	5.7	81.0	606	0 days 00:00:44.949833	2021-03-15 0:30:24	2021-03-15 0:30:02	2021-03-15 0:30:47
*snow-on, in-track	WV01_20210418_10200100AFB36000_10200100AFC76700	0.56	40.9	0.8	16.4	63.6	389	0 days 00:00:56.129041	2021-04-18 00:39:57	2021-04-18 00:39:.30	2021-04-18 00:40:25

snow-on, in-track	WV03WV03_20220311_104001007474290 0_1040010074A02E00	0.32	30.4	0.5	19.5	68.9	181	0 days 00:00:49.550400	2022-03-11 21:36:26	2022-03-11 21:36:01	2022-03-11 21:36:51
snow-on, in-track	WV03WV03_20220404_104001007448B20 0_10400100764CC800	0.34	31.8	0.6	10.3	70.2	172	0 days 00:00:49.749600	2022-04-04 21:13:54	2022-04-04 21:14:19	2022-04-04 21:13:29
dusting, in-track	WV03WV03_20221028_104001007C556C 00_104001007C7E2600	0.37	32.5	0.6	3.3	66.2	232	0 days 00:00:51.449800	2022-10-28 21:07:38	2022-10-28 21:08:04	2022-10-28 21:07:12
snow-on, cross-track	WV03WV02_20230304_104001008268B50 0_10300100E27CC000	0.3	22.9	0.4	9.6	80.4	206.7	1 days 23:46:35.581400	2023-03-04 21:28:31	2023-03-05 21:21:49	2023-03-03 21:35:13
snow-on, cross-track	WV02WV02_20230304_10300100E3D636 00_10300100E27CC000	0.5	26.8	0.5	11.2	78.8	235.5	2 days 00:26:08.455000	2023-03-04 21:48:17	2023-03-05 22:01:22	2023-03-03 21:35:13
snow-on, cross-track	WV03WV02_20230305_104001008268B50 0_10300100E3D63600	0.3	26.3	0.5	3.2	71.1	221.1	0 days 00:39:32.873600	2023-03-05 21:41:35	2023-03-05 21:21:49	2023-03-05 22:01:22
snow-on, cross-track	WV02WV03_20230313_10300100E33E9D 00_1040010081655C00	0.3	10.9	0.2	2.6	82.6	195.9	3 days 00:13:40.341800	2023-03-13 9:25:40	2023-03-14 21:32:31	2023-03-11 21:18:50
snow-on, cross-track	WV03WV02_20230313_1040010081655C0 0_10300100E4AB6200	0.3	7.9	0.1	5.6	82.6	195.5	3 days 00:13:59.691200	2023-03-13 9:25:50	2023-03-11 21:18:50	2023-03-14 21:32:50
snow-on, cross-track	WV02WV03_20230313_10300100E33E9D 00_10400100849A8200	0.3	12.5	0.2	0.4	86.5	191.8	1 days 23:58:17.192400	2023-03-13 21:33:22	2023-03-14 21:32:31	2023-03-12 21:34:13
snow-on, cross-track	WV03WV02_20230313_10400100849A82 00_10300100E4AB6200	0.3	10.5	0.2	1.0	87.0	191.4	1 days 23:58:36.541800	2023-03-13 21:33:32	2023-03-12 21:34:13	2023-03-14 21:32:50
snow-on, cross-track	WV02WV02_20230314_10300100E33E9D 00_10300100E4AB6200	0.5	10.7	0.2	1.4	86.5	237.2	0 days 00:00:19.349400	2023-03-14 21:32:40	2023-03-14 21:32:31	2023-03-14 21:32:50
snow-on, cross-track	WV02WV02_20230315_10300100E33E9D 00_10300100E4BE2100	0.5	31.1	0.6	8.8	80.5	241.3	2 days 00:27:13.352400	2023-03-15 21:46:07	2023-03-14 21:32:31	2023-03-16 21:59:44
snow-on, cross-track	WV03WV02_20230316_1040010083470D 00_10300100E4AB6200	0.4	16.4	0.3	13.2	76.7	205.7	2 days 23:42:35.152800	2023-03-16 9:24:08	2023-03-17 21:15:25	2023-03-14 21:32:50
snow-on, cross-track	WV02WV03_20230316_10300100E33E9D 00_1040010083173200	0.3	14.3	0.3	0.4	89.6	191.6	3 days 23:57:59.651800	2023-03-16 21:31:30	2023-03-14 21:32:31	2023-03-18 21:30:30
snow-on, cross-track	WV03WV02_20230316_104001008317320 0_10300100E4AB6200	0.3	6.2	0.1	1.5	84.8	191.3	3 days 23:57:40.302400	2023-03-16 21:31:40	2023-03-18 21:30:30	2023-03-14 21:32:50
snow-on, cross-track	WV02WV03_20230317_10300100E4BE21 00_1040010083470D00	0.4	36.2	0.7	2.1	75.3	213.7	0 days 23:15:41.149800	2023-03-17 9:37:35	2023-03-16 21:59:44	2023-03-17 21:15:25
snow-on, cross-track	WV02WV03_20230317_10300100E4BE21 00_1040010083173200	0.3	19.7	0.4	14.3	74.8	192.3	1 days 23:30:46.299400	2023-03-17 21:45:07	2023-03-16 21:59:44	2023-03-18 21:30:30
snow-on, cross-track	WV03WV03_20230318_1040010083470D 00_1040010083173200	0.3	19.8	0.4	10.5	77.4	192.3	1 days 00:15:05.149600	2023-03-18 9:22:58	2023-03-17 21:15:25	2023-03-18 21:30:30
snow-on, cross-track	WV02WV02_20230318_10300100E4BE21 00_10300100E4181D00	0.5	10.4	0.2	19.3	70.6	253.2	2 days 23:50:04.854200	2023-03-18 9:54:46	2023-03-16 21:59:44	2023-03-19 21:49:49
snow-on, cross-track	WV03WV02_20230318_1040010083470D 00_10300100E4181D00	0.4	28.9	0.5	4.5	78.7	207.8	2 days 00:34:23.704400	2023-03-18 21:32:37	2023-03-17 21:15:25	2023-03-19 21:49:49

snow-on, cross-track	WV02WV02_20230318_10300100E4BE2100_10300100E20EDB00	0.5	44.3	0.8	2.0	84.7	275.7	3 days 23:13:31.155800	2023-03-18 21:36:30	2023-03-16 21:59:44	2023-03-20 21:13:15
snow-on, cross-track	WV03WV02_20230319_1040010083470D00_10300100E20EDB00	0.4	18.0	0.3	0.9	70.8	214.3	2 days 23:57:50.006000	2023-03-19 9:14:20	2023-03-17 21:15:25	2023-03-20 21:13:15
snow-on, cross-track	WV03WV02_20230319_1040010083173200_10300100E4181D00	0.3	10.0	0.2	7.9	80.0	192.3	1 days 00:19:18.554800	2023-03-19 9:40:10	2023-03-18 21:30:30	2023-03-19 21:49:49
snow-on, cross-track	WV01WV01_20230319_10200100D4810C00_10200100D68F9500	0.6	40.8	0.7	2.7	86.9	370.8	3 days 00:27:46.493000	2023-03-19 12:25:30	2023-03-21 0:39:23	2023-03-18 0:11:37
snow-on, cross-track	WV03WV02_20230319_1040010083173200_10300100E20EDB00	0.3	24.6	0.4	7.8	80.7	192.34	1 days 23:42:44.856400	2023-03-19 21:21:53	2023-03-18 21:30:30	2023-03-20 21:13:15
snow-on, cross-track	WV02WV02_20230320_10300100E20EDB00_10300100E4181D00	0.5	34.3	0.6	3.3	85.4	253.05	0 days 23:23:26.301600	2023-03-20 9:31:32	2023-03-20 21:13:15	2023-03-19 21:49:49

Table C. 2 Snow depth comparison metrics by evaluation area. In-track stereopairs have one noted WorldView (WV) sensor listed in the `sensors` column. Cross-track stereo pairs have two sensors and are categorized as follows: asterisks (*) approach in-track accuracy (filters: snow depth NMAD and elevation difference NMAD over stable surfaces < 0.6 m), all other pairs did not meet filters, and carets (^) indicate pairings that additionally had visible speckling (i.e., poor reconstruction of surface elevation).

				Aggregate		***Woody wetlands***		***Field***		***Creamer's Field***		***Stable***
Water Year	***Sensors***	***Date***	***SNOTEL [m]***	***Median [m]***	***NMAD***	***Median [m]***	***NMAD***	***Median [m]***	***NMAD***	***Median [m]***	***NMAD***	***NMAD***
*WY 2017	WV02WV02	20170303_0926	0.74	0.89	0.60	1.09	0.40	0.53	0.29	0.36	0.22	0.51
*WY 2017	WV02WV02	20170326_2159	0.64	0.75	0.43	0.90	0.24	0.47	0.31	0.31	0.32	0.32
WY 2018	WV02	20180322_2205	0.76	0.66	0.44	0.86	0.29	0.69	0.29	0.17	0.24	0.44
WY 2019	WV02	20190228_2155	0.38	0.56	0.33	0.71	0.20	0.40	0.26	0.32	0.18	0.40
WY 2020	WV02WV03	20200224_0956	0.48	0.60	0.46	0.80	0.34	0.55	0.21	0.06	0.20	0.67
WY 2020	WV02WV03	20200224_2137	0.48	0.50	0.64	0.75	0.45	-0.12	0.20	0.09	0.19	0.89
*WY 2020	WV03WV03	20200224_0955	0.48	0.44	0.34	0.56	0.25	0.23	0.21	0.21	0.14	0.45
WY 2020	WV02WV02	20200225_0945	0.48	0.70	0.82	0.98	0.55	0.10	0.29	0.07	0.25	0.65
WY 2020	WV01WV01	20200305_1220	0.58	0.57	0.48	0.72	0.41	0.30	0.30	0.30	0.24	0.65

WY 2021	WV01	20210315_0030	-	0.57	0.28	0.72	0.21	0.39	0.18	0.49	0.15	0.38
WY 2021	WV01	20210418_0039	-	0.50	0.27	0.54	0.22	0.30	0.19	0.43	0.21	0.43
WY 2022	WV03	20220311_2136	0.74	0.69	0.24	0.79	0.16	0.61	0.20	0.53	0.15	0.40
WY 2022	WV03	20220404_2113	0.66	0.61	0.25	0.67	0.17	0.55	0.19	0.42	0.17	0.33
*WY 2023	WV02WV02	20230304_2148	0.53	0.68	0.58	0.20	0.27	0.32	0.26	0.88	0.34	0.55
WY 2023	WV03WV02	20230304_2128	0.53	0.52	0.49	0.24	0.23	0.13	0.29	0.62	0.32	0.70
*WY 2023	WV03WV02	20230305_2141	0.53	0.82	0.43	0.44	0.25	0.82	0.30	0.90	0.34	0.42
^WY 2023	WV02WV03	20230313_0925	0.58	1.27	0.96	0.57	0.70	0.64	0.81	1.48	0.63	1.46
^WY 2023	WV02WV03	20230313_2133	0.58	1.13	0.97	1.23	1.48	1.36	1.57	1.08	0.72	1.47
^WY 2023	WV03WV02	20230313_2133	0.58	1.29	1.19	1.30	1.58	1.84	2.03	1.08	0.86	1.51
^WY 2023	WV03WV02	20230313_0925	0.58	1.33	1.23	0.75	0.93	0.61	0.87	1.47	0.95	2.04
^WY 2023	WV02WV02	20230314_2132	0.56	0.66	0.71	0.54	0.67	0.51	0.82	0.54	0.49	1.29
^WY 2023	WV02WV02	20230315_2146	0.56	0.77	0.61	0.21	0.28	0.38	0.46	0.91	0.36	0.73
WY 2023	WV03WV02	20230316_2131	0.56	0.71	0.87	0.54	0.82	1.17	1.39	0.67	0.70	2.46
WY 2023	WV03WV02	20230316_0924	0.56	1.07	0.77	0.30	0.42	0.72	0.66	1.39	0.44	1.00
WY 2023	WV02WV03	20230316_2131	0.56	0.59	0.63	0.10	0.40	0.42	0.65	0.76	0.45	1.29
WY 2023	WV02WV03	20230317_2145	0.56	0.89	0.52	0.65	0.38	0.50	0.36	0.97	0.37	0.94
*WY 2023	WV02WV03	20230317_0937	0.56	0.76	0.53	0.38	0.21	0.46	0.26	0.96	0.34	0.40
WY 2023	WV03WV02	20230318_2132	0.56	0.62	0.49	0.25	0.24	0.36	0.27	0.77	0.34	0.92
WY 2023	WV03WV03	20230318_0922	0.56	0.82	0.57	0.29	0.29	0.51	0.36	0.99	0.35	0.72
WY 2023	WV02WV02	20230318_2136	0.56	0.44	0.57	-0.03	0.22	0.06	0.29	0.62	0.40	1.41
WY 2023	WV02WV02	20230318_0954	0.56	0.22	0.45	0.07	0.42	0.04	0.40	0.21	0.34	2.73

WY 2023	WV03WV02	20230319_0914	0.56	0.55	0.48	0.25	0.29	0.07	0.30	0.79	0.25	1.30
WY 2023	WV03WV02	20230319_0940	0.56	0.87	0.70	0.79	0.60	0.49	0.59	0.84	0.57	1.91
WY 2023	WV03WV02	20230319_2121	0.56	0.51	0.55	-0.02	0.27	0.10	0.37	0.63	0.35	2.37
WY 2023	WV01WV01	20230319_1225	0.56	0.51	0.56	0.02	0.27	0.17	0.26	0.72	0.38	1.01
WY 2023	WV02WV02	20230320_0931	0.56	0.41	0.56	-0.06	0.24	0.04	0.29	0.55	0.43	1.38

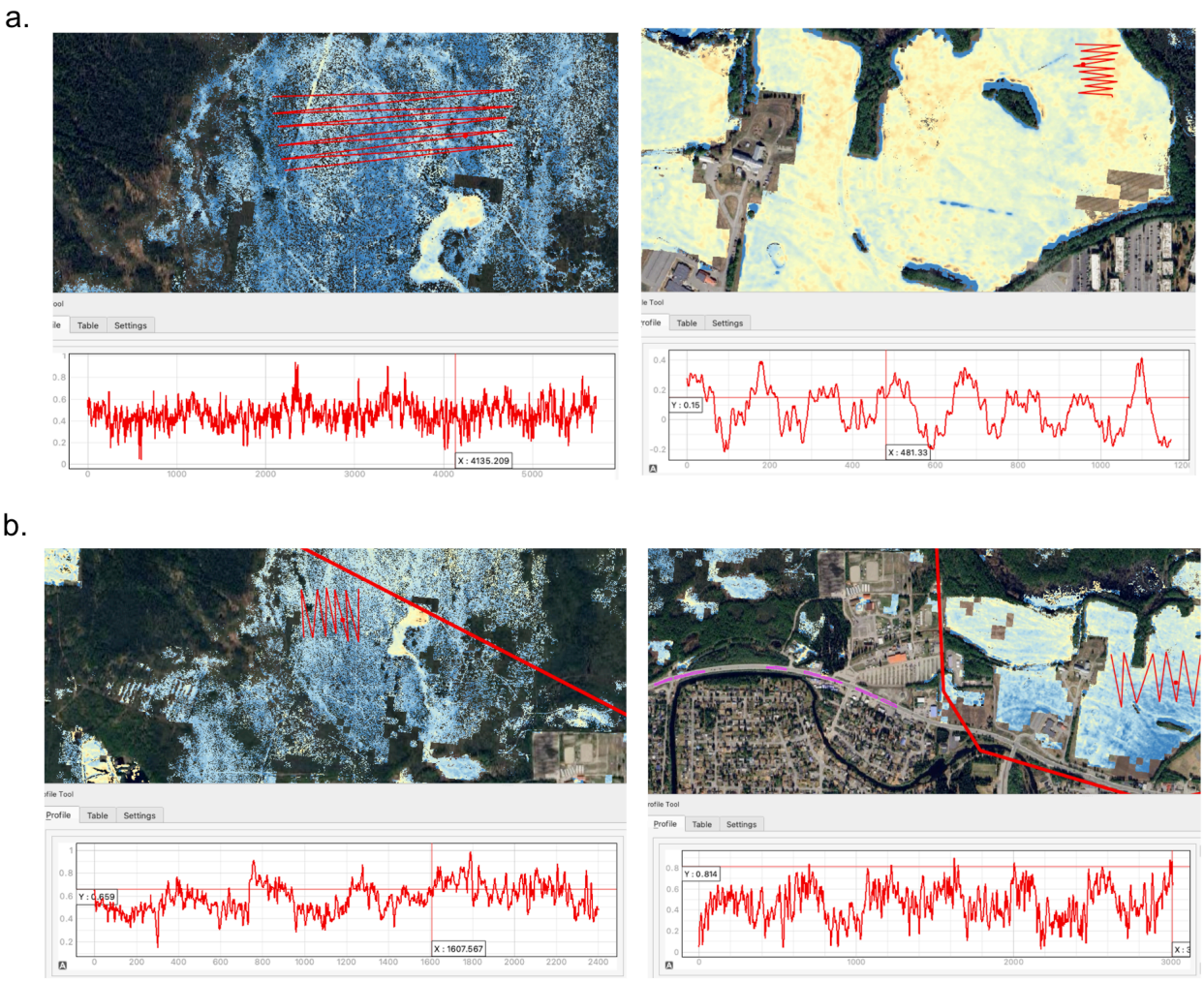

Figure C. 1 Residual artifacts in the form of (a) CCD artifacts and (b) jitter over woody wetlands (left) and Creamer's Field (right).

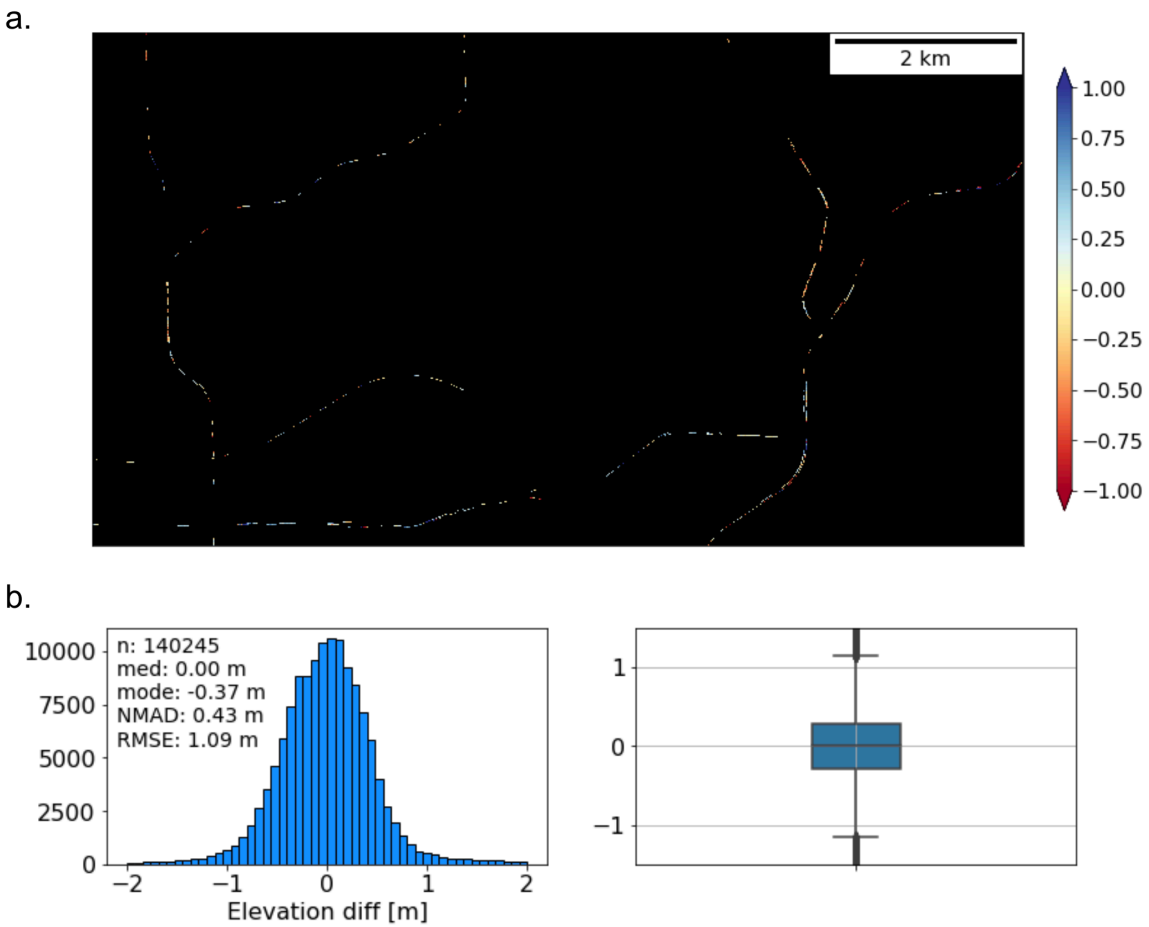

Figure C. 2 (a) map of elevation change values (WV DSM – 3DEP DTM) over stable surfaces for 11 March 2022, and (b) corresponding histogram and boxplot.

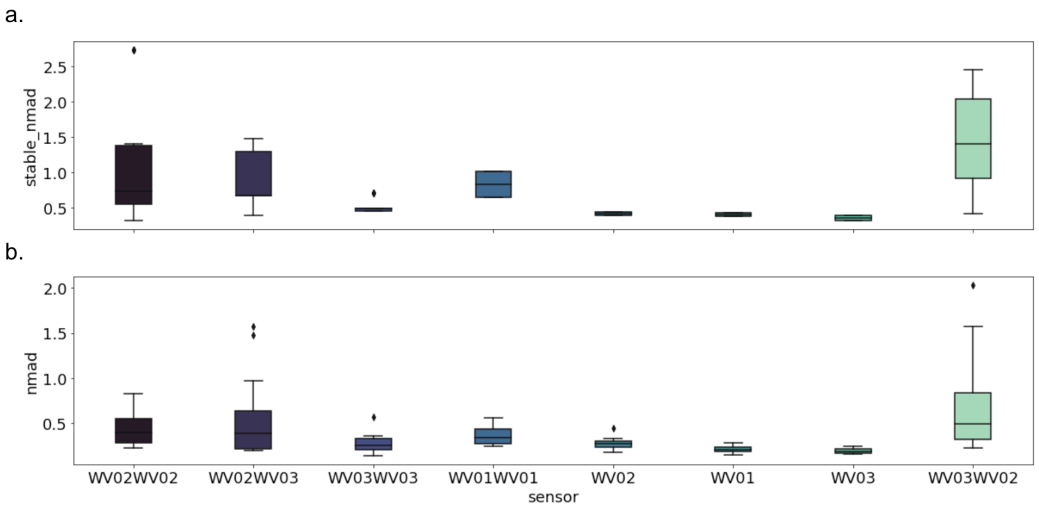

Figure C. 3 (a) Elevation difference NMAD over stable surfaces ('stable_nmad') and (b) snow depth NMAD ('nmad') by stereo pair sensors.

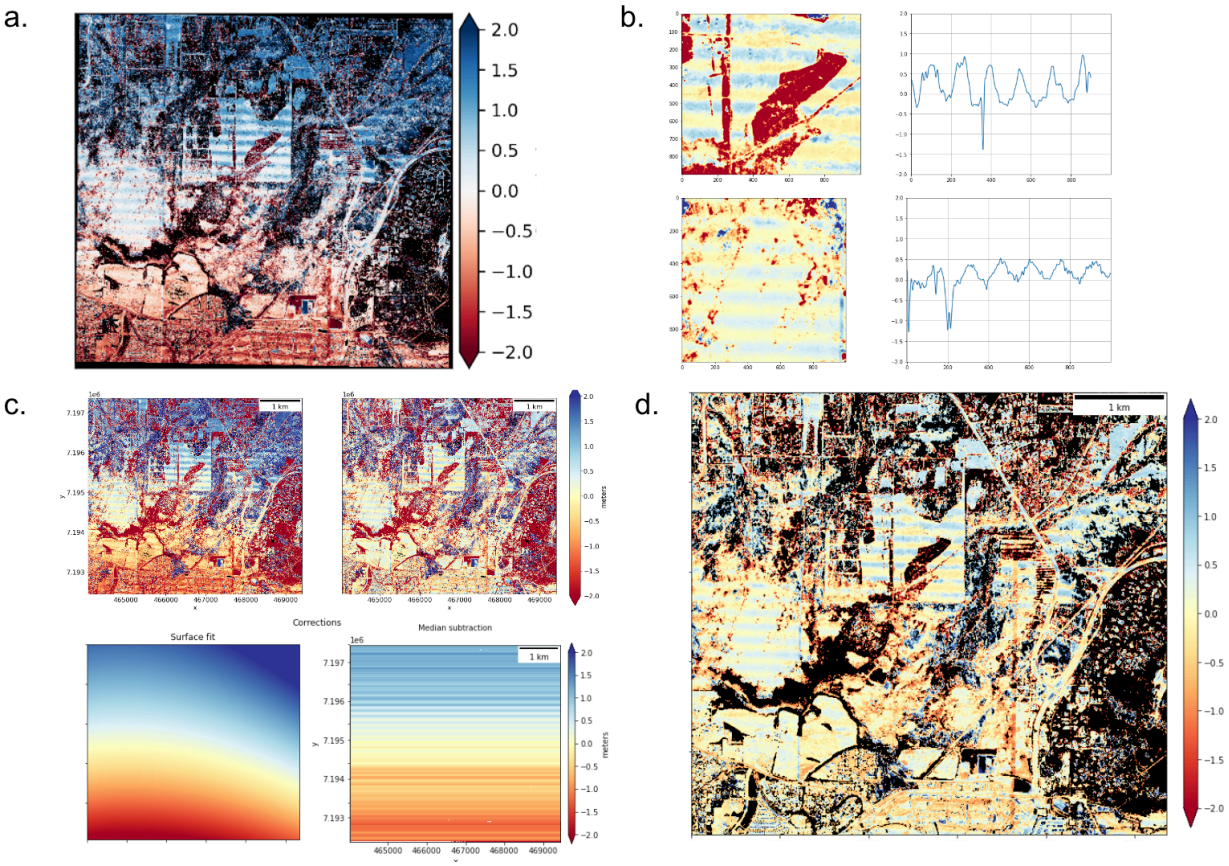

Figure C. 4 Visible artifacts in elevation difference map using (a) Pléiades Neo DSM from 13 March 2023 and 3DEP snow-free DTM. (b) high frequency linear artifacts of 0.5–1m magnitude with ~180 m wavelength over agricultural field and woody wetlands. (c) applied polynomial surface and median row-wise corrections and remaining error. (d) remaining error post-corrections.

Chapter 5. CONCLUSIONS

In a warming world with shifting precipitation regimes, the imperative to develop innovative tools and methodologies for effective preparation, management, and response has never been more pronounced. The anticipation of heightened intensity and variability in extreme weather events, characterized by prolonged periods of abnormal dryness, wetness, or other atypical meteorological conditions, necessitates a proactive approach. The ability to make informed decisions regarding resource management and risk mitigation hinges upon a comprehensive understanding of current and future conditions.

In this dissertation, extensive effort has been devoted to establishing an optimized workflow for the systematic retrieval of stereo snow depth, going beyond sporadic proof-of-concept endeavors. This dissertation considers a multitude of reference perspectives, highlighting the strengths and nuances of mapping seasonal snow depth with VHR optical stereo images. Particular attention has been dedicated to addressing limitations that can either be corrected (i.e., residual artifacts) or anticipated and integrated into interpretation (i.e., snow-free references, stable surfaces, place-based understanding).

Ch. 2 employed VHR multispectral images to train Random Forest machine learning models, enabling the classification of snow and priority land cover classes at the meter scale. Though model performance demonstrated high in-region accuracy, generalization (i.e., transfer) tests showed that a broader library of spectral reflectance is needed to improve classification across regions and seasons. Notably, the field of earth observation and dynamic land cover classification for "rare" feature classes is actively evolving. Comparisons between downsampled WorldView snow cover with coarser operational fractional snow cover products (i.e., Landsat fSCA, MODSCAG fSCA) reinforced the necessity of matching observation scale with process scales and features (such as canopy cover) to reduce misrepresentation of snow cover variability and distribution.

In Chapter 3, extensive validation datasets from the SnowEx 2017 campaign were leveraged to evaluate multi-view stereo snow depth maps over montane forest snow in Grand Mesa, Colorado. What began as a "many to one" assessment evolved into a comprehensive "many to

many" evaluation, highlighting the strengths and limitations of each measurement technique. In the first multi-year assessment of stereo snow depth maps, we demonstrated the repeatability of stereo snow depth mapping precision and accuracy. This work, conducted with a single snow-free lidar reference collected from 2016 is noteworthy, particularly for regions lacking annual lidar collections. However, we were not able to determine the feasibility of stereo snow depth mapping in the absence of a snow-free lidar reference. Nevertheless, this effort demonstrated that accurate stereo snow depth mapping was feasibly in an anomalously low snow year with median snow depths less than 1 m.

Chapter 4 pushes the boundaries of stereo snow depth mapping through evaluation over a shallow boreal forest snowpack in interior Alaska. Six in-track stereo snow depth maps and an additional 30 cross-track stereo pairs were produced and evaluated, revealing a tendency for stereo retrieval to underestimate snow depth by ~0.3 m when compared with airborne lidar. An assessment of ArcticDEM DSM products and Pléiades Neo imagery further highlighted the prevalence of residual artifacts in situations with sub-meter depth signals.

This body of work affirms the viability of stereo retrieval for snow depth estimation, while underscoring the need for continued efforts to secure adequate snow-free references and to develop a scalable method for delineating stable surfaces for accurate co-registration. In addition to continued assessment across remaining snow climates and environments, future endeavors should prioritize public stereo missions with the development of comprehensively vetted data products to enhance global access. Simultaneously, privacy and data sovereignty considerations must be taken seriously, with the direct inclusion of community representatives in decision-making processes.

BIBLIOGRAPHY

Alexandrov, O., Beyer, R., ScottMcMichael, Broxton, M., Lundy, M., Husmann, K., Edwards, L., Nefian, A., SmithB, Shean, D., Smith, T., Mstyer, Annex, A., Moratto, Z., Harguess, Dehecq, A., Aravkin, A., Meyer, J., PicoJr, … Jlaura. (2023). *NeoGeographyToolkit/StereoPipeline: ASP 3.2.0* (3.2.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7497499

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Anghileri, D., Voisin, N., Castelletti, A., Pianosi, F., Nijssen, B., & Lettenmaier, D. P. (2016). Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments. *Water Resources Research*, *52*(6), 4209–4225. https://doi.org/10.1002/2015WR017864

Arguez, A., Durre, I., Applequist, S., Vose, R. S., Squires, M. F., Yin, X., Heim, R. R., & Owen, T. W. (2012). NOAA's 1981–2010 U.S. Climate Normals: An Overview. *Bulletin of the American Meteorological Society*, *93*(11), 1687–1697. https://doi.org/10.1175/BAMS-D-11-00197.1

Arvidson, T., Gasch, J., & Goward, S. N. (2001). Landsat 7's long-term acquisition plan—An innovative approach to building a global imagery archive. *Remote Sensing of Environment*, *78*(1), 13–26. https://doi.org/10.1016/S0034-4257(01)00263-2

Austin, G. (2008). Fens of Grand Mesa, Colorado: Characterization, impacts from human activities, and restoration [M.A., Prescott College]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/304815929/abstract/5ACF1F7C43784546PQ/1

Bach, A. (2002). Snowshed Contributions to the Nooksack River Watershed, North Cascades Range, Washington. *Geographical Review*, *92*(2), 192–212. https://doi.org/10.2307/4140970

Bair, E. H., Stillinger, T., & Dozier, J. (2020). Snow Property Inversion From Remote Sensing (SPIReS): A Generalized Multispectral Unmixing Approach With Examples From MODIS and Landsat 8 OLI. *IEEE Transactions on Geoscience and Remote Sensing*, 1–15. https://doi.org/10.1109/TGRS.2020.3040328

Bales, R. C., Molotch, N. P., Painter, T. H., Dettinger, M. D., Rice, R., & Dozier, J. (2006). Mountain hydrology of the western United States. *Water Resources Research*, *42*(8). https://doi.org/10.1029/2005WR004387

Berhane, T. M., Costa, H., Lane, C. R., Anenkhonov, O. A., Chepinoga, V. V., & Autrey, B. C. (2019). The Influence of Region of Interest Heterogeneity on Classification Accuracy in Wetland Systems. *Remote Sensing*, *11*(5), Article 5. https://doi.org/10.3390/rs11050551

Berman, E. E., Coops, N. C., Kearney, S. P., & Stenhouse, G. B. (2019). Grizzly bear response to fine spatial and temporal scale spring snow cover in Western Alberta. *PLOS ONE*, *14*(4), e0215243. https://doi.org/10.1371/journal.pone.0215243

Beyer, R. A., Alexandrov, O., & McMichael, S. (2018). The Ames Stereo Pipeline: NASA's Open Source Software for Deriving and Processing Terrain Data. *Earth and Space Science*, *5*(9), 537–548. https://doi.org/10.1029/2018EA000409

Bhushan, S., & Shean, D. (2021). *Chamoli Disaster Pre-event 2-m DEM Composite: September 2015* (1.0) [dataset]. Zenodo. https://doi.org/10.5281/zenodo.4554647

Billings, W. D., & Bliss, L. C. (1959). An Alpine Snowbank Environment and Its Effects on Vegetation, Plant Development, and Productivity. *Ecology*, *40*(3), 388–397. https://doi.org/10.2307/1929755

Björk, R. G., & Molau, U. (2007). Ecology of Alpine Snowbeds and the Impact of Global Change. *Arctic, Antarctic, and Alpine Research*, *39*(1), 34–43. https://doi.org/10.1657/1523-0430(2007)39[34:EOASAT]2.0.CO;2

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, *13*(14–15), 2149–2175. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8

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

Broxton, P. D., Harpold, A. A., Biederman, J. A., Troch, P. A., Molotch, N. P., & Brooks, P. D. (2015). Quantifying the effects of vegetation structure on snow accumulation and ablation in mixed-conifer forests. *Ecohydrology*, *8*(6), 1073–1094. https://doi.org/10.1002/eco.1565

Brucker, L., Hiemstra, C., Marshall, H. P., & Elder, K. (2018). *SnowEx17 Community Snow Depth Probe Measurements, Version 1* (Version 1) [dataset]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/WKC6VFMT7JTF

Cannistra, A. F., Shean, D. E., & Cristea, N. C. (2021). High-resolution CubeSat imagery and machine learning for detailed snow-covered area. *Remote Sensing of Environment*, *258*, 112399. https://doi.org/10.1016/j.rse.2021.112399

Carpino, O. A., Berg, A. A., Quinton, W. L., & Adams, J. R. (2018). Climate change and permafrost thaw-induced boreal forest loss in northwestern Canada. *Environmental Research Letters*, *13*(8), 084018. https://doi.org/10.1088/1748-9326/aad74e

Carroll, T., Cline, D., Fall, G., Nilsson, A., Li, L., & Rost, A. (2001). NOHRSC operations and the simulation of snow cover properties for the coterminous. *U.S. Paper Presented at 69th Annual Western Snow Conference*, 1–10.

Cho, E., McCrary, R. R., & Jacobs, J. M. (2021). Future Changes in Snowpack, Snowmelt, and Runoff Potential Extremes Over North America. *Geophysical Research Letters*, *48*(22), e2021GL094985. https://doi.org/10.1029/2021GL094985

Colorado Hazard Mapping Portal. (2016). *Delta County lidar* [dataset]. https://coloradohazardmapping.com/ LidarDownload

Cosgrove, C. L., Wells, J., Nolin, A. W., Putera, J., & Prugh, L. R. (2021). Seasonal influence of snow conditions on Dall's sheep productivity in Wrangell-St Elias National Park and Preserve. *PLOS ONE*, *16*(2), e0244787. https://doi.org/10.1371/journal.pone.0244787

Creamer's Field Migratory Waterfowl Refuge Resource Inventory (p. 32). (n.d.). Retrieved October 17, 2023, from https://www.adfg.alaska.gov/static/lands/protectedareas/creamersfield/pdfs/creamers_resource_inventory.pdf

Cristea, N. C., Breckheimer, I., Raleigh, M. S., HilleRisLambers, J., & Lundquist, J. D. (2017). An evaluation of terrain-based downscaling of fractional snow covered area data sets based on LiDAR-derived snow data and orthoimagery. *Water Resources Research*, *53*(8), 6802–6820. https://doi.org/10.1002/2017WR020799

Currier, W. R., Pflug, J., Mazzotti, G., Jonas, T., Deems, J. S., Bormann, K. J., Painter, T. H., Hiemstra, C. A., Gelvin, A., Uhlmann, Z., Spaete, L., Glenn, N. F., & Lundquist, J. D. (2019). Comparing Aerial Lidar Observations With Terrestrial Lidar and Snow-Probe Transects From NASA's 2017 SnowEx Campaign. *Water Resources Research*, *55*(7), 6285–6294. https://doi.org/10.1029/2018WR024533

Currier, W. R., Sun, N., Wigmosta, M., Cristea, N., & Lundquist, J. D. (2022). The Impact of Forest-Controlled Snow Variability on Late-Season Streamflow Varies by Climatic Region and Forest Structure. *Hydrological Processes*, *n/a*(n/a), e14614. https://doi.org/10.1002/hyp.14614

Dai, C., & Howat, I. M. (2018). Detection of Saturation in High-Resolution Pushbroom Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(5), 1684–1693. https://doi.org/10.1109/JSTARS.2018.2814543

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Deschamps-Berger, C., Gascoin, S., Berthier, E., Deems, J., Gutmann, E., Dehecq, A., Shean, D., & Dumont, M. (2020). Snow depth mapping from stereo satellite imagery in mountainous terrain: Evaluation using airborne laser-scanning data. *The Cryosphere*, *14*(9), 2925–2940. https://doi.org/10.5194/tc-14-2925-2020

Dewitz, J. (2021). *National Land Cover Database (NLCD) 2019 Products* [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54

Dobrowski, S. Z. (2011). A climatic basis for microrefugia: The influence of terrain on climate. *Global Change Biology*, *17*(2), 1022–1035. https://doi.org/10.1111/j.1365-2486.2010.02263.x

Douglas, T. A., Hiemstra, C. A., Anderson, J. E., Barbato, R. A., Bjella, K. L., Deeb, E. J., Gelvin, A. B., Nelsen, P. E., Newman, S. D., Saari, S. P., & Wagner, A. M. (2021). Recent degradation of interior Alaska permafrost mapped with ground surveys, geophysics, deep drilling, and repeat airborne lidar. *The Cryosphere*, *15*(8), 3555–3575. https://doi.org/10.5194/tc-15-3555-2021

Dozier, J., & Painter, T. H. (2004). Multispectral and hyperspectral remote sensing of alpine snow properties. *Annual Review of Earth and Planetary Sciences*, *32*(1), 465–494. https://doi.org/10.1146/annurev.earth.32.101802.120404

Eberhard, L. A., Sirguey, P., Miller, A., Marty, M., Schindler, K., Stoffel, A., & Bühler, Y. (2021). Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping. *The Cryosphere*, *15*(1), 69–94. https://doi.org/10.5194/tc-15-69-2021

Elder, K., Brucker, L., Hiemstra, C., & Marshall, H. P. (2018). *SnowEx17 Community Snow Pit Measurements, Version 1* (Version 1) [dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/Q0310G1XULZS

Elder, K., Dozier, J., & Michaelsen, J. (1991). Snow accumulation and distribution in an Alpine Watershed. *Water Resources Research*, *27*(7), 1541–1552. https://doi.org/10.1029/91WR00506

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Laso Bayas, J. C., Lunga, D., McCarty, J. L., Pontius, R. G., Reinmann, A. B., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z.-F., & Estes, L. (2020). Accounting for Training Data Error in Machine Learning Applied to Earth Observations. *Remote Sensing*, *12*(6), Article 6. https://doi.org/10.3390/rs12061034

Essery, R., & Pomeroy, J. (2004). Implications of spatial distributions of snow mass and melt rate for snow-cover depletion: Theoretical considerations. *Annals of Glaciology*, *38*, 261–265. https://doi.org/10.3189/172756404781815275

Foody, G. M. (2020). Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification. *Remote Sensing of Environment*, *239*, 111630. https://doi.org/10.1016/j.rse.2019.111630

Ford, K. R., Ettinger, A. K., Lundquist, J. D., Raleigh, M. S., & Lambers, J. H. R. (2013). Spatial Heterogeneity in Ecologically Important Climate Variables at Coarse and Fine Scales in a High-Snow Mountain Landscape. *PLOS ONE*, *8*(6), e65008. https://doi.org/10.1371/journal.pone.0065008

Gascoin, S., Barrou Dumont, Z., Deschamps-Berger, C., Marti, F., Salgues, G., López-Moreno, J. I., Revuelto, J., Michon, T., Schattan, P., & Hagolle, O. (2020). Estimating Fractional Snow Cover in Open Terrain from Sentinel-2 Using the Normalized Difference Snow Index. *Remote Sensing*, *12*(18), 2904. https://doi.org/10.3390/rs12182904

Giles, P. (2001). Remote Sensing and Cast Shadows in Mountainous Terrain. *Photogrammetric Engineering & Remote Sensing*, *67*(7), 833–839.

Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F. S., Carroll, R. W. H., McNamara, J. P., & Harpold, A. A. (2022). Why does snowmelt-driven streamflow response to warming vary? A data-driven review and predictive framework. *Environmental Research Letters*, *17*(5), 053004. https://doi.org/10.1088/1748-9326/ac64b4

Grünewald, T., Stötter, J., Pomeroy, J. W., Dadic, R., Moreno Baños, I., Marturià, J., Spross, M., Hopkinson, C., Burlando, P., & Lehning, M. (2013). Statistical modelling of the snow

depth distribution in open alpine terrain. *Hydrology and Earth System Sciences*, *17*(8), 3005–3021. https://doi.org/10.5194/hess-17-3005-2013

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., & Zhao, F. (2021). Globally observed trends in mean and extreme river flow attributed to climate change. *Science*, *371*(6534), 1159–1162. https://doi.org/10.1126/science.aba3996

Hale, K. E., Jennings, K. S., Musselman, K. N., Livneh, B., & Molotch, N. P. (2023). Recent decreases in snow water storage in western North America. *Communications Earth & Environment*, *4*(1), Article 1. https://doi.org/10.1038/s43247-023-00751-3

Hall, D. K., Riggs, G. A., DiGirolamo, N. E., & Román, M. O. (2019). Evaluation of MODIS and VIIRS cloud-gap-filled snow-cover products for production of an Earth science data record. *Hydrology and Earth System Sciences*, *23*(12), 5227–5241. https://doi.org/10.5194/hess-23-5227-2019

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, *54*(2), 127–140. https://doi.org/10.1016/0034-4257(95)00137-P

Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., & Bayr, K. J. (2002). MODIS snow-cover products. *Remote Sensing of Environment*, *83*(1), 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0

Ham, J., Chen, Y., Crawford, M. M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), 492–501. https://doi.org/10.1109/TGRS.2004.842481

Hao, S., Jiang, L., Shi, J., Wang, G., & Liu, X. (2019). Assessment of MODIS-Based Fractional Snow Cover Products Over the Tibetan Plateau. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(2), 533–548. https://doi.org/10.1109/JSTARS.2018.2879666

Haugerud, R. A., & Tabor, R. W. (2009). *Geologic Map of the North Cascade Range, Washington: U.S. Geological Survey Scientific Investigations Map 2940, 2 sheets, scale 1:200,000; 2 pamphlets, 29 p. And 23 p.* https://pubs.usgs.gov/sim/2940/

Hermosilla, T., Wulder, M. A., White, J. C., & Coops, N. C. (2022). Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. *Remote Sensing of Environment*, *268*, 112780. https://doi.org/10.1016/j.rse.2021.112780

Hill, D. F., Burakowski, E. A., Crumley, R. L., Keon, J., Hu, J. M., Arendt, A. A., Wikstrom Jones, K., & Wolken, G. J. (2019). Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, *13*(7), 1767–1784. https://doi.org/10.5194/tc-13-1767-2019

Hodgson, M. E., & Bresnahan, P. (2004). Accuracy of Airborne Lidar-Derived Elevation. *Photogrammetric Engineering & Remote Sensing*, *70*(3), 331–339. https://doi.org/10.14358/PERS.70.3.331

Hodúl, M., Bird, S., Knudby, A., & Chénier, R. (2018). Satellite derived photogrammetric bathymetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, *142*, 268–277. https://doi.org/10.1016/j.isprsjprs.2018.06.015

Hu, J. M., & Shean, D. (2022). Improving Mountain Snow and Land Cover Mapping Using Very-High-Resolution (VHR) Optical Satellite Images and Random Forest Machine Learning Models. *Remote Sensing*, *14*(17), Article 17. https://doi.org/10.3390/rs14174227

Hu, J. M., & Shean, D. (2023). *Land Cover Classification, Snow Cover, and Fractional Snow-Covered Area Maps from Maxar WorldView Satellite Images, Version 1* [dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/USXB6X9CD4Q2

Hu, J. M., Shean, D., & Bhushan, S. (2023a). Six Consecutive Seasons of High-Resolution Mountain Snow Depth Maps From Satellite Stereo Imagery. *Geophysical Research Letters*. https://doi.org/10.1029/2023GL104871

Hu, J. M., Shean, D., & Bhushan, S. (2023b). *SnowEx20 3m Snow Depth DEM Time Series from High-Resolution Satellite Image Pairs, Version 1* (Version 1) [dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/7QCNCHVQMCI8

Hugonnet, R., Brun, F., Berthier, E., Dehecq, A., Mannerfelt, E. S., Eckert, N., & Farinotti, D. (2022). Uncertainty Analysis of Digital Elevation Models by Spatial Inference From Stable Terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 6456–6472. https://doi.org/10.1109/JSTARS.2022.3188922

Jan, A., & Painter, S. L. (2020). Permafrost thermal conditions are sensitive to shifts in snow timing. *Environmental Research Letters*, *15*(8), 084026. https://doi.org/10.1088/1748-9326/ab8ec4

Jennings, K., Barnhart, T. B., & Molotch, N. P. (2018). *SnowEx17 Time Series Sonic Snow Depth Measurement Array, Version 1* (Version 1) [dataset]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/5YJEYNLS1YK4

Jeong, J., & Kim, T. (2014). Analysis of Dual-Sensor Stereo Geometry and Its Positioning Accuracy. *Photogrammetric Engineering & Remote Sensing*, *80*(7), 653–661. https://doi.org/10.14358/PERS.80.7.653

Jeong, J., & Kim, T. (2016). Quantitative Estimation and Validation of the Effects of the Convergence, Bisector Elevation, and Asymmetry Angles on the Positioning Accuracies of Satellite Stereo Pairs. *Photogrammetric Engineering & Remote Sensing*, *82*(8), 625–633. https://doi.org/10.14358/PERS.82.8.625

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., & Guo, Q. (2018). The Transferability of Random Forest in Canopy Height Estimation from Multi-Source Remote Sensing Data. *Remote Sensing*, *10*(8), Article 8. https://doi.org/10.3390/rs10081183

John, A., Cannistra, A. F., Yang, K., Tan, A., Shean, D., Hille Ris Lambers, J., & Cristea, N. (2022). High-Resolution Snow-Covered Area Mapping in Forested Mountain Ecosystems Using PlanetScope Imagery. *Remote Sensing*, *14*(14), Article 14. https://doi.org/10.3390/rs14143409

Jones, H. G. (1999). The ecology of snow-covered systems: A brief overview of nutrient cycling and life in the cold. *Hydrological Processes*, *13*(14–15), 2135–2147. https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2135::AID-HYP862>3.0.CO;2-Y

Kim, E., Gatebe, C., Hall, D., Newlin, J., Misakonis, A., Elder, K., Marshall, H. P., Hiemstra, C., Brucker, L., De Marco, E., Crawford, C., Kang, D. H., & Entin, J. (2017). NASA's snowex campaign: Observing seasonal snow in a forested environment. *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1388–1390. https://doi.org/10.1109/IGARSS.2017.8127222

Koenig, J., & Gueguen, L. (2016). A comparison of land use land cover classification using superspectral WorldView-3 vs hyperspectral imagery. *2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5. https://doi.org/10.1109/WHISPERS.2016.8071721

Krauß, T., D'Angelo, P., & Wendt, L. (2019). Cross-track satellite stereo for 3D modelling of urban areas. *European Journal of Remote Sensing*, *52*(sup2), 89–98. https://doi.org/10.1080/22797254.2018.1551069

Kruckeberg, A. R. (1991). *The Natural History of Puget Sound Country*. University of Washington Press.

Kuester, Mi. (2017, April 3). *Absolute Radiometric Calibration: 2016v0*. https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/209/ABSRADCAL_FLEET_2016v0_Rel20170606.pdf

Kulakowski, D., & Veblen, T. (2006). *Historical range of variability of forest vegetation of Grand Mesa National Forest, Colorado. USDA Forest Service, Rocky Mountain Region and the Colorado Forest Restoration Institute, Fort Collins. 84 pages. (Refereed).* https://doi.org/10.13140/RG.2.1.2523.4326

Lagomasino, D., Fatoyinbo, T., Lee, S.-K., & Simard, M. (2015). High-resolution forest canopy height estimation in an African blue carbon ecosystem. *Remote Sensing in Ecology and Conservation*, *1*(1), 51–60. https://doi.org/10.1002/rse2.3

Lamare, M., Dumont, M., Picard, G., Larue, F., Tuzet, F., Delcourt, C., & Arnaud, L. (2020). Simulating optical top-of-atmosphere radiance satellite images over snow-covered rugged terrain. *The Cryosphere*, *14*(11), 3995–4020. https://doi.org/10.5194/tc-14-3995-2020

Leffler, R. J., Horvitz, A., Downs, R., Changery, M., Redmond, K. T., & Taylor, G. (2001). *Evaluation of a national seasonal snowfall record at the Mount Baker, Washington, ski*

area. (25; pp. 15–20). National Weather Digest. https://www.ncdc.noaa.gov/monitoring-content/extremes/ncec/reports/national-seasonal-snowfall-record.pdf

Lehning, M., Löwe, H., Ryser, M., & Raderschall, N. (2008). Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resources Research*, *44*(7). https://doi.org/10.1029/2007WR006545

Liu, C., Huang, X., Li, X., & Liang, T. (2020). MODIS Fractional Snow Cover Mapping Using Machine Learning Technology in a Mountainous Area. *Remote Sensing*, *12*(6), Article 6. https://doi.org/10.3390/rs12060962

Lundquist, J. D., & Flint, A. L. (2006). Onset of Snowmelt and Streamflow in 2004 in the Western United States: How Shading May Affect Spring Streamflow Timing in a Warmer World. *Journal of Hydrometeorology*, *7*(6), 1199–1217. https://doi.org/10.1175/JHM539.1

Lundquist, J. D., & Lott, F. (2008). Using inexpensive temperature sensors to monitor the duration and heterogeneity of snow-covered areas. *Water Resources Research*, *44*(4). https://doi.org/10.1029/2008WR007035

Ly, V. (2020). *Mapping snow sensor usability in the northern hemisphere with google earth engine* [ProQuest]. https://www.proquest.com/pqdtglobal1/dissertations-theses/mapping-snow-sensor-usability-northern-hemisphere/docview/2458500040/sem-2?accountid=14784

Macander, M. J., Swingley, C. S., Joly, K., & Raynolds, M. K. (2015). Landsat-based snow persistence map for northwest Alaska. *Remote Sensing of Environment*, *163*, 23–31. https://doi.org/10.1016/j.rse.2015.02.028

Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, *10*(11), 114016. https://doi.org/10.1088/1748-9326/10/11/114016

Marshall, A. M., Link, T. E., Abatzoglou, J. T., Flerchinger, G. N., Marks, D. G., & Tedrow, L. (2019). Warming Alters Hydrologic Heterogeneity: Simulated Climate Sensitivity of Hydrology-Based Microrefugia in the Snow-to-Rain Transition Zone. *Water Resources Research*, *55*(3), 2122–2141. https://doi.org/10.1029/2018WR023063

Marston, C. G., Aplin, P., Wilkinson, D. M., Field, R., & O'Regan, H. J. (2017). Scrubbing Up: Multi-Scale Investigation of Woody Encroachment in a Southern African Savannah. *Remote Sensing*, *9*(5), Article 5. https://doi.org/10.3390/rs9050419

Marti, R., Gascoin, S., Berthier, E., de Pinel, M., Houet, T., & Laffly, D. (2016). Mapping snow depth in open alpine terrain from stereo satellite imagery. *The Cryosphere*, *10*(4), 1361–1380. https://doi.org/10.5194/tc-10-1361-2016

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mazzotti, G., Essery, R., Moeser, C. D., & Jonas, T. (2020). Resolving Small-Scale Forest Snow Patterns Using an Energy Balance Snow Model With a One-Layer Canopy. *Water Resources Research*, *56*(1), e2019WR026129. https://doi.org/10.1029/2019WR026129

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McGrath, D., Webb, R., Shean, D., Bonnell, R., Marshall, H.-P., Painter, T. H., Molotch, N. P., Elder, K., Hiemstra, C., & Brucker, L. (2019). Spatially Extensive Ground-Penetrating Radar Snow Depth Observations During NASA's 2017 SnowEx Campaign: Comparison With In Situ, Airborne, and Satellite Observations. *Water Resources Research*, *55*(11), 10026–10036. https://doi.org/10.1029/2019WR024907

Millard, K., & Richardson, M. (2015). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, *7*(7), Article 7. https://doi.org/10.3390/rs70708489

Modi, P. A., Small, E. E., Kasprzyk, J., & Livneh, B. (2022). Investigating the Role of Snow Water Equivalent on Streamflow Predictability during Drought. *Journal of Hydrometeorology*, *23*(10), 1607–1625. https://doi.org/10.1175/JHM-D-21-0229.1

Molotch, N. P., & Bales, R. C. (2005). Scaling snow observations from the point to the grid element: Implications for observation network design. *Water Resources Research*, *41*(11). https://doi.org/10.1029/2005WR004229

Montesano, P. M., Neigh, C., Sun, G., Duncanson, L., Van Den Hoek, J., & Ranson, K. J. (2017). The use of sun elevation angle for stereogrammetric boreal forest height in open canopies. *Remote Sensing of Environment*, *196*, 76–88. https://doi.org/10.1016/j.rse.2017.04.024

Moore, T. R. (1987). Thermal regime of peatlands in subarctic eastern Canada. *Canadian Journal of Earth Sciences*, *24*(7), 1352–1359. https://doi.org/10.1139/e87-129

Mott, R., Schirmer, M., & Lehning, M. (2011). Scaling properties of wind and snow depth distribution in an Alpine catchment. *Journal of Geophysical Research: Atmospheres*, *116*(D6). https://doi.org/10.1029/2010JD014886

Mott, R., Schlögl, S., Dirks, L., & Lehning, M. (2017). Impact of Extreme Land Surface Heterogeneity on Micrometeorology over Spring Snow Cover. *Journal of Hydrometeorology*, *18*(10), 2705–2722. https://doi.org/10.1175/JHM-D-17-0074.1

Mott, R., Vionnet, V., & Grünewald, T. (2018). The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes. *Frontiers in Earth Science*, *6*. https://doi.org/10.3389/feart.2018.00197

Natali, S. M., Schuur, E. a. G., Trucco, C., Hicks Pries, C. E., Crummer, K. G., & Baron Lopez, A. F. (2011). Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra. *Global Change Biology*, *17*(3), 1394–1407. https://doi.org/10.1111/j.1365-2486.2010.02303.x

Native Land Digital. (2022). *Native Land territories map* [Map]. https://native-land.ca/

Noh, M.-J., & Howat, I. M. (2015). Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions. *GIScience & Remote Sensing*, *52*(2), 198–217. https://doi.org/10.1080/15481603.2015.1008621

Nolin, A. W. (2010). Recent advances in remote sensing of seasonal snow. *Journal of Glaciology*, *56*(200), 1141–1150. https://doi.org/10.3189/002214311796406077

Nolin, A. W., Sproles, E. A., Rupp, D. E., Crumley, R. L., Webb, M. J., Palomaki, R. T., & Mar, E. (2021). New snow metrics for a warming world. *Hydrological Processes*, *35*(6), e14262. https://doi.org/10.1002/hyp.14262

Nuth, C., & Kääb, A. (2011). Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change. *The Cryosphere*, *5*(1), 271–290. https://doi.org/10.5194/tc-5-271-2011

O'Neel, S., McNeil, C., Sass, L. C., Florentine, C., Baker, E. H., Peitzsch, E., McGrath, D., Fountain, A. G., & Fagre, D. (2019). Reanalysis of the US Geological Survey Benchmark Glaciers: Long-term insight into climate forcing of glacier mass balance. *Journal of Glaciology*, *65*(253), 850–866. https://doi.org/10.1017/jog.2019.66

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, Frank, Joyce, Michael, Laidlaw, Ross, Mattmann, Chris, McGurk, Bruce, Ramirez, Paul, Richardson, Megan, & Skiles, S. McKenzie. (2018). *ASO L4 Lidar Snow Depth 3m UTM Grid, Version 1* (Version 1) [dataset]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/KIE9QNVG7HP0

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Hedrick, A., Joyce, M., Laidlaw, R., Marks, D., Mattmann, C., McGurk, B., Ramirez, P., Richardson, M., Skiles, S. M., Seidel, F. C., & Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, *184*, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Painter, T. H., Dozier, J., Roberts, D. A., Davis, R. E., & Green, R. O. (2003). Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, *85*(1), 64–77. https://doi.org/10.1016/S0034-4257(02)00187-6

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, *113*(4), 868–879. https://doi.org/10.1016/j.rse.2009.01.001

Parr, C., Sturm, M., & Larsen, C. (2020). Snowdrift Landscape Patterns: An Arctic Investigation. *Water Resources Research*, *56*(12), e2020WR027823. https://doi.org/10.1029/2020WR027823

PDAL Contributors. (2022). *PDAL Point Data Abstraction Library* (2.5.2) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2616780

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.

(2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 6.

Pelto, B. M., Menounos, B., & Marshall, S. J. (2019). Multi-year evaluation of airborne geodetic surveys to estimate seasonal mass balance, Columbia and Rocky Mountains, Canada. *The Cryosphere*, *13*(6), 1709–1727. https://doi.org/10.5194/tc-13-1709-2019

Persson, H. J., & Perko, R. (2016). Assessment of boreal forest height from WorldView-2 satellite stereo images. *Remote Sensing Letters*, *7*(12), 1150–1159. https://doi.org/10.1080/2150704X.2016.1219424

Pestana, S., Chickadel, C. C., Harpold, A., Kostadinov, T. S., Pai, H., Tyler, S., Webster, C., & Lundquist, J. D. (2019). Bias Correction of Airborne Thermal Infrared Observations Over Forests Using Melting Snow. *Water Resources Research*, *55*(12), 11331–11343. https://doi.org/10.1029/2019WR025699

Pflug, J. M., Hughes, M., & Lundquist, J. D. (2021). Downscaling Snow Deposition Using Historic Snow Depth Patterns: Diagnosing Limitations From Snowfall Biases, Winter Snow Losses, and Interannual Snow Pattern Repeatability. *Water Resources Research*, *57*(8), e2021WR029999. https://doi.org/10.1029/2021WR029999

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18321-y

Pomeroy, J. W., & Brun, E. (2001). Pomeroy.pdf. In *Physical Properties of Snow* (p. 126). https://www.inscc.utah.edu/~campbell/snowdynamics/reading/Pomeroy.pdf

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Porter, C., Howat, I., Noh, M.-J., Husby, E., Khuvis, S., Danish, E., Tomko, K., Gardiner, J., Negrete, A., Yadav, B., Klassen, J., Kelleher, C., Cloutier, M., Bakker, J., Enos, J., Arnold, G., Bauer, G., & Morin, P. (2022). *ArcticDEM - strips, version 4.1*. Harvard Dataverse. https://doi.org/10.7910/DVN/C98DVS

Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., & Zhu, Z. (2019). Making Landsat Time Series Consistent: Evaluating and Improving Landsat Analysis Ready Data. *Remote Sensing*, *11*(1), Article 1. https://doi.org/10.3390/rs11010051

Rasmussen, L. A., & Tangborn, W. V. (1976). Hydrology of the North Cascades Region, Washington: 1. Runoff, precipitation, and storage characteristics. *Water Resources Research*, *12*(2), 187–202. https://doi.org/10.1029/WR012i002p00187

Revuelto, J., Billecocq, P., Tuzet, F., Cluzet, B., Lamare, M., Larue, F., & Dumont, M. (2020). Random forests as a tool to understand the snow depth distribution and its evolution in mountain areas. *Hydrological Processes*, *34*(26), 5384–5401. https://doi.org/10.1002/hyp.13951

Rittger, K., Bormann, K. J., Bair, E. H., Dozier, J., & Painter, T. H. (2021). Evaluation of VIIRS and MODIS Snow Cover Fraction in High-Mountain Asia Using Landsat 8 OLI. *Frontiers in Remote Sensing*, *2*, 8. https://doi.org/10.3389/frsen.2021.647154

Rittger, K., Raleigh, M. S., Dozier, J., Hill, A. F., Lutz, J. A., & Painter, T. H. (2020). Canopy Adjustment and Improved Cloud Detection for Remotely Sensed Snow Cover Mapping. *Water Resources Research*, *56*(6), e2019WR024914. https://doi.org/10.1029/2019WR024914

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rosenthal, W., & Dozier, J. (1996). Automated Mapping of Montane Snow Cover at Subpixel Resolution from the Landsat Thematic Mapper. *Water Resources Research*, *32*(1), 115–130. https://doi.org/10.1029/95WR02718

Sannel, A. B. K., Hugelius, G., Jansson, P., & Kuhry, P. (2016). Permafrost Warming in a Subarctic Peatland – Which Meteorological Controls are Most Important? *Permafrost and Periglacial Processes*, *27*(2), 177–188. https://doi.org/10.1002/ppp.1862

Schirmer, M., Wirz, V., Clifton, A., & Lehning, M. (2011). Persistence in intra-annual snow depth distribution: 1. Measurements and topographic control. *Water Resources Research*, *47*(9). https://doi.org/10.1029/2010WR009426

Selkowitz, D. J., Forster, R. R., & Caldwell, M. K. (2014). Prevalence of Pure Versus Mixed Snow Cover Pixels across Spatial Resolutions in Alpine Environments. *Remote Sensing*, *6*(12), Article 12. https://doi.org/10.3390/rs61212478

Selkowitz, D., Painter, T., Rittger, K., Schmidt, G., & Forster, R. (2017). *The USGS Landsat Snow Covered Area Products: Methods and Preliminary Validation* (p. 76).

Shaw, T. E., Deschamps-Berger, C., Gascoin, S., & McPhee, J. (2020). Monitoring Spatial and Temporal Differences in Andean Snow Depth Derived From Satellite Tri-Stereo Photogrammetry. *Frontiers in Earth Science*, *8*. https://www.frontiersin.org/article/10.3389/feart.2020.579142

Shaw, T. E., Gascoin, S., Mendoza, P. A., Pellicciotti, F., & McPhee, J. (2020). Snow Depth Patterns in a High Mountain Andean Catchment from Satellite Optical Tristereoscopic Remote Sensing. *Water Resources Research*, *56*(2), e2019WR024880. https://doi.org/10.1029/2019WR024880

Shean, D. (2017). *High Mountain Asia 8-meter DEMs Derived from Cross-track Optical Imagery, Version 1* [dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/0MCWJJH5ABYO

Shean, D. (2018). *Dshean/snowtools* [Python]. https:// github.com/dshean/snowtools

Shean, D., Bhushan, S., Lilien, D., Knuth, F., Schwat, E., Meyer, J., Sharp, M., & Hu, M. (2021). *dshean/demcoreg: V1.1.0* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.5733347

Shean, D. E., Alexandrov, O., Moratto, Z. M., Smith, B. E., Joughin, I. R., Porter, C., & Morin, P. (2016). An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 101–117. https://doi.org/10.1016/j.isprsjprs.2016.03.012

Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, *231*, 111199. https://doi.org/10.1016/j.rse.2019.05.018

Steltzer, H., Landry, C., Painter, T. H., Anderson, J., & Ayres, E. (2009). Biological consequences of earlier snowmelt from desert dust deposition in alpine landscapes. *Proceedings of the National Academy of Sciences*, *106*(28), 11629–11634. https://doi.org/10.1073/pnas.0900758106

Stewart, I. T. (2009). Changes in snowpack and snowmelt runoff for key mountain regions. *Hydrological Processes*, *23*(1), 78–94. https://doi.org/10.1002/hyp.7128

Sturm, M., & Holmgren, J. (2018). An Automatic Snow Depth Probe for Field Validation Campaigns. *Water Resources Research*, *54*(11), 9695–9701. https://doi.org/10.1029/2018WR023559

Sturm, M., & Liston, G. E. (2021). Revisiting the Global Seasonal Snow Classification: An Updated Dataset for Earth System Applications. *Journal of Hydrometeorology*, *22*(11), 2917–2938. https://doi.org/10.1175/JHM-D-21-0070.1

Sumbul, G., Charfuelan, M., Demir, B., & Markl, V. (2019). Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 5901–5904. https://doi.org/10.1109/IGARSS.2019.8900532

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *2017 IEEE International Conference on Computer Vision (ICCV)*, 843–852. https://doi.org/10.1109/ICCV.2017.97

Tanniru, S., & Ramsankaran, R. (2023). Passive Microwave Remote Sensing of Snow Depth: Techniques, Challenges and Future Directions. *Remote Sensing*, *15*(4), Article 4. https://doi.org/10.3390/rs15041052

Tarnocai, C., Canadell, J. G., Schuur, E. a. G., Kuhry, P., Mazhitova, G., & Zimov, S. (2009). Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochemical Cycles*, *23*(2). https://doi.org/10.1029/2008GB003327

Thibault, I., & Ouellet, J.-P. (2005). Hunting behaviour of eastern coyotes in relation to vegetation cover, snow conditions, and hare distribution. *Écoscience*, *12*(4), 466–475. https://doi.org/10.2980/i1195-6860-12-4-466.1

Thorn, C. E., & Hall, K. (2002). Nivation and cryoplanation: The case for scrutiny and integration. *Progress in Physical Geography: Earth and Environment*, *26*(4), 533–550. https://doi.org/10.1191/0309133302pp351ra

Trujillo, E., & Molotch, N. P. (2014). Snowpack regimes of the Western United States. *Water Resources Research*, *50*(7), 5611–5623. https://doi.org/10.1002/2013WR014753

Tsai, Y.-L. S., Dietz, A., Oppelt, N., & Kuenzer, C. (2019). Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique. *Remote Sensing*, *11*(8), Article 8. https://doi.org/10.3390/rs11080895

Updike, T., & Comp, C. (2010). *Radiometric Use of WorldView-2 Imagery* (p. 17).

U.S. Geological Survey. (2017a). *1/3rd arc-second Digital Elevation Models (DEMs)—USGS National Map 3DEP Downloadable Data Collection* [dataset]. https://www.sciencebase.gov/catalog/item/4f70aa9fe4b058caae3f8de5

U.S. Geological Survey. (2017b). *USGS lidar point cloud* [dataset]. https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/ Elevation/LPC/Projects/CO MesaCounty 2015/

U.S. Geological Survey, Earth Resources Observation And Science Center. (2018). *Collection-1 Landsat Level-3 Fractional Snow Covered Area (FSCA) Science Product* [Tiff]. U.S. Geological Survey. https://doi.org/10.5066/F7XK8DS5

Valovcin, F. R. (1976). *Snow/cloud discrimination* (AFGL-TR-766-1780174, ADA 032385). AIR FORCE GEOPHYSICS LABORATORY. https://apps.dtic.mil/dtic/tr/fulltext/u2/a032385.pdf

Vikhamar, D., & Solberg, R. (2003). Subpixel mapping of snow cover in forests by optical remote sensing. *Remote Sensing of Environment*, *84*(1), 69–82. https://doi.org/10.1016/S0034-4257(02)00098-6

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., & Weingartner, R. (2007). Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005653

Walker, D. A., Halfpenny, J. C., Walker, M. D., & Wessman, C. A. (1993). Long-term Studies of Snow-Vegetation Interactions: A hierarchic geographic information system helps examine links between species distributions and regional patterns of greenness. *BioScience*, *43*(5), 287–301. https://doi.org/10.2307/1312061

Walters, R. D., Watson, K. A., Marshall, H.-P., McNamara, J. P., & Flores, A. N. (2014). A physiographic approach to downscaling fractional snow cover data in mountainous regions. *Remote Sensing of Environment*, *152*, 413–425. https://doi.org/10.1016/j.rse.2014.07.001

Watson, A., Davison, R. W., & French, D. D. (1994). Summer Snow Patches and Climate in Northeast Scotland, U.K. *Arctic and Alpine Research*, *26*(2), 141–151. https://doi.org/10.1080/00040851.1994.12003050

Webb, R., McGrath, D., Hale, K., & Molotch, N. P. (2019). *SnowEx17 Ground Penetrating Radar, Version 2* (Version 2) [dataset]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/G21LGCNLFSC5

Webb, R., Raleigh, M. S., McGrath, D., Molotch, N. P., Elder, K., Hiemstra, C., Brucker, L., & Marshall, H. P. (2020). Within-Stand Boundary Effects on Snow Water Equivalent Distribution in Forested Areas. *Water Resources Research*, *56*(10), e2019WR024905. https://doi.org/10.1029/2019WR024905

Winkler, D. E., Butz, R. J., Germino, M. J., Reinhardt, K., & Kueppers, L. M. (2018). Snowmelt Timing Regulates Community Composition, Phenology, and Physiological Performance of Alpine Plants. *Frontiers in Plant Science*, *0*. https://doi.org/10.3389/fpls.2018.01140

Winstral, A., Elder, K., & Davis, R. E. (2002). Spatial Snow Modeling of Wind-Redistributed Snow Using Terrain-Based Parameters. *Journal of Hydrometeorology*, *3*(5), 524–538. https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2

Wrzesien, M. L., Pavelsky, T. M., Durand, M. T., Dozier, J., & Lundquist, J. D. (2019). Characterizing Biases in Mountain Snow Accumulation From Global Data Sets. *Water Resources Research*, *55*(11), 9873–9891. https://doi.org/10.1029/2019WR025350

Wu, W., Li, M.-F., Xu, X., Tang, X.-P., Yang, C., & Liu, H.-B. (2021). The transferability of random forest and support vector machine for estimating daily global solar radiation using sunshine duration over different climate zones. *Theoretical and Applied Climatology*, *146*(1), 45–55. https://doi.org/10.1007/s00704-021-03726-6

Xin, Q., Woodcock, C. E., Liu, J., Tan, B., Melloh, R. A., & Davis, R. E. (2012). View angle effects on MODIS snow mapping in forests. *Remote Sensing of Environment*, *118*, 50–59. https://doi.org/10.1016/j.rse.2011.10.029

Yeend, W. E. (1969). Quaternary geology of the Grand and Battlement Mesas area, Colorado. In *Professional Paper* (617). U.S. G.P.O.,. https://doi.org/10.3133/pp617

Yin, T., Montesano, P. M., Cook, B. D., Chavanon, E., Neigh, C. S. R., Shean, D., Peng, D., Lauret, N., Mkaouar, A., Regaieg, O., Zhen, Z., Qin, R., Gastellu-Etchegorry, J.-P., & Morton, D. C. (2023). Modeling forest canopy surface retrievals using very high-resolution spaceborne stereogrammetry: (II) optimizing acquisition configurations. *Remote Sensing of Environment*, *298*, 113824. https://doi.org/10.1016/j.rse.2023.113824

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *5*(1), 44–53. https://doi.org/10.1093/nsr/nwx106

Zong, S., Lembrechts, J. J., Du, H., He, H. S., Wu, Z., Li, M., & Rixen, C. (2022). Upward range shift of a dominant alpine shrub related to 50 years of snow cover change. *Remote Sensing of Environment*, *268*, 112773. https://doi.org/10.1016/j.rse.2021.112773

VITA

Jyun-Yi Michelle Hu was born in Madison, Wisconsin, the ancestral land of the Ho-Chunk. She spent her formative years embracing cold, snowy Wisconsin winters and the frozen expanse of the Yahara chain of lakes. Michelle earned her Bachelor of Science in Biology and International Studies, complemented by a certificate (minor) in Environmental studies, and a GIS Professional Development Certificate from the University of Wisconsin–Madison. She attained a Master of Science in Water Resources Science from Oregon State University, where she was advised by Dr. Anne W. Nolin.

Looking ahead, Michelle plans to lead interdisciplinary investigations and cultivate extensive collaborations across diverse fields of application. She is particularly interested in improving accessibility to technical methodologies through a commitment to a reproducible and open approach. Michelle's primary objective is to develop actionable knowledge, driven by the overarching goal to catalyze the convergence of fundamental exploratory research with practical applications for societal benefit.