Evaluating Camera Traps as Ground-Based Remote Sensing Networks, Linking Snow and Wildlife

Catherine Marina Breen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee

Laura Prugh, Chair

Jessica Lundquist

L. Monika Moskal

Program Authorized to Offer Degree:

School of Environmental and Forest Sciences

University of Washington

Abstract

Evaluating Camera Traps as Ground-Based Remote Sensing Networks, Linking Snow and Wildlife

Catherine Marina Breen

Chair of the Supervisory Committee:

Laura Prugh

School of Environmental and Forest Sciences

Seasonal snow covers 31% of the Northern Hemisphere in midwinter, playing an influential role in wildlife communities and water storage for hydrological processes. Monitoring seasonal snow comprises three main approaches, including remote sensing, hydrological modeling, and field techniques. However, current methods can be costly and less accurate in complex terrain, forests, and on cloudy days, and methods present trade-offs in spatial and temporal resolution. Using a large-scale camera network of over 1000+ cameras based in Norway from a project called Scandcam (viltkamera.nina.no), my dissertation sought to determine how camera traps (also referred to as remote cameras, wildlife cameras, or game cameras) can connect on-the-ground and remote sensing snow monitoring techniques for

ecological and hydrological applications. Chapter 2 initialized a conceptual and analytical framework for using wildlife cameras to supplement and validate optical remote sensing products for snow cover. I identified advantages and disadvantages and presented wildlife camera traps as a novel method to improve snow monitoring on cloudy days and at high latitudes, regions where satellites are less accurate. In chapter 3, I provided a methodological example in which I used computer vision and machine learning to extract snow depth from remote cameras from NASA and University of Washington snow measurement campaigns, serving as a case study for snow information extraction from imagery as well as a pipeline for future model development. In chapter 4, I demonstrated how information from cameras can answer questions about animal activity in response to changes in snow conditions. I found that information from the camera, including location and time of day, can predict the strength of the snow surface, and subsequently changes in diel activity in roe deer (*Capreolus capreolus*) but not for mountain hares (*Lepus timidus*). In chapter 5, I examined how we can leverage wildlife cameras to supplement passive microwave remote sensing from the Advanced Microwave Scanning Radiometer 2 (AMSR2) for rain-on-snow mapping. I developed a model to detect rain in camera trap imagery and demonstrated a 1% improvement in AMSR2-detected rain-on-snow events. In chapter 6, I combined data from passive acoustic recorders (AudioMoths from Open Acoustic Devices) with camera data to improve the ability to track rain-on-snow events. I demonstrated that this technique improves human detection of and neural network predictions for rain-on-snow events. These findings illustrate that wildlife cameras can operate as “mini weather stations” for winter snow monitoring, and that by providing on-the-ground information, cameras demonstrate the ability to improve our snow monitoring techniques and understanding of wildlife communities in a changing climate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am forever grateful to my committee, Jessica Lundquist, L. Monika Moskal, John Odden, Laura Prugh, and Joseph Hellerstein, for their continuous support and guidance over the past five years. Thank you especially to Laura, who I have learned so much from and whose consistent insights and feedback enabled me to grow tremendously as a researcher. From our initial email exchange, when Laura asked about my graduate school intentions and my cross-country skills, I knew I not only admired her career but her mentorship style as well. Since that day, I have been grateful for her efforts to sharpen my skills and appreciated her deep commitment to my success. I am also deeply grateful to my NASA technical adviser, Carrie Vuyovich, for her support, insights, and inclusion in the NASA mission.

Living and working in Norway has been a highlight of my PhD career and enriched both the research and my graduate school experience. A special thank you to John Odden and the entire team at the Norwegian Institute for Nature Research for supporting the multiple international trips, filling out the forms required for my visas, and helping me to integrate into Norwegian office and city life. Thank you to all the colleagues at NINA who made my ability to conduct work in Norway possible. Sunniva Bahlk and Nina Myhr provided valuable access to camera trap data. Neri von Hornstedt, John Linnell, and Inger Maren Rivrud deepened my knowledge of Norwegian ecology and wildlife modeling. Thank you to Zea Walton for initiating the collaboration between NINA and the University of Washington, including me as part of her Erasmus+ mobility exchange grant, and subsequently beginning my research journey in Norway. An additional thank you to my fellow NINA PhD officemates, Calle Johannesson and Elina Lungrin, for your friendship and support. I am excited to see where your research takes you, and I look forward to many more adventures and US-Europe cross-overs with you both.

A thank you to the members of the Prugh lab that I had the pleasure of overlapping with: Taylor Ganz, Sam Kreling, Yasmine Hentati, Ben Sullender, Ellie Reese, Jiangyue Wang, Prashant Mahajan, Calum Cunningham, Tristan Nuñez, Michael Procko, Kayla Shivley, Peter Mahoney, and Becca Windell. I always looked forward to lab meetings to spend time with you all, and I benefitted tremendously from the diverse perspectives in the lab. I am inspired by all of you. Thank you for your collaboration and friendship.

Finally, a thank you to my friends and family. From friendships in Norway, Seattle, and all over the U.S., I am grateful to the many friends who relentlessly supported me, and conveniently planned dinners and trips to make sure I stepped away from my computer from time to time. You made life fulfilling and fun. The acknowledgements would not be complete without a thank you to my parents, for their relentless support, and consistent preparation of fresh coffee whenever I came to visit for a respite from work. To my siblings, Betsy, Andrew, Peter, and Mollie, thank you for setting examples of what it looks like to follow your passions fervently. I look up to all of you very much. I'd also like to thank my identical twin sister Mollie for finding me around the world while I was conducting research in both Seattle and Oslo. Thanks for bringing that half-eaten chocolate chip cookie across the Atlantic.

Chapter 1. INTRODUCTION

1.1 Background

Snow is an essential component of habitat quality for wildlife, water storage for hydrological processes, and human uses such as agriculture, forestry, and tourism (Mankin et al., 2015; Rizzi et al., 2018). However, monitoring snow conditions, especially in hard-to-reach and complex terrain, is difficult, due to the lack of weather stations and the field efforts required for manual measurements. Remote sensing from sources such as optical satellites, active and passive microwave satellites, and radar sensing have increased spatial and temporal coverage for snow observations, but trees, clouds, and the polar night still limit measurements depending on the source (Dietz et al., 2012; Nolin, 2010; Tsang et al., 2022). Light detection and ranging (lidar) has made recent advances for high-resolution snow maps, but observations can be costly and limited to specific areas (Jacobs et al., 2021). Recent work has called for new tools for monitoring snow (Penczykowski et al., 2017), especially tools optimized for fine-spatial and temporal scales (Reinking et al., 2022). Specifically for connecting the impacts of snow on wildlife, recent work has called for improved tools for snow measurements that are optimized at the spatial and temporal scale of wildlife activity (Boelman et al., 2019; Reinking et al., 2022).

Camera traps, sometimes referred to as “wildlife cameras” or “remote cameras,” are a popular tool for noninvasive animal monitoring, including animal abundance surveys, monitoring movement, and observing behavior (Bassing et al., 2023; Burton et al., 2015). Recently, researchers have called for increased use of the camera’s “bycatch information,” or the ancillary environmental information in the camera image (Brown et al., 2016; Sonnentag et al., 2012; Steenweg et al., 2017). Cameras automatically record temperature and date-time metadata,

and visually indicate the evidence of meteorological phenomena, including snow, hail, and vegetation green-up (Alison et al., 2024; Brown et al., 2016). With the ability to capture environmental data and scale up to networks that range from ten to hundreds of cameras, camera traps hold great potential as a ground-based remote sensing tool that can simultaneously supplement and validate remote sensing observations for snow as well as provide insight on the snow conditions simultaneous to the wildlife detection.

One extensive camera trap network is the Scandcam project, hosted by the Norwegian Institute for Nature Research (project website: viltkamera.nina.no). The first cameras were installed in 2010, and over the course of the next decade, the network expanded into a larger camera trap network spanning 1000+ cameras, with the aim to survey the movement, behavior, and population metrics of the recolonizing Eurasian lynx (*Lynx lynx*) (Carricondo-Sanchez et al., 2017; Hofmeester et al., 2019). The cameras span a 10° (58° - 68°) latitudinal gradient and a maritime-montane forest snow classification gradient (Fig. 1.1A; Sturm & Liston, 2021), offering a unique chance to ask questions related to the utility of camera traps to monitor snow, and the ability to connect that information to improve remotely sensed snow maps as well as wildlife understanding. These cameras reflect the growing use of camera networks worldwide (Burton et al., 2015), and we leverage networks from the 2019-2020 NASA SnowEx Campaign and a snow measurement campaign from the University of Washington for developing methods for snow property extraction across various snow conditions and ecosystems (Fig. 1.1B and C). With the growing use of cameras combined with the limitations of current snow monitoring techniques, my dissertation sought to build a framework for combining remote cameras with other remotely sensed datasets, develop new methods for extracting snow properties from

cameras, and demonstrate the use of wildlife cameras to answer novel ecological questions on the behavior of wildlife in a changing climate.

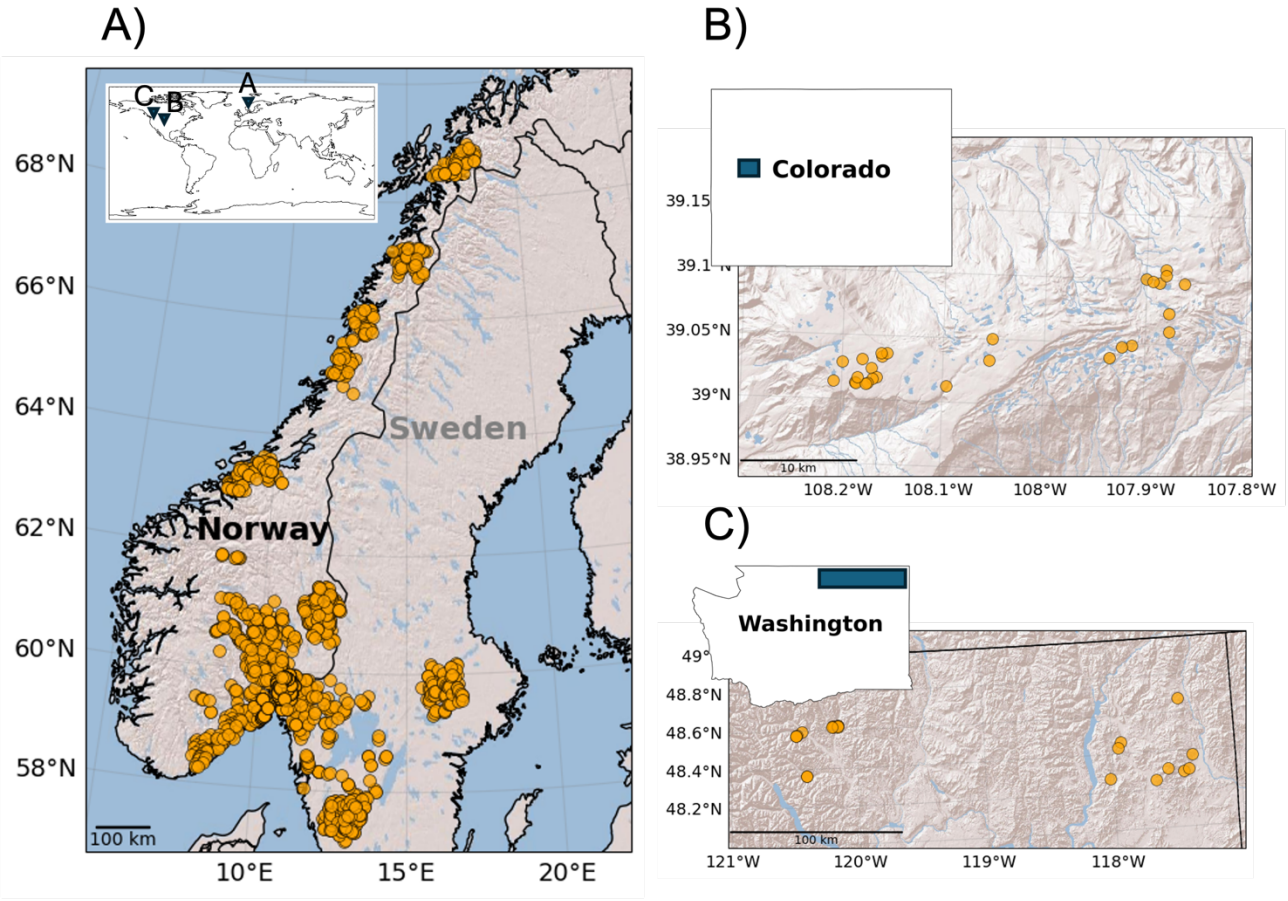


Figure 1.1. A) Map of cameras deployed for wildlife monitoring from the Scandcam wildlife camera network (project page: viltkamera.nina.no) used in this paper for comparison with satellite observation and the study of wildlife activity. We leveraged camera networks from B) the 2019-20 NASA SnowEx campaigns in Grand Mesa, CO, and C) 2020 and 2021 snow measurement campaigns from the University of Washington in Okanogan and Chewelah County to develop a model to detect snow depth from camera imagery. We used the World Shaded Relief map from the ArcGIS REST Services Directory to show general terrain shading (Esri 2014).

1.2 Goals

Camera traps currently represent an untapped resource for snow monitoring but establishing them as a method involves new frameworks for extracting snow information and connecting the devices to satellite observations. In Chapter 2, I developed a conceptual and analytical framework for validating remote sensing products for snow cover using camera traps. Specifically, I compared the agreement between snow cover detected from camera traps using visual inspection and snow detected from the Moderate Resolution Imaging Spectroradiometer (MODIS). First, I identified factors that affect agreement, demonstrating the potential of cameras to offer insight when optical satellites fall short. Second, I used the camera traps to identify the threshold to convert the continuous snow cover index provided by MODIS (ranging from 0 to indicate no snow cover to 100 to indicate complete snow cover), demonstrating the utility of cameras to convert a MODIS map to binary "snow" or "no snow" snow product. Third, I compared the cameras to the MODIS cloud-gap-filled product, identifying a decrease in accuracy on day 3, suggesting the day after which to supplement MODIS maps with camera values.

The ability to tap into camera information relies on efficient data extraction pipelines for snow extraction. Computer vision and machine learning offer efficient data extraction from camera traps (Beery et al., 2019; Tuia et al., 2022). However, no model for extracting snow information from camera traps exists that can deploy to multiple networks. In Chapter 3, I developed a keypoint detection model for extracting snow pole depth information when installed in front of time-lapse cameras. I leveraged a camera network from the 2019-20 NASA SnowEx campaign, and I examined cases when the model was successful and unsuccessful in estimating snow depth in the image. Then, I assessed how well the model worked on a dataset from 2020

and 2021 field campaigns from the University of Washington, mimicking a pipeline for a future user by which they deploy the model on their snow pole network of interest. By analyzing how many images the model needed to see during training to achieve similar accuracy relative to other snow depth measurement devices, I identified the utility of this model as a step towards a generalizable algorithm.

Snow monitoring from camera traps offers new insight into the impacts of snow on wildlife at fine-spatial and temporal scales. Recent work has emphasized the role of environmental factors in changing wildlife diel activity (Gaynor et al., 2018). Snow conditions can change as often as hourly (United States, Army Corps of Engineers, 1956), and in Chapter 4, I analyzed hourly changing snow conditions on animal activity. I built a model that leveraged information from the camera, including location, hour of day, and time of year, to predict the strength of the snow and the subsequent impacts on wildlife diel activity. I compared changes in diel cycles for two Norwegian mammals, the roe deer (*Capreolus capreolus*) and mountain hare (*Lepus timidus*), which face similar predator pressures from lynx but have different morphology adaptations to snow (Hackländer & Zachos, 2020; Kelsall, 1969). This work highlighted the capability of camera traps to capture fine-scale, sub-daily snow conditions. When linked to hourly animal activity, this study identified differences in environmental pressures among animals depending on their locomotive adaptations to snow and diel plasticity.

Similar to snow hardness, prior research has identified severe snow icing caused by rain-on-snow (ROS) events as detrimental to wildlife due to impacts on movement and foraging abilities (Hansen et al., 2011). Passive microwave remote sensing using the Advance Microwave Sensor Radiometer 2 (AMSR2) has had success in detecting rain-on-snow events in Alaska and Quebec, as well as deriving freeze-thaw estimates (van de Kerk, 2018; Langlois et al., 2017; Pan

et al., 2018). However, more work is needed in cloudy, maritime snowpacks. In Chapter 5, I compared the AMSR2 ROS record to weather stations, establishing an accuracy benchmark and investigating factors that affected the accuracy of the AMSR2 relative to weather stations. I found impacts from trees and proximity to water, and I demonstrated that by incorporating information from cameras, we can improve the ROS record, furthering their use in dense forests and cloudy regions.

Precipitation detection continues to be a difficult problem for ecologists, hydrologists, and remote sensing specialists, and even broadening out to other researchers that rely on accurate precipitation, such as climate modelers (Asadieh & Krakauer, 2015). Given the difficulties in detecting rain-on-snow events from both cameras and satellites, in Chapter 6, I analyzed a new method, combining cameras with low-cost passive acoustic recorders (AudioMoths) to detect winter precipitation and differentiate between rain and snow. I demonstrated that combining the audio and visual data can improve differentiation of rain and snow, which is useful for two main applications: 1) improving the ability to distinguish rain and snow by human inspection and 2) improving neural network models. The combination of AudioMoths and camera traps also showed that two tools from wildlife ecology yielded new insights into meteorological observations.

1.3 Implications

Snow is changing worldwide (Hammond et al., 2018), with impacts on wildlife populations (Berteaux et al., 2017; Boelman et al., 2019) and hydrology (Woods, 2009). This research contributes to new tools for ecologists by providing tools for detecting snow cover, depth, and precipitation utilizing tools they already employ. This research also contributes to

new tools for hydrologists, by providing novel sources for snow observations from remote cameras that can offer insight into snow processes that happen at fine-spatial and temporal scales. The models developed through this research set accuracy benchmarks for machine learning applied to snow and will expedite extracting snow information in future studies, reducing labeling and training time. Camera traps have increased in popularity over the last several decades (Rowcliffe & Carbone, 2008; Steenweg et al., 2017), and this work underscores the utility of cameras for environmental monitoring, particularly for snow in cloudy, maritime regions. With the increase in available machine learning models for science and publicly available data repositories for camera trap data (e.g., wildlifeinsights.org), it is evident that we are sitting on a wealth of fine-scale snow information that offers advantages to manually-obtained ground observations (Tuia et al., 2022). This research has developed new methods for extracting snow information from camera imagery, showing that wildlife cameras can operate as ground-based remote sensing networks.

1.4 References

Alison, J., Payne, S., Alexander, J. M., Bjorkman, A. D., Clark, V. R., Gwate, O., et al. (2024). Deep learning to extract the meteorological by-catch of wildlife cameras. *Global Change Biology*, *30*(1), e17078. https://doi.org/10.1111/gcb.17078

Asadieh, B., & Krakauer, N. Y. (2015). Global trends in extreme precipitation: climate models versus observations. *Hydrology and Earth System Sciences*, *19*(2), 877–891. https://doi.org/10.5194/hess-19-877-2015

Bassing, S. B., DeVivo, M., Ganz, T. R., Kertson, B. N., Prugh, L. R., Roussin, T., et al. (2023).

Are we telling the same story? Comparing inferences made from camera trap and telemetry data for wildlife monitoring. *Ecological Applications*, *33*(1). https://doi.org/10.1002/eap.2745

Beery, S., Wu, G., Rathod, V., Votel, R., & Huang, J. (2019). Long Term Temporal Context for Per-Camera Object Detection. *arXiv:1912.03538 [Cs, Eess, q-Bio]*. Retrieved from http://arxiv.org/abs/1912.03538

Berteaux, D., Gauthier, G., Domine, F., Ims, R. A., Lamoureux, S. F., Lévesque, E., & Yoccoz, N. (2017). Effects of changing permafrost and snow conditions on tundra wildlife: critical places and times. *Arctic Science*, *3*(2), 65–90. https://doi.org/10.1139/as-2016-0023

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., et al. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, *14*(1), 010401. https://doi.org/10.1088/1748-9326/aaeec1

Brown, T. B., Hultine, K. R., Steltzer, H., Denny, E. G., Denslow, M. W., Granados, J., et al. (2016). Using phenocams to monitor our changing Earth: toward a global phenocam network. *Frontiers in Ecology and the Environment*, *14*(2), 84–93. https://doi.org/10.1002/fee.1222

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., et al. (2015). REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Carricondo-Sanchez, D., Odden, M., Linnell, J. D. C., & Odden, J. (2017). The range of the

mange: Spatiotemporal patterns of sarcoptic mange in red foxes (Vulpes vulpes) as revealed by camera trapping. *PLOS ONE*, *12*(4), e0176200. https://doi.org/10.1371/journal.pone.0176200

Dietz, A. J., Kuenzer, C., Gessner, U., & Dech, S. (2012). Remote sensing of snow – a review of available methods. *International Journal of Remote Sensing*, *33*(13), 4094–4134. https://doi.org/10.1080/01431161.2011.640964

Esri. (2014). World Shaded Relief map. ArcGIS REST Services Directory. Retrieved from https://services.arcgisonline.com/arcgis/rest/services/World_Shaded_Relief/MapServer

Gaynor, K. M., Hojnowski, C. E., Carter, N. H., & Brashares, J. S. (2018). The influence of human disturbance on wildlife nocturnality. *Science*, *360*(6394), 1232–1235. https://doi.org/10.1126/science.aar7121

Hackländer, K., & Zachos, F. E. (Eds.). (2020). *Handbook of the Mammals of Europe*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65038-8

Hammond, J. C., Saavedra, F. A., & Kampf, S. K. (2018). Global snow zone maps and trends in snow persistence 2001–2016. *International Journal of Climatology*, *38*(12), 4369–4383. https://doi.org/10.1002/joc.5674

Hansen, B. B., Aanes, R., Herfindal, I., Kohler, J., & Sæther, B.-E. (2011). Climate, icing, and wild arctic reindeer: past relationships and future prospects. *Ecology*, *92*(10), 1917–1923. https://doi.org/10.1890/11-0095.1

Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., Andrén, H., et al. (2019). Using by-catch data from wildlife surveys to quantify climatic parameters and the timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, rse2.136. https://doi.org/10.1002/rse2.136

Kelsall, J. P. (1969). Structural Adaptations of Moose and Deer for Snow. *Journal of Mammalogy*, *50*(2), 302. https://doi.org/10.2307/1378347

van de Kerk, M. (2018). Range-wide variation in the effect of spring snow phenology on Dall sheep population dynamics. *Environ. Res. Lett.*, 14.

Langlois, A., Johnson, C.-A., Montpetit, B., Royer, A., Blukacz-Richards, E. A., Neave, E., et al. (2017). Detection of rain-on-snow (ROS) events and ice layer formation using passive microwave radiometry: A context for Peary caribou habitat in the Canadian Arctic. *Remote Sensing of Environment*, *189*, 84–95. https://doi.org/10.1016/j.rse.2016.11.006

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

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

Pan, C. G., Kirchner, P. B., Kimball, J. S., Kim, Y., & Du, J. (2018). Rain-on-snow events in Alaska, their frequency and distribution from satellite observations. *Environmental Research Letters*, *13*(7), 075004. https://doi.org/10.1088/1748-9326/aac9d3

Penczykowski, R. M., Connolly, B. M., & Barton, B. T. (2017). Winter is changing: Trophic interactions under altered snow regimes. *Food Webs*, *13*, 80–91. https://doi.org/10.1016/j.fooweb.2017.02.006

Reinking, A. K., Højlund Pedersen, S., Elder, K., Boelman, N. T., Glass, T. W., Oates, B. A., et

al. (2022). Collaborative wildlife–snow science: Integrating wildlife and snow expertise to improve research and management. *Ecosphere*, *13*(6). https://doi.org/10.1002/ecs2.4094

Rizzi, J., Nilsen, I. B., Stagge, J. H., Gisnås, K., & Tallaksen, L. M. (2018). Five decades of warming: impacts on snow cover in Norway. *Hydrology Research*, *49*(3), 670–688. https://doi.org/10.2166/nh.2017.051

Rowcliffe, J. M., & Carbone, C. (2008). Surveys using camera traps: are we looking to a brighter future? *Animal Conservation*, *11*(3), 185–186. https://doi.org/10.1111/j.1469-1795.2008.00180.x

Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B. H., et al. (2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, *152*, 159–177. https://doi.org/10.1016/j.agrformet.2011.09.009

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., et al. (2017). Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, *15*(1), 26–34. https://doi.org/10.1002/fee.1448

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

Tsang, L., Durand, M., Derksen, C., Barros, A. P., Kang, D.-H., Lievens, H., et al. (2022).

Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing. *The Cryosphere*, *16*(9), 3531–3573. https://doi.org/10.5194/tc-16-3531-2022

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), 792. https://doi.org/10.1038/s41467-022-27980-y

United States, Army, Corps of Engineers. (1956). *Snow hydrology: Summary report of the snow investigations*. North Pacific Division, Corps of Engineers, U.S. Army.

Woods, R. A. (2009). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. https://doi.org/10.1016/j.advwatres.2009.06.011

Chapter 2. EVALUATING MODIS SNOW PRODUCTS USING AN EXTENSIVE WILDLIFE CAMERA NETWORK

Publication history: A version of this chapter was previously published in *Remote Sensing of Environment*. The full citation for this publication is:

Breen, C., Vuyovich, C., Odden, J., Hall, D., & Prugh, L. (2023). Evaluating MODIS snow products using an extensive wildlife camera network. *Remote Sensing of Environment*, *295*, 113648. https://doi.org/10.1016/j.rse.2023.113648

Abstract: Snow covers a maximum of 47 million km^2 of Earth's northern hemisphere each winter and is an important component of the planet's energy balance, hydrology cycles, and ecosystems. Monitoring regional and global snow cover has increased in urgency in recent years due to warming temperatures and declines in snow cover extent. Optical satellite instruments provide large-scale observations of snow cover, but cloud cover and dense forest canopy can reduce accuracy in mapping snow cover. Remote camera networks deployed for wildlife monitoring operate below cloud cover and in forests, representing a virtually untapped source of snow cover observations to supplement satellite observations. Using images from 1181 wildlife cameras deployed by the Norwegian Institute for Nature Research (NINA), we compared snow cover extracted from camera images to Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover products during winter months of 2018–2020. Ordinal snow classifications (scale = 0–4) from cameras were closely related to normalized difference snow index (NDSI) values from the MODIS Terra Snow Cover Daily L3 Global 500 m (MOD10A1) Collection 6

product ($R^2 = 0.70$). Tree canopy cover, the normalized difference vegetation index (NDVI), and image color mode influenced agreement between camera images and MOD10A1 NDSI values. For MOD10A1F, MOD10A1's corresponding cloud-gap filled product, agreement with cloud-gap filled values decreased from 78.5% to 56.4% in the first three days of cloudy periods and stabilized thereafter. Using our camera data as validation, we derived a threshold to create daily binary maps of snow cover from the MOD10A1 product. The threshold corresponding to snow presence was an NDSI value of 40.50, which closely matched a previously defined global binary threshold of 40 using the MOD10A2 8-day product. These analyses demonstrate the utility of camera trap networks for validation of snow cover products from satellite remote sensing, as well as their potential to identify sources of inaccuracy.

Keywords: validation, Norway, remote cameras, gap-filling, MODIS, snow

2.1 Introduction

Seasonal snow covers 31% of the Earth's land surface each year, playing an integral role in habitat quality for wildlife, water storage for hydrological processes, and human uses such as agriculture, forestry, and tourism (Mankin et al., 2015; Rizzi et al., 2018). Warming temperatures have reduced snow cover extent globally, but these changes vary strongly among regions (Brown and Mote, 2009; Solomon et al., 2007). Accurate snow cover mapping within and across years is thus needed to inform regional forecasting and climate change mitigation efforts.

Snow cover is typically measured using ground observations, modeling, and remote sensing at scales that range from point measurements (e.g., ground observations) to kilometers (e.g., passive microwave sensors at 25-km resolution). Remote sensing from satellites is a powerful tool because satellites provide information across broad spatial coverages and at fine

temporal scales, enabling global and regional snow cover maps where *in situ* measurements may not be possible (Nolin, 2010). NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 6 product provides a daily or every other day 500-m resolution optical image from which snow maps are derived. Daily MODIS snow observations are highly suitable for continuous snow monitoring, which is desirable for many applications, including wildlife science (Boelman et al., 2019). For example, daily MODIS snow maps have been used to successfully detect changes in bird nesting success and shifts in the timing of mammal migrations (Laforge et al., 2021; Madsen et al., 2007).

The most recent version of the MODIS products (Collection 6.1) includes a daily 500-m global snow product, MOD10A1, and a daily cloud-gap filled (CGF) 500-m global snow product, MOD10A1F. Both are suitable for use as inputs in hydrological, ecological, and climate models (Bokhorst, 2016; Dong and Menzel, 2016). MOD10A1 and MOD10A1F provide normalized difference snow index (NDSI) values based on the high reflectance of snow in the visible band and low reflectance in the near-infrared band, ranging from 0 (snow-free) to 100 (completely snow-covered). NDSI values lower than 100 can be completely snow-covered (Klein et al., 1998), but adjusting NDSI values to a fractional snow cover is no longer included in MODIS products as it is region-dependent and other factors may affect when MODIS underestimates snow. The overall accuracy of the MOD10A1 product is estimated to fall between 79.5- 96% depending on the tree cover density, snow depth, and solar zenith angle in the region of interest (Coll and Li, 2018; Franklin, 2020; Hall et al., 2019a; Hall and Riggs, 2007). Optical sensors are obstructed by tree cover, and shallow snow might not be bright enough to reflect solar radiation since the underlying material is likely to be darker (Liang et al., 2008). At high solar zenith angles, chances are higher that sensors will be obstructed by clouds

and experience higher atmospheric distortion (Xin et al., 2012), both of which can also obscure or scatter light, decreasing the accuracy of observations.

Cloud masking in MOD10A1 greatly reduces coverage (Hall et al., 2019a), and MOD10A1F improves coverage by filling all cloud-masked pixels. Each cloud-masked pixel is given the most recently observed snow cover value, along with a corresponding "cloud persistence" value for the age in days of the snow observation. This product has been used in applications such as hydrological snow trend studies (Hao et al., 2022) and analyses of snow cover impacts on wildlife (Mahoney et al., 2018). The cloud-gap filled product has been shown to return similar accuracy as MOD10A1 in the western US where cloudy periods are typically brief (Hall et al., 2019a), whereas accuracy is lower in the northeastern and northwestern US where longer cloudy periods are common (Gao et al., 2011; Hall et al., 2010). Beyond the US, validation of the MOD10A1F product is sparse due to the recency of the product availability. Weather stations and other sensors improved MOD10A1F maps in China (Hao et al., 2022), but more work in diverse areas with longer cloudy episodes, such as high latitude regions, is needed to understand the accuracy of the MOD10A1F product in those areas. Understanding accuracy may inform a region-dependent threshold after which additional cloudy days may result in unreliable snow cover estimates, and indicate when alternative sources for snow cover, such as weather stations or other ground observations, should be used instead of gap-filled values.

Binary products can be developed from the current MODIS snow-cover products and may be used to map snow presence/absence. Early MODIS snow-cover products categorized pixels as "snow" if the NDSI was greater than 40, using Landsat fractional snow-covered area maps from Prince Albert National Park in Saskatchewan, Canada (Klein et al., 1998). Later, a binary map developed from MOD10A2 categorized a pixel as "snow" if any pixel within an 8-

day period had an NDSI value greater than 10 (Hall et al., 2002). The lower threshold increased snow detection but at the cost of increased false positives. Now, the threshold for snow presence is considered region dependent (Thapa et al., 2019; Zhang et al., 2019), and the end-user is recommended to determine the threshold above which the corresponding pixel should be identified as snow covered (Riggs et al., 2017). Given the utility of binary snow products for monitoring snow phenology and subsequent applicability to wildlife studies (Curk et al., 2020; Madsen et al., 2007; Thapa et al., 2019), more work is needed to develop daily binary snow maps for specific regions.

In this study, we use cameras deployed in remote locations for wildlife monitoring, often referred to as "camera traps," to evaluate the MODIS/Terra MOD10A1 and MOD10A1F products and derive a regional threshold for daily binary snow-covered maps in Scandinavia. Wildlife camera trap networks have underutilized potential for satellite validation that could be a valuable supplement to traditional validation methods based on other satellites (Crawford, 2015; Huang et al., 2011), weather stations, and ground collection (Negi et al., 2007). Cameras provide environmental monitoring (Brown et al., 2016; Sonnentag et al., 2012), with visual information about environmental conditions along with a timestamp. While single cameras have a limited field of view, they can be set up in networks of up to many hundreds of cameras across large regions (Forrester et al., 2016; Garvelmann et al., 2013). Databases are increasingly available to archive camera images across networks, furthering the potential for global camera networks to improve environmental monitoring (Steenweg et al., 2017). For example, Wildlife Insights currently hosts over 35 million images from 20,000 camera deployments worldwide (https://www.wildlifeinsights.org/home). Cameras operate below cloud cover and tree canopy,

and they are particularly advantageous for monitoring snow cover because they can operate for months or years in sub-freezing conditions and difficult-to-reach locations (Tobler et al., 2015).

Wildlife camera traps have been used successfully to evaluate satellite measures of greenness (Sun et al., 2021) and to provide information on snowpack dynamics at localized spatial scales (Hofmeester et al., 2021; Sirén et al., 2018). Hofmeester et al. (2021) visually categorized snow cover from camera trap images to assess changes in spring and autumn molting of mountain hare (*Lepus timidus*). Sirén et al. (2018) found strong correlations between depth readings on snow poles and data from the Snow Data Assimilation System (SNODAS) at 80 cameras in Vermont. However, extracting information from camera images can be challenging. Camera traps use an infrared flash in low light settings, resulting in grey-scale images that can make differentiating among objects more difficult (Beery et al., 2020). Camera traps therefore have great potential but require more work investigating their utility as ground-based remote sensing networks for monitoring snow at broader scales.

Using three years of camera trap images from a network of 1,181 cameras in Norway and Sweden managed by the Norwegian Institute for Nature Research (NINA), we compared snow data extracted from camera images to MOD10A1 and MOD10A1F NDSI snow cover products. We quantified agreement between snow cover values from cameras and MODIS NDSI, examining factors we hypothesized *a priori* would affect agreement. We predicted the following:

1. Agreement would be higher between cameras and NDSI at extreme values for snow cover, whereas agreement would be lower when the snow is patchy (i.e., moderate NDSI values) due to differences in scales between MODIS pixels (500 m) and camera fields of view (~20-80 m).

2. Factors that have been shown to affect MODIS accuracy will affect camera and MODIS agreement, such that agreement will be lower when canopy cover and latitude are higher (Xiao et al., 2022; Xin et al., 2012).
3. Factors that have been shown to affect image quality will affect camera and MODIS agreement, such that images with low lighting taken in grey-scale (i.e., with infrared flash) will have lower agreement with NDSI than images taken in full color.
4. Camera observations should agree more with MODIS observations on clear sky days compared to cloudy days, and cloud persistence should decrease the agreement between cameras and the cloud gap filled NDSI product.

We derived a binary MOD10A1 product of snow cover, using camera data to identify a NDSI threshold corresponding to snow presence.

2.2 Methods

2.2.1 *Study Area*

We used images from camera traps in the Scandcam network. Scandcam is a long-term, year-round study established in 2010 by the Norwegian Institute for Nature Research to monitor recovering Eurasian lynx (*Lynx lynx*). Our dataset includes images from three winter seasons: 1) January 1 – March 2018, 2) October 1, 2018 – March 2019, and 3) October 2019 – March 2020. Scandcam camera trap locations are optimized for lynx detection across Norway and southern Sweden (59° – 69° N, 8° – 16° E), with no more than one camera per 2 km^2 area across a 350,000 km^2 area (Fig. 2.1; Carricondo-Sanchez et al., 2017). Cameras span a 10° latitudinal gradient, with deeper snow generally occurring in the north and inland than along the coast (Saloranta, 2012). Snow usually arrives in Norway and Sweden in early October at high

elevations and northern areas and melts by early April, although sites farther north can remain snow-covered into summer (Saloranta, 2012). Because the cameras were deployed to detect lynx, they were placed in lynx habitat such as forests and sub-alpine areas, but they varied in whether they were under closed-canopy or open-canopy areas. Southern Norway and Sweden are characterized by boreal coniferous forest dominated by Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*). In the north, forest composition transitions to alpine vegetation dominated by birch species (*Betula pendula* and *Betula pubescens)* (Bouyer et al., 2015).

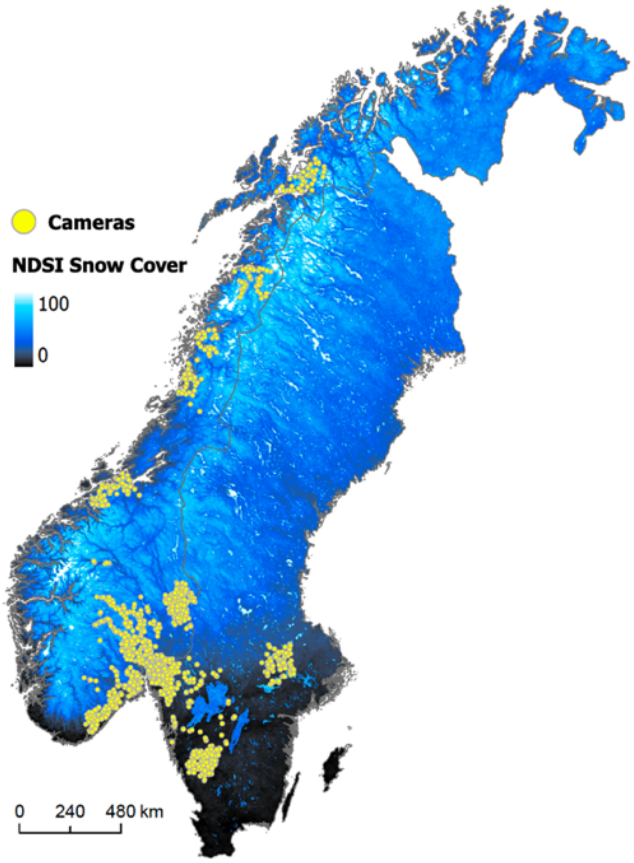


Figure 2.1 Locations of Scandcam cameras (yellow points, n = 1,181) in Norway and Sweden shown over a composite snow cover map created from MOD10A1 Version 6 that shows mean NDSI snow cover values across the three winters of this study (January – March 2018, October 2018 – April 2019, October 2019 – April 2020).

2.2.2 *Data*

2.2.2.1 MODIS Data

NDSI values were extracted at all camera locations (n = 1,181) for all days in the study period from the MOD10A1 product on the Google Earth Engine public data archive (n = 770 days; Hall et al., 2016). To quantify the percentage of usable MOD10A1 NDSI values during our study period, we divided the number of non-null NDSI values by the total number of values (including cloud-masked pixels with "NA" values). The MOD10A1F product was downloaded as GeoTiffs for the same days from the EarthData platform (https://search.earthdata.nasa.gov/) and uploaded to Google Earth Engine (GEE). MOD10A1F NDSI and corresponding cloud persistence values were extracted for all cameras for the same days after ensuring both MODIS products matched projections (Appendix A1). Since MOD10A1F is not offered in the GEE archive, MOD10A1F was uploaded as individual tiles. In total, we processed 3,392 tiles for the MOD10A1F product. We used the GEE Collection 6 MOD10A1 product rather than Collection 6.1 from EarthData, because Collection 6 is commonly used in other studies, and GEE has limits on the number of original assets one can store on the server. We used Collection 6.1 for MOD10A1F (cloud-gap filled) NDSI and cloud persistence products to make use of the most up-to-date version. Previous work has demonstrated that Collection 6.1 and Collection 6 have 99% correspondence, with revisions considered minor (Riggs et al., 2019).

2.2.2.2 Camera Images

Cameras with either infrared flash or white flash (Reconyx model HC500, HC600, PC800, or PC900) were secured to trees approximately 1 m above the ground. Cameras were programmed to take a daily "timelapse" image at 8 AM or 12 AM, as well as anytime the camera

was triggered by motion (e.g., from an animal walking by). For all cameras, for every day in our study years that there was a corresponding non-null MOD10A1 value, we selected one image per day from the Scandcam image inventory. The vast majority of photos were taken under low-light conditions resulting in a grey-scale image. To achieve a more balanced dataset to evaluate the effect of image color mode on snow labeling accuracy, we manually inspected all of the images to select a color image if available. We deferred to the timelapse image when taken in white flash or daylight hours, or a daytime motion-triggered image, when available. Images from the prior day or the day after the image of interest were also inspected and labeled if it was hard to discern snow due to lighting conditions.

To assess the effects of cloud cover, we labeled a subset of images that corresponded to 250 random days from the MOD10A1F product (100 days for each of the full winter seasons, and 50 days for the partial season). Additionally, we included images that were inspected while labeling MOD10A1 images (n = 510 images). These included both before and after images corresponding to MOD10A1 values to help confirm the amount of snow. While there is potentially a bias that these images would favor lower cloud persistence values, we examined a histogram and found a similar distribution of cloud persistence values compared to the distribution of cloud persistence values from the full MOD10A1F dataset (Appendix A, Figure A2).

Images were labeled using Timelapse (http://saul.cpsc.ucalgary.ca/time-lapse/), a freely available camera trap labeling software for wildlife ecologists. The software automatically extracts metadata including time and date, and it provides a customizable interface that observers use to label photos. All data can then be exported as a .csv file. Snow cover was manually labelled using the software's user interface on an ordinal scale that ranged from 0 (no snow) to 4

(full snow coverage). These categories matched those used for snow cover classification at Norwegian weather stations (Lussana et al., 2018): 0 corresponded to 0% snow cover, 1 to ~25% snow cover, 2 to ~50%, 3 to ~75%, and 4 to ~100% (Fig. 2.2). Images were initially labeled by two people, but testing of a double-labeled subset revealed low agreement among observers (kappa coefficient $\kappa = 0.45$; McHugh, 2012). There was complete agreement at label 0, moderate agreement for values 1 – 3 ($\kappa = 0.51$) and low agreement for label 4 ($\kappa = 0.10$). The low agreement at label 4 was a result of the less-experienced labeler incorrectly labelling low-light images with snow as "no snow." Thus, the more-experienced observer (C. B.) labeled all images.

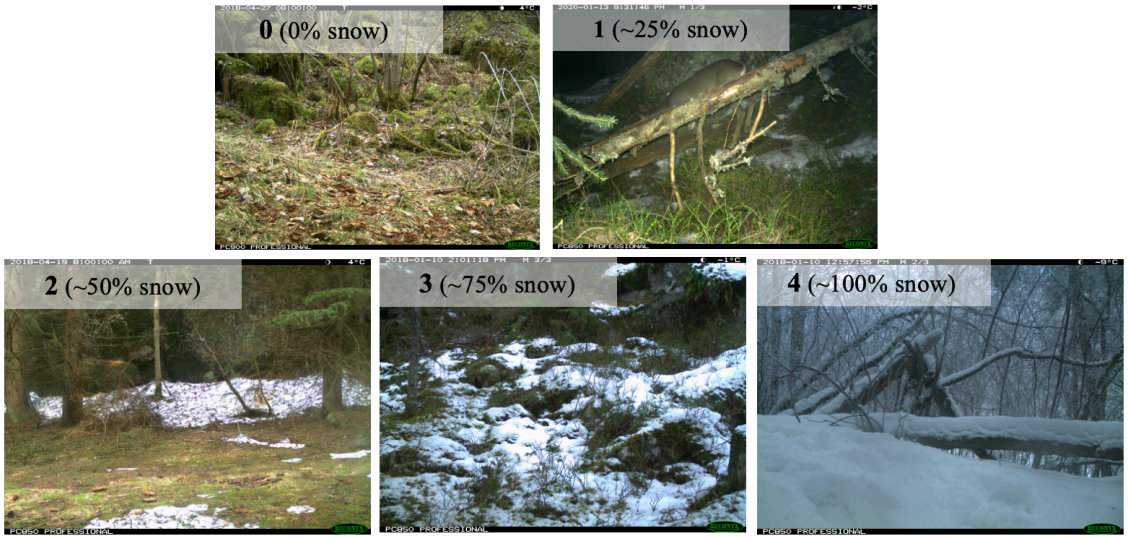


Figure 2.2. Example remote camera images for snow classification. Snow cover was classified using an ordinal scale from 0 – 4, where 0 = 0% snow cover, 1 = ~25%, 2 = ~50%, 3 = ~75%, and 4 = ~100%.

2.2.3 *Assessing Agreement between Camera Images and MODIS Snow Values*

To evaluate the relationship between image labels and MOD10A1 (H1), we fit a general linear model using the ordinal image labels as a continuous predictor variable and MOD10A1

NDSI as the response variable. Since NDSI values have been noted to "plateau" at higher snow values depending on the normalized difference vegetation index (NDVI) at that pixel (Klein et al., 1998), a polynomial term was included to account for potential non-linearity. All models were fit using program R (version 4.2.1).

To test our prediction that agreement between MODIS and images would be highest at extreme values (H1), we compared agreement between MODIS NDSI snow cover values and snow cover from labeled camera images (hereafter called "image labels") across the ordinal image labels. We calculated agreement as:

$$Agreement = 100 - |MODIS - Camera| \quad (2.1)$$

Where *MODIS* is the NDSI value and *Camera* is the labeled image value after converting ordinal labels (0-4) to their corresponding percent cover values (0, 25, 50, 75, and 100). Agreement could range from 0 (i.e., complete disagreement) to 100 (i.e., complete agreement). Some amount of disagreement was expected from comparing ordinal image labels to continuous NDSI values. Thus, we caution that agreement levels should not be compared directly to R^2 values from traditional validations. Other studies that assessed MODIS NDSI accuracy using cameras and other ground sources converted NDSI values to binary snow and no snow values using a threshold and confusion matrix (Thapa et al., 2019; Zhang et al., 2019). We made use of the full range of NDSI values by not thresholding the values for agreement assessment, in order to statistically assess covariates that affected the level of agreement. We equate NDSI to a scale of 0-100% snow cover to represent the relationship between NDSI and snow cover in the absence of factors that may affect satellite accuracy. Taking the absolute value of agreement allowed for clearer interpretation of how different covariates affected the magnitude of disagreement regardless of its direction (see 2.4). We expected agreement to be highest at the

extremes (i.e., labels ~0% and ~100%) and lowest for intermediate labels (i.e., labels ~25%, 50%, and 75%), so we fit a linear model with a polynomial term to allow for a parabolic shape.

2.2.4. *Assessing Agreement between MODIS Snow Products and Factors Influencing Agreement*

To identify factors affecting agreement between snow cover from image labels and the MOD10A1 product (H2 and H3), we used a general linear mixed-effects model to determine how tree canopy cover, latitude (a proxy for solar zenith), and image color mode affected the agreement between image labels and MOD10A1 NDSI values (Table 2.1; Eqn. 2). We first

Table 2.1. Covariates used to analyze agreement between MODIS and image-labeled snow values. Range of each factor is provided. MODIS cloud persistence values were only used to assess MOD10A1F (i.e., the cloud-gap filled product) agreement with camera images.

Covariate	Range	Resolution	Hypothesized effect on agreement
Daily MODIS NDVI	-1.0 – 1.0	500 m	Increasing vegetation will prevent MODIS obs., decreasing agreement with ground obs.
Landsat tree canopy cover	0 – 100%	30 m	Increasing tree canopy cover will prevent MODIS obs., decreasing agreement with ground obs.
Image color mode	Color (1) or Grey-scale (0)	20-30 m[1]	The infrared red flash will decrease the saturation of the image (converting it to grey-scale), increasing the difficulty of differentiating snow from other aspects of the landscape.
Latitude	59.0 – 69.0	1 degree	Increasing latitude increases angle of MODIS obs., increasing angular distortion and decreasing agreement with ground obs.
MODIS Cloud Persistence	0 – 40 days	500 m	Increasing cloud cover days increases possibility of missed accumulation or melt events, decreasing agreement with ground obs.

[1] Resolution derived from the approximate range that wildlife cameras detect (Urbanek et al., 2019).

tested covariates for correlation to avoid overfitting the model. We used Pearson's method for correlation between continuous variables and Kendall's method for correlation between continuous and our categorical variable (i.e., image color mode) and found that all correlations were below the commonly-used threshold of 0.7 (Dormann et al., 2013; Appendix A, Table A3). All correlations were also below the threshold for moderate correlations ($|r| = 0.4$), except for tree canopy cover and latitude, which was -0.404. To further examine multicollinearity among predictors, we implemented the variance inflation factor (VIF) test. All factors were below 1.2, lower than the conservative threshold of 3 (Zuur et al., 2010; Appendix A, Figure A4). Temporal and spatial autocorrelation in snow datasets can inflate parameter estimates and type 1 error (Reinking et al., 2022). To evaluate spatial autocorrelation, we conducted Moran's I test using the *spdep* package in R (Bivand, 2022). We failed to detect spatial autocorrelation (Moran's I statistic = -0.007, $p = 0.55$), but we included Camera ID as a random effect to account for lack of independence among images taken from the same camera. To test for temporal autocorrelation, we followed the approach of Sirén et al. (2018) and created a relative date variable for each observation using the *timeDate* package in R (Wuertz et al., 2023). The package contains a function to convert a date to a relative number of days from a specified origin, defaulting to January 1, 1970. We tested for improved model fit using Akaike Information Criterion (AIC) values with and without including the relative date in an auto-regressive correlation structure (i.e., an "*ar1*" term) with camera station ID included as a grouping variable. Incorporating the *ar1* correlation structure had a lower AIC score [(ΔAIC = -1830.2 compared to the model without a correlation structure]. We therefore proceeded to use this structure for modeling agreement in Eqn. (2). We included all covariates in a general linear mixed effects model with a Gaussian family using the *glmmTMB* package in R (Brooks et al., 2023):

$$Agreement \sim (1 \mid Camera\ ID) + daily\ NDVI + Tree\ Canopy\ Cover + Latitude + Image\ Color\ Mode + ar1(relative\ date + 0 \mid Camera\ ID) \quad (2.2)$$

Agreement was calculated as described above in Eqn. (1). Image color mode was classified as "Color" or "Grey-scale" by inspecting image saturation. Images taken with infrared flash have low light saturation and appear as black-and-white, grey-scale images (Fig. 2.3). After inspecting a histogram of saturation values from a subset of 60 images, there was a clear break between images in grey-scale and color at saturation values of 0.02 (Appendix A, Table A5). We then evaluated this threshold using a random subset of 1,000 images and found 100% accuracy, so we labeled all images with values below 0.02 as grey-scale and above 0.02 as color.

Figure 2.3. A grey-scale and color image from the camera on 22 November 2018 illustrates how light saturation affects the ability of an observer to identify snow cover. The image on the left was the daily timelapse photo taken at 08:00h during low light conditions, which triggered the camera to take the image in grey-scale (i.e., with infrared flash). The image on the right was triggered by a wolf (*Canis lupus*) passing by at 14:03h, when there was enough light for a color image. The amount of snow in the color image is much easier to see.

Previous studies found that dense forests affected MODIS NDSI by causing an underestimation of the snow cover, using daily NDVI as a proxy for forest canopy (Hall and Riggs, 2007; Klein et al., 1998). MODIS NDVI is a vegetation index that provides information on vegetation canopy greenness, along with leaf area, and chlorophyll and canopy structure (Didan, 2015). NDVI in Norway varies spatially due to differences in vegetation from boreal, deciduous trees in southern Norway to alpine shrubs in northern Norway. Within a winter season, NDVI is highest in October and November and lowest in February and March, likely reflecting both deciduous trees losing canopy leaves in the fall, and seasonal snow covering ground vegetation in January to March (Appendix A, Table A6 & Fig. A3). To test the efficacy of NDVI as a proxy for tree canopy cover, we extracted the corresponding daily MODIS NDVI value at 500-m for each labeled image. We also extracted tree canopy cover from the 30-m Landsat Vegetation Continuous Fields tree cover layer, which estimates the percentage of horizontal ground covered by woody vegetation greater than 5 meters in height from 2015 (Townshend, 2016). Continuous predictor variables – tree canopy cover, latitude, and NDVI – were normalized by subtracting by the mean and dividing by the standard deviation. Model fit was evaluated by examining residuals for dispersion and outliers from the *DHARMa* package in R (Hartig, 2022; Appendix A, Table A8).

To test our prediction that agreement between MODIS and camera data would decline as the number of cloudy days (i.e., cloud persistence) increased (H4), we modeled the agreement between snow cover from image labels and the MOD10A1F product as a function of the cloud persistence value. Because we expected the relationship between agreement and the number of cloudy days to be non-linear, we ran a generalized additive mixed model with camera ID included as a random effect using the *mgcv* package in R (Eqn. 3; Wood, 2017). We selected

eight knots for the model, following recommendations for knots to be larger than the degrees of freedom (i.e., 6) plus 1 (Wood, 2017). Cloud persistence values equal to 0 (MOD10A1 values) were included to allow agreement comparison to clear sky days.

$$Agreement \sim (1 \mid Camera\ ID) + MOD10A1F\ Cloud\ Persistence \quad (2.3)$$

Agreement was calculated as described in 2.3 (Eqn. 1). Data was sparse for persistence times > two weeks, (3% of data), so we limited analysis to 14 days.

2.2.5 *Deriving a Threshold for Daily Binary Snow Mapping in Norway*

Image labels were converted from the 5-class ordinal scale to a binary classification by reclassifying all images labeled 1 – 4 as "snow" (with a corresponding 1 label), and all image labeled with a 0 as "no snow" (with a corresponding 0 label). We identified an optimal threshold for the MOD10A1 product by counting the number of true positives and false positives when converting to a binary product at each NDSI value. We plotted the true positive rate against the false positive rate at each threshold value to create a receiver-operating-characteristic (ROC) curve using the *pROC* package in R (Robin et al., 2011). The top left corner of the ROC curve is known as Youden's Index, or the maximum difference between the true positive and false positive rate (Youden, 1950). Because it weighs both true positive and false positive rates equally, it is considered the optimum threshold for a classifier when there is equal preference for both classes (Liu, 2012). In addition to generating a threshold for all cameras, we repeated this analysis separately for cameras within closed canopy (> 20% canopy cover; n = 6,229 images) and open canopy ($\leq$ 20% cover; n = 2,731 images) because thresholds tend to be lower in areas of closed canopy cover (Chokmani et al., 2010).

2.3. Results

2.3.1. *Labeled Image and MODIS Comparisons*

Of the 1,703,702 MOD10A1 snow cover values obtained at all 1,181 cameras during winters 2018 – 2020, 1,311,249 (76%) were null (cloud-masked). Daily labeled images corresponding to non-null values from MOD10A1 spanned 665 cameras (n = 8,918 images). Cameras not included either had no corresponding non-null MODIS value or did not have images on file during our study period. There was a strong correlation between snow classification from the images and MOD10A1 NDSI values (R^2 = 0.70, NDSI = -3.50**image*^{2} + 28.02**image* + 10.90, where *image* is the labeled value on the 0 - 4 scale), but the NDSI values from MODIS products plateaued at about 75 (Fig. 2.4A). We found overall strong agreement between snow cover from MODIS NDSI and camera images ($\bar{x}$ = 80.5%, 95% CI = 80.1 – 80.8; Fig. 2.4B). Consistent with H1, agreement was highest for images with label 0 (corresponding to ~0% snow cover; agreement $\bar{x}$ = 89.2%, 95% CI = 88.6 – 89.8). Contrary to H1, however, agreement was lowest for images with label 4 (corresponding to ~100% snow cover; agreement $\bar{x}$ = 67.1%, 95% CI = 66.7 – 67.5, Fig. 2.4B).

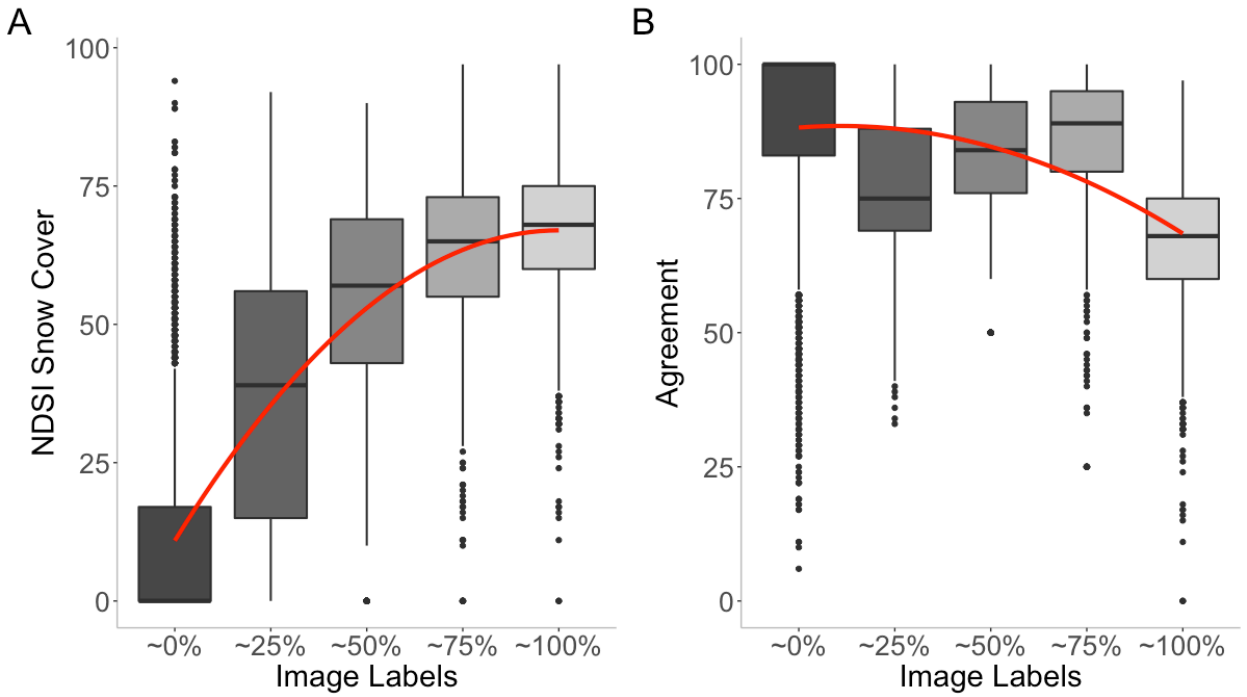


Figure 2.4. A) Distribution of MOD10A1 NDSI values within each snow cover classification from labeled camera images, and B) agreement of snow cover values between MODIS and images within each snow cover classification. Images were labeled using an ordinal classification with 5 levels (0 – 4) corresponding to snow cover percentages shown. Agreement was defined as 100 minus the absolute difference between the image label and MOD10A1 NDSI snow value. Red lines show the best fit using linear models with polynomial terms.

2.3.2 *Factors that influence agreement between cameras and MODIS*

As predicted by H3, latitude and tree canopy cover negatively affected agreement between snow cover derived from cameras and MOD10A1. However, only canopy cover had a statistically significant effect (Table 2.2). Although significant, the effect was relatively weak, and mapping the agreement at each camera relative to tree canopy cover showed that agreement was high in many areas with closed canopies (Fig. 2.5A-D). Contrary to expectations, NDVI was

not strongly correlated with tree canopy cover (r = 0.09, Appendix, Table A9) and had a significant positive effect on agreement: image labels and MODIS-derived snow cover were in better agreement in areas with higher daily NDVI. Average NDVI values in October were twice as high as any other month (Appendix A, Table A6), and October likewise had a relatively high proportion of 0 values with high agreement (Fig. 2.4B). Thus, we examined the effect of removing October observations from our model and found the effect of NDVI on agreement changed from strongly positive (coefficient value = 6.60) to weakly negative (coefficient = -0.075; Appendix A, Table A7). Our dataset was roughly split between color (n = 4,184 images) and grey-scale (n = 4,733 images), and image color mode positively affected agreement as predicted by H4 (Table 2.2).

Table 2.2 Coefficient estimates, standard error (SE), t-values, and *p*-values from a general linear mixed model assessing factors that affect MODIS and camera agreement (n = 8,808) for the three winter seasons: 1) January 1 - March 2018, 2) October 1, 2018 – March 2019, and 3) October 2019 – March 2020. Continuous variables were normalized by subtracting the mean and dividing by the standard deviation prior to analysis. Image color mode is a categorical variable (1: color image; 0: grey-scale image). Camera identification was included as a random effect (n = 658). Model results without observations from October 2018 and October 2019 are included in Appendix A, Table A7. Results from the model without October data are similar, except that the effect size of NDVI changes from strongly positive to weakly negative.

Parameter	Estimate	SE	t-value	*p*-value
Intercept	78.88	0.40	196.37	<0.005
Latitude	-0.48	0.35	-1.37	0.17
NDVI	6.60	0.35	29.12	<0.005
Tree canopy cover	-0.93	0.32	-2.84	<0.005
Image color mode (color image)	1.73	0.47	3.63	<0.005

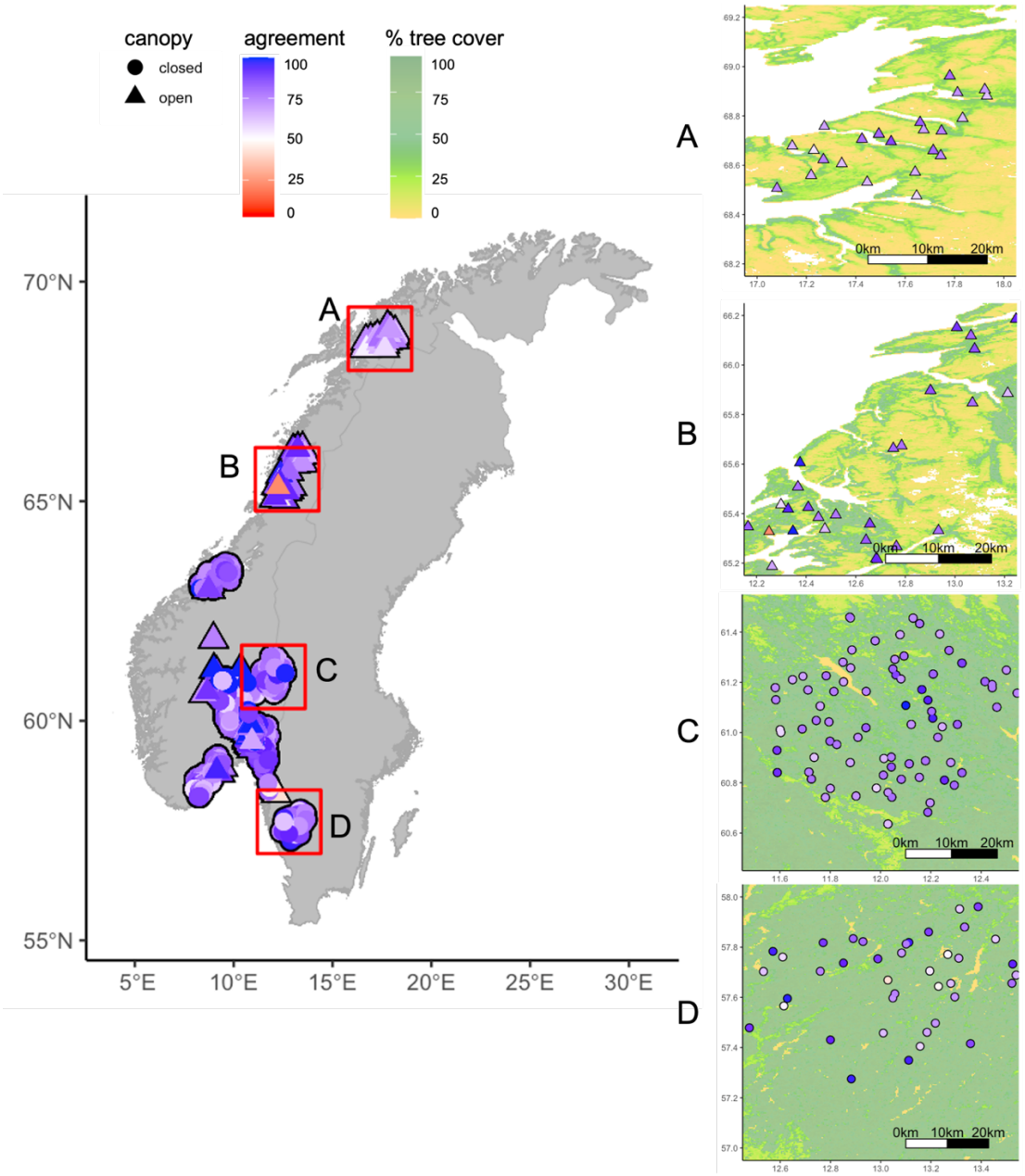


Figure 2.5 Average agreement between snow cover from labeled images and MOD10A1 snow cover at Scandcam cameras between winter months for 2018 – 2020. The four boxes correspond to four example clusters in counties from north to south: A) north Nordland and Troms og Finnmark; B) south Nordland; C) Innlandet; and D) south Viken. The base map is tree canopy cover from 30-m Landsat. Triangles represent cameras within closed canopy areas (≥ 20%) and circles represent cameras within open canopy areas (< 20%).

2.3.3 *Image Labels and MOD10A1F Product Comparison*

Cloud persistence was a significant predictor for agreement between image labels and snow values from the MOD10A1F product. Agreement was highest (78.5%) on clear sky days (i.e. cloud persistence = 0) and decreased by almost one third (to 56.4%) within the first 3 days before leveling off just after (Fig. 2.6).

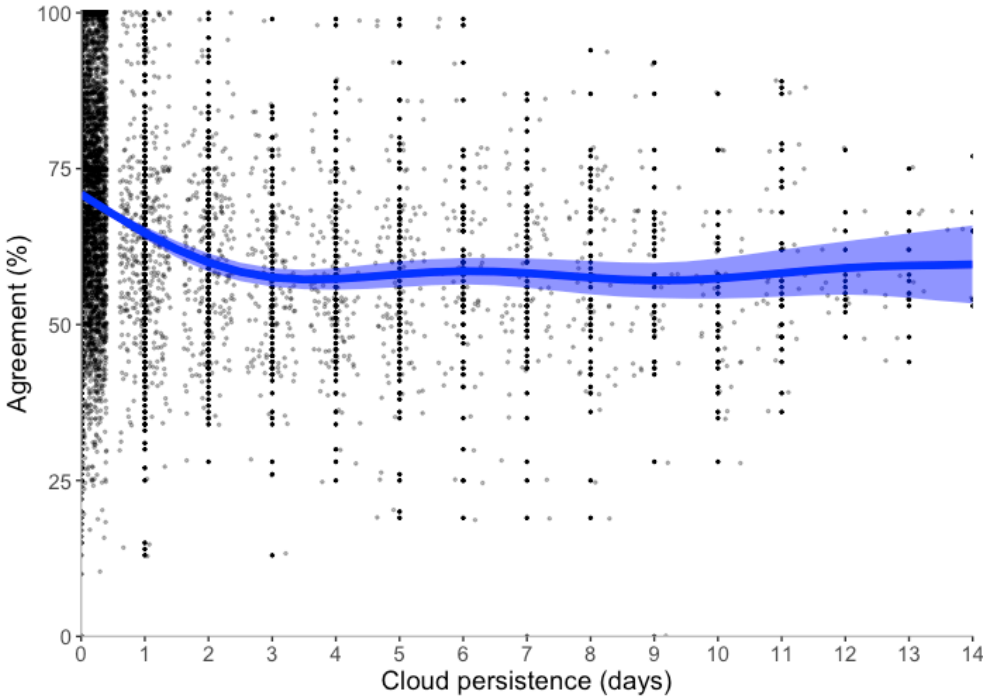


Figure 2.6. Agreement between image labels and MOD10A1F NDSI snow values as a function of number of cloudy days (i.e., cloud persistence) using a generalized additive model. Agreement was defined as 100 minus the absolute difference between the image label and MOD10A1F NDSI snow value. A jitter was applied (alpha = 0.3) to show data more easily.

2.3.4. *Optimal Threshold Derivation for Binary Snow Cover Mapping*

At the Youden's Index point of the ROC curve, the true positive rate was 88% and the false positive rate was 11%. This point corresponded to a MOD10A1 NDSI snow cover value of 40.5

(Fig. 2.7). At the commonly used threshold value of 40 (Hall et al., 2019a), the true positive rate was 89% and the false positive rate was 11%, showing that for a slightly higher true positive rate, there is not much difference in the false positive rate. The current MOD10A2 product employs a threshold of 10, which has a 97% true positive rate and 31% false negative rate. When

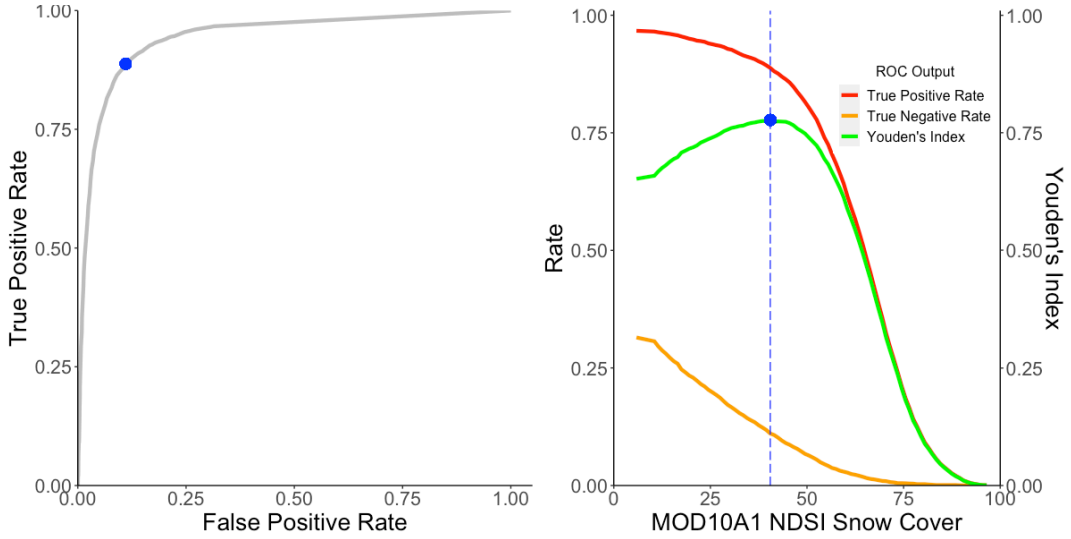


Figure 2.7: A) A Receiver-Operator Characteristic (ROC) curve when images are reclassified for snow or no-snow by cutting the data with a label >=1 as 'snow.' The ROC curve shows the performance of the classifier at each threshold, in this case the value of the NDSI snow cover. The closer the curve is to the top left corner, the better the performance of the model. The blue point closest to the top left corner is (0.11, 0.88) is referred to as Youden's Index. B) The true negative rate (orange) and the true positive rate (red) graphed separately for every MOD10A1 NDSI snow cover value alongside the Youden's Index, the difference in between (green). The MOD10A1 value at the maximum value of the Youden index is 40.50. The maximum value of the Youden index is the minimum between the true positive rate and true negative rate when both classes are given equal weight. The blue points on both graphs represent the same cut point in the data.

we conducted separate analyses for closed canopy (≥ 20%) and open canopy (< 20%) sites, the threshold was the same for closed canopy locations (40.5) and slightly higher for open canopy locations (41.5). Appendix A8 shows the change in true positive and false positive rates with different threshold values, along with the results for open and closed canopy analyses.

2.4 Discussion

In this study, we identified strong agreement between snow information obtained from wildlife cameras and MODIS at a regional scale, demonstrating the ability of cameras to supplement MODIS snow observations. Previous studies have found strong agreement using fewer than 100 cameras in tandem with satellites at localized spatial scales (Raleigh et al., 2013; Sugiura et al., 2013), and our findings show this relationship holds across a large region and multiple winter seasons. As predicted, we found strongest agreement at low snow cover values, but agreement was worse than expected at high snow cover values because NDSI values plateaued around 75 instead of 100. Cameras, thus, demonstrated that an NDSI value of 75 represents 100% snow cover for this region. We also demonstrated the ability to customize MOD10A1 to create binary snow maps using a camera-derived threshold of 40.5, which was nearly identical to the commonly used 40 threshold from previous MODIS products (Klein et al., 1998). These findings highlight that despite large differences in scales, wildlife camera networks have potential to improve satellite monitoring for snow and create new products at fine temporal scales.

Our finding of strong agreement between camera image snow values and MODIS snow values may be attributed in part to our method of classifying snow cover into five classes. While not a continuous measure, a 5-class ordinal labeling scheme for images extracts more

information about the amount of snow cover than previous work using binary labels (Berman et al., 2018; Sugiura et al., 2013). There are caveats to this classification scheme, as agreement was lower at labels 1 (i.e., ~25% snow cover), 2 (i.e., ~50%), and 4 (i.e., ~100%), which may highlight MODIS uncertainties. For example, MODIS is less accurate when snow is thin or patchy, such as labels 1 and 2 (Berman et al., 2018; Dong and Menzel, 2016). Similarly, low agreement at label 4 highlights the tendency of MODIS to underestimate snow cover in boreal regions (Klein et al., 1998). In this region, maximum NDSI equating to 100% snow cover appears to be 75, when MODIS plateaus. Cameras can thus be used to adjust NDSI for fractional snow-covered maps. However, discrepancies in agreement at different classification schemes highlights drawbacks to using cameras in tandem with satellite products: patchy snow in cameras may be missed or interpreted as complete snow cover. Furthermore, labeling for snow cover values can be subjective and have uncertainty, as highlighted by low correspondence among labels by two observers found during pilot testing. We recommend a single, experienced labeler when labeling wildlife photos, and testing for agreement among labelers early on. Despite lower agreement within certain classes and among observers, strong agreement overall suggests that cameras can be an effective method of snow classification when used in tandem with satellites.

We predicted that latitude (i.e., solar zenith angles), ground vegetation, and the image color mode (i.e., grey or color scale) would limit MODIS and camera image agreement (Xiao et al., 2022; Xin et al., 2012). Although mapping the agreement at each of our cameras showed a general decline in agreement as latitude increased, the effect was not significant, and there was still strong agreement even at high latitudes. Overall, our findings indicated that latitude and canopy cover had relatively minor effects on the accuracy of MODIS snow cover, highlighting its robustness for monitoring snow trends across Scandinavia. Images in grey-scale had lower

overall agreement with satellites, and they took much longer to label due to the need to study the image more carefully to separate snow from vegetation and rocks. Humans and artificial intelligence have more difficulty extracting information about environmental conditions and wildlife from grey-scale images (Beery et al., 2019; Favorskaya and Buryachenko, 2019). Nighttime images are inevitable when using motion-triggered wildlife cameras for environmental monitoring, but we recommend maximizing the number of color images either through prioritizing color photos as we did here, or by scheduling timelapse photos to occur during daylight hours. Because low-light images were also the main reason why images from one labeler had to be relabeled, prioritizing color photos may increase both agreement between camera and satellite as well as agreement among labelers.

Using cameras to assess agreement demonstrated drawbacks of using NDVI alongside MODIS NDSI. Contrary to our hypothesis, NDVI positively affected the agreement. While daily NDVI is often included to account for the effects of vegetation on MODIS snow detection (Hall et al., 2002; Klein et al., 1998; Xin et al., 2012), NDVI has multiple interpretations, including green-up, biomass, and plant stress (Huang et al., 2021). The positive effect of NDVI on snow cover agreement suggests that daily NDVI during winter may not have represented vegetation that was obscuring the sensor, but rather the absence of snow. We included snow values ranging from 0 – 100, but values equal to 0 for both camera images and MODIS will have exact agreement whereas our estimates for the other snow labels were approximations. When we excluded images from the month October, the month that also has the highest average NDVI at the camera locations, we found the expected negative relationship between NDVI and agreement for months between November and March. October data was important to include in our study because the “snow-on” date typically occurs during October in Norway, and this date is critical

for deriving snow cover phenology metrics used by wildlife ecologists studying migration timing and other seasonal phenomena. However, the strong effect of October on the NDVI estimate reinforces that NDVI was reflecting the absence of snow rather than canopy cover. We also examined maximum NDVI over each snow-covered season as a covariate instead of daily NDVI, and we found similar results (Appendix A7). In contrast, the tree canopy cover covariate had a negative effect on agreement as expected, even with October data included. The Landsat tree canopy cover product is a more direct measure of obstructing vegetation than NDVI (Potapov et al., 2021; Sexton et al., 2013), and our findings indicate that direct canopy products may be preferable to NDVI for snow mapping applications.

Agreement was also affected by cloudy days, supporting previous literature on limitations of cloud-gap filled products in cloudy regions (Gao et al., 2011; Hall et al., 2019b). However, agreement did not decrease linearly with time, but instead decreased rapidly and then leveled off after 3 days. This result is likely due to clouds changing the snow conditions, such as snowstorms increasing snow cover or increased humidity accelerating snowmelt (Zhang et al., 1996). Backfilling pixels with the most recent cloud-free value thus has limitations even for short cloud persistence durations. In cases when clouds persist for much of the winter, our results show that gap-filled products may be highly inaccurate, and wildlife camera data in these regions is especially valuable. While cloud-masked MOD10A1 values had substantially higher agreement with camera images than gap-filled MOD10A1F values, use of the MOD10A1 product comes at the cost of substantial data loss, as only 23% of pixels were usable due to cloud masking. Similarly, a study examining how snow properties affect movements of GPS-collared Dall sheep (*Ovis dalli dalli*) in Lake Clark National Park, Alaska, was only able to use 2.2% of their dataset when using cloud-masked MODIS products (Mahoney et al., 2018). Ultimately,

spatial products of snow cover may be able to automate the inclusion of snow values from camera networks when satellite values are not accurate or available, utilizing AI and machine learning to spatially and temporally fill gaps.

Gap-filling accuracy with camera-labeled values will depend on classification accuracy, and image classification error may be further reduced by using a binary classification, although some information is lost. However, binary maps can be especially useful for identifying snow-on and snow-off dates, with important applications for monitoring changing snow phenology and impacts on seasonal migrations and breeding seasons. The threshold NDSI value of 40.5 we identified using wildlife cameras in Scandinavia was remarkably similar to the value of 40 derived for MODIS from Landsat fractional snow-covered area maps in Canada (Klein et al., 1998). Thresholds in forested areas tend to be lower than open canopy thresholds because some snow visibility is blocked by the trees (Chokmani et al., 2010). Our findings were consistent with these trends, but the effect of canopy cover was minor (40.5 vs 41.5 for closed vs open canopy sites, respectively). By employing Youden's index to select the optimal threshold, we assumed equal weight to both snow and no snow classes. However, depending on the mapping needs, other threshold values could be used. For example, higher thresholds for snow might be desirable when making maps of the first "snow on" date in the fall to prioritize snow detection. Other studies have found adjustments to the threshold can increase regional accuracy (Chokmani et al., 2010; Da Ronco et al., 2020; Luo et al., 2022). While our study found that MODIS detected 88% of snow-covered pixels, Luo et al. (2022) found that MODIS identified just 14-18% of snow-covered pixels in forests when using conventional MODIS thresholds. MODIS snow detection tends to be less accurate in steep areas with complex topography (Rittger et al., 2021), and the Luo et al. (2022) study occurred in alpine terrain with sites > 2700 m a.s.l. and slopes between 19

and 34 degrees. Our study occurred at much lower elevations (0 – 800 m a.s.l.), with moderate slopes between 0.5 to 20 degrees. These differences reinforce our findings that agreement between camera and satellite may depend on environmental factors, and when using the two for validation or in-tandem, it is important to account for external context. Generally, a threshold of 40 is robust for this region, similar to other studies creating binary maps from forested ecosystems. A threshold of 10 from MOD10A2 would be low for this region, thus researchers should be aware that deriving their own binary thresholds is an important step for MODIS Collection 6 products. Future studies could employ this approach to create custom thresholds from cameras in their regions of interest.

Because our cameras were optimized for lynx detection, we did not control for field of view. Previous work suggests that wider field of views are more advantageous for snow cover monitoring (Parajka et al., 2012). Our results suggest that even narrow fields of view offer insight into snow conditions, but wider fields should provide a better observation of snow conditions at a scale more similar to satellite remote sensing. Additionally, we did not control for possible observation delays, which could be up to 24 hours depending on when the satellite passes over the area of interest and when the camera image is taken (Sugiura et al., 2013). One camera trap image per day appeared sufficient to connect to MODIS, but we recommend multiple images per camera each day to increase labeling options. Examining the outliers from our model evaluations aligns with these recommendations, because outlier images consisted primarily of those with narrow fields of view and active weather (Appendix A9). Continuous indices of vegetation greenness have been derived from camera images using RGB values as proxies for vegetation (Sun et al., 2021), but to our knowledge, no automated method of extracting continuous snow cover indices from camera images has been developed. AI

algorithms for automated snow detection from camera images are a promising area of development to increase the utility of wildlife camera networks for environmental monitoring.

Our study focused on comparing snow cover from cameras to MODIS snow products, and we found surprisingly strong agreement considering differences in spatial resolution. The Visible Infrared Imaging Radiometer Suite (VIIRS) instrument has a snow product similar to MODIS at 375-m spatial resolution (Riggs et al., 2017). Future work could explore incorporating multiple cameras in one satellite pixel to improve snow monitoring of patchy snow conditions, such as during snow accumulation and snow melt. Alternatively, camera images could be matched to finer-resolution snow products derived from satellites such as Landsat, Sentinel, and Planet CubeSat (Cannistra et al., 2021; Chokmani et al., 2010, 2010; Riggs et al., 2017). Snow maps must be derived by manually creating the NDSI maps from Landsat, Sentinel, and Planet sensors, but these products have spatial resolutions at 30 m, 10 m, and 0.7-3 m, respectively, closer to the camera field of view (Cannistra et al., 2021).

2.5 Conclusion

As the remote sensing community continues to develop new global products, the wildlife ecology community continues to expand camera trap networks for continuous biodiversity monitoring (Pettorelli et al., 2014; Steenweg et al., 2017). Connecting camera traps to satellite data represents an important step towards an interconnected network of ground-based remote sensing data that can improve researchers' and the public's ability to determine environmental changes and subsequent impacts on sensitive species. In Norway, snow cover extent has decreased by more than 20,000 km^2 (6% of the country area) since 1961 due to changes in temperature and precipitation (Rizzi et al., 2018; Skaugen et al., 2012). When these trends are

incorporated into climate impact models, predictions suggest accelerated rates of local extinctions across 273 species of Norwegian vegetation (Niittynen et al., 2018). With the increasing number of cameras operating as environmental monitoring devices, we can improve our understanding of both environmental and wildlife responses in a changing climate.

2.6 Data Availability

A selection of photos is publicly available at https://viltkamera.nina.no. Analysis code can be found at https://github.com/catherine-m-breen/MODIS-Snow-Cover-to-Binary-Snow-Covered-Area. GEE assets are available at the following links: https://code.earthengine.google.com/?asset=users/catherinembreen/MODIS_Norway

2.7 References

Beery, S., Wu, G., Rathod, V., Votel, R., & Huang, J. (2019). Long Term Temporal Context for Per-Camera Object Detection. *arXiv:1912.03538 [Cs, Eess, q-Bio]*. Retrieved from http://arxiv.org/abs/1912.03538

Berman, E. E., Bolton, D. K., Coops, N. C., Mityok, Z. K., Stenhouse, G. B., & Moore, R. D. (Dan). (2018). Daily estimates of Landsat fractional snow cover driven by MODIS and dynamic time-warping. *Remote Sensing of Environment*, *216*, 635–646. https://doi.org/10.1016/j.rse.2018.07.029

Bivand, R. (2022). R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data. *Geographical Analysis*, *54*(3), 488–518. https://doi.org/10.1111/gean.12319

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., et

al. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, *14*(1), 010401. https://doi.org/10.1088/1748-9326/aaeec1

Bokhorst, S. (2016). Changing Arctic snow cover: A review of recent developments and assessment of future needs for observations, modelling, and impacts, 22.

Bouyer, Y., San Martin, G., Poncin, P., Beudels-Jamar, R. C., Odden, J., & Linnell, J. D. C. (2015). Eurasian lynx habitat selection in human-modified landscape in Norway: Effects of different human habitat modifications and behavioral states. *Biological Conservation*, *191*, 291–299. https://doi.org/10.1016/j.biocon.2015.07.007

Brooks, M., E., Kristensen, K., Benthem, K., J. ,van, Magnusson, A., Berg, C., W., Nielsen, A., et al. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378. https://doi.org/10.32614/RJ-2017-066

Brown, R. D., & Mote, P. W. (2009). The Response of Northern Hemisphere Snow Cover to a Changing Climate*. *Journal of Climate*, *22*(8), 2124–2145. https://doi.org/10.1175/2008JCLI2665.1

Brown, T. B., Hultine, K. R., Steltzer, H., Denny, E. G., Denslow, M. W., Granados, J., et al. (2016). Using phenocams to monitor our changing Earth: toward a global phenocam network. *Frontiers in Ecology and the Environment*, *14*(2), 84–93. https://doi.org/10.1002/fee.1222

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

Carricondo-Sanchez, D., Odden, M., Linnell, J. D. C., & Odden, J. (2017). The range of the mange: Spatiotemporal patterns of sarcoptic mange in red foxes (Vulpes vulpes) as revealed by camera trapping. *PLOS ONE*, *12*(4), e0176200. https://doi.org/10.1371/journal.pone.0176200

Chokmani, K., Dever, K., Bernier, M., Gauthier, Y., & Paquet, L.-M. (2010). Adaptation of the SNOWMAP algorithm for snow mapping over eastern Canada using Landsat-TM imagery. *Hydrological Sciences Journal*, *55*(4), 649–660. https://doi.org/10.1080/02626661003747374

Coll, J., & Li, X. (2018). Comprehensive accuracy assessment of MODIS daily snow cover products and gap filling methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, 435–452. https://doi.org/10.1016/j.isprsjprs.2018.08.004

Crawford, C. J. (2015). MODIS Terra Collection 6 fractional snow cover validation in mountainous terrain during spring snowmelt using Landsat TM and ETM+: MODIS TERRA C6 FRACTIONAL SNOW COVER VALIDATION USING LANDSAT. *Hydrological Processes*, *29*(1), 128–138. https://doi.org/10.1002/hyp.10134

Curk, T., Pokrovsky, I., Lecomte, N., Aarvak, T., Brinker, D. F., Burnham, K., et al. (2020). Arctic avian predators synchronise their spring migration with the northern progression of snowmelt. *Scientific Reports*, *10*(1), 7220. https://doi.org/10.1038/s41598-020-63312-0

Da Ronco, P., Avanzi, F., De Michele, C., Notarnicola, C., & Schaefli, B. (2020). Comparing MODIS snow products Collection 5 with Collection 6 over Italian Central Apennines. *International Journal of Remote Sensing*, *41*(11), 4174–4205. https://doi.org/10.1080/01431161.2020.1714778

Didan, K. (2015). MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13A1.006

Dong, C., & Menzel, L. (2016). Producing cloud-free MODIS snow cover products with conditional probability interpolation and meteorological data. *Remote Sensing of Environment*, *186*, 439–451. https://doi.org/10.1016/j.rse.2016.09.019

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Favorskaya, M., & Buryachenko, V. (2019). Selecting Informative Samples for Animal Recognition in the Wildlife. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent Decision Technologies 2019* (Vol. 143, pp. 65–75). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-8303-8_6

Forrester, T., O'Brien, T., Fegraus, E., Jansen, P., Palmer, J., Kays, R., et al. (2016). An Open Standard for Camera Trap Data. *Biodiversity Data Journal*, *4*, e10197. https://doi.org/10.3897/BDJ.4.e10197

Franklin, S. E. (2020). Interpretation and use of geomorphometry in remote sensing: a guide and review of integrated applications. *International Journal of Remote Sensing*, *41*(19), 7700–7733. https://doi.org/10.1080/01431161.2020.1792577

Gao, Y., Lu, N., & Yao, T. (2011). Evaluation of a cloud-gap-filled MODIS daily snow cover product over the Pacific Northwest USA. *Journal of Hydrology*, *404*(3–4), 157–165. https://doi.org/10.1016/j.jhydrol.2011.04.026

Garvelmann, J., Pohl, S., & Weiler, M. (2013). From observation to the quantification of snow processes with a time-lapse camera network. *Hydrology and Earth System Sciences*, *17*(4), 1415–1429. https://doi.org/10.5194/hess-17-1415-2013

Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, *21*(12), 1534–1547. https://doi.org/10.1002/hyp.6715

Hall, D. K., Kelly, R. E. J., Riggs, G. A., Chang, A. T. C., & Foster, J. L. (2002). Assessment of the relative accuracy of hemispheric-scale snow-cover maps. *Annals of Glaciology*, *34*, 24–30. https://doi.org/10.3189/172756402781817770

Hall, D. K., Riggs, G. A., Foster, J. L., & Kumar, S. V. (2010). Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product. *Remote Sensing of Environment*, *114*(3), 496–503. https://doi.org/10.1016/j.rse.2009.10.007

Hall, D.K., Riggs, G.A., Solomonson, V., 2016. MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA. https://doi.org/10.5067/MODIS/ MOD10A1.006.

Hall, D. K., Riggs, G. A., DiGirolamo, N. E., & Román, M. O. (2019a). MODIS Cloud-Gap Filled Snow-Cover Products: *Advantages and Uncertainties* (preprint). Snow and Ice/ Remote Sensing and GIS. https://doi.org/10.5194/hess-2019-123.

Hall, D. K., Riggs, G. A., DiGirolamo, N. E., & Román, M. O. (2019b). 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

Hao, X., Huang, G., Zheng, Z., Sun, X., Ji, W., Zhao, H., et al. (2022). Development and

validation of a new MODIS snow-cover-extent product over China. *Hydrology and Earth System Sciences*, *26*(8), 1937–1952. https://doi.org/10.5194/hess-26-1937-2022

Hartig, F. (2022). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. Retrieved from https://CRAN.R-project.org/package=DHARMa

Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., Andrén, H., et al. (2019). Using by-catch data from wildlife surveys to quantify climatic parameters and the timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, rse2.136. https://doi.org/10.1002/rse2.136

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research*, *32*(1), 1–6. https://doi.org/10.1007/s11676-020-01155-1

Huang, X., Liang, T., Zhang, X., & Guo, Z. (2011). Validation of MODIS snow cover products using Landsat and ground measurements during the 2001–2005 snow seasons over northern Xinjiang, China. *International Journal of Remote Sensing*, *32*(1), 133–152. https://doi.org/10.1080/01431160903439924

Klein, A. G., Hall, D. K., & Riggs, G. A. (1998). Improving snow cover mapping in forests through the use of a canopy reflectance model. *HYDROLOGICAL PROCESSES*, *12*, 22.

Laforge, M. P., Bonar, M., & Vander Wal, E. (2021). Tracking snowmelt to jump the green wave: phenological drivers of migration in a northern ungulate. *Ecology*, *102*(3). https://doi.org/10.1002/ecy.3268

Liang, T., Huang, X., Wu, C., Liu, X., Li, W., Guo, Z., & Ren, J. (2008). An application of

MODIS data to snow cover monitoring in a pastoral area: A case study in Northern Xinjiang, China. *Remote Sensing of Environment*, *112*(4), 1514–1526. https://doi.org/10.1016/j.rse.2007.06.001

Liu, X. (2012). Classification accuracy and cut point selection. *Statistics in Medicine*, *31*(23), 2676–2686. https://doi.org/10.1002/sim.4509

Luo, J., Dong, C., Lin, K., Chen, X., Zhao, L., & Menzel, L. (2022). Mapping snow cover in forests using optical remote sensing, machine learning and time-lapse photography. *Remote Sensing of Environment*, *275*, 113017. https://doi.org/10.1016/j.rse.2022.113017

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., & Andersen, J. (2018). seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, *10*(1), 235–249. https://doi.org/10.5194/essd-10-235-2018

Madsen, J., Tamstorf, M., Klaassen, M., Eide, N., Glahder, C., Rigét, F., et al. (2007). Effects of snow cover on the timing and success of reproduction in high-Arctic pink-footed geese Anser brachyrhynchus. *Polar Biology*, *30*(11), 1363–1372. https://doi.org/10.1007/s00300-007-0296-9

Mahoney, P. J., Liston, G. E., LaPoint, S., Gurarie, E., Mangipane, B., Wells, A. G., et al. (2018). Navigating snowscapes: scale-dependent responses of mountain sheep to snowpack properties. *Ecological Applications*, *28*(7), 1715–1729. https://doi.org/10.1002/eap.1773

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

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Negi, H. S., Thakur, N. K., & Mishra, V. D. (2007). Estimation and validation of snow surface temperature using modis data for snow-avalanche studies in NW-Himalaya. *Journal of the Indian Society of Remote Sensing*, *35*(4), 287–299. https://doi.org/10.1007/BF02990785

Niittynen, P., Heikkinen, R. K., & Luoto, M. (2018). Snow cover is a neglected driver of Arctic biodiversity loss. *Nature Climate Change*, *8*(11), 997–1001. https://doi.org/10.1038/s41558-018-0311-x

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

Parajka, J., Haas, P., Kirnbauer, R., Jansa, J., & Blöschl, G. (2012). Potential of time-lapse photography of snow for hydrological purposes at the small catchment scale: POTENTIAL OF TIME-LAPSE PHOTOGRAPHY OF SNOW FOR HYDROLOGICAL PURPOSES. *Hydrological Processes*, *26*(22), 3327–3337. https://doi.org/10.1002/hyp.8389

Pettorelli, N., Laurance, W. F., O'Brien, T. G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, *51*(4), 839–848. https://doi.org/10.1111/1365-2664.12261

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., et al.

(2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Raleigh, M. S., Rittger, K., Moore, C. E., Henn, B., Lutz, J. A., & Lundquist, J. D. (2013). Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada. *Remote Sensing of Environment*, *128*, 44–57. https://doi.org/10.1016/j.rse.2012.09.016

Reinking, A. K., Højlund Pedersen, S., Elder, K., Boelman, N. T., Glass, T. W., Oates, B. A., et al. (2022). Collaborative wildlife–snow science: Integrating wildlife and snow expertise to improve research and management. *Ecosphere*, *13*(6). https://doi.org/10.1002/ecs2.4094

Riggs, G.A., Hall, D.K., Rom´an, M.O.. (2019). MODIS Snow Products Collection 6.1 User Guide, Version 1.0. NASA Goddard Space Flight Center. https://modis-snow-ice-gsfc. nasa.gov/?c=userguides.

Riggs, G. A., Hall, D. K., & Román, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, *9*(2), 765–777. https://doi.org/10.5194/essd-9-765-2017

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*, 647154. https://doi.org/10.3389/frsen.2021.647154

Rizzi, J., Nilsen, I. B., Stagge, J. H., Gisnås, K., & Tallaksen, L. M. (2018). Five decades of warming: impacts on snow cover in Norway. *Hydrology Research*, *49*(3), 670–688. https://doi.org/10.2166/nh.2017.051

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

Saloranta, T. M. (2012). Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model. *The Cryosphere*, *6*(6), 1323–1337. https://doi.org/10.5194/tc-6-1323-2012

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., et al. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, *6*(5), 427–448. https://doi.org/10.1080/17538947.2013.786146

Sirén, A. P. K., Somos-Valenzuela, M., Callahan, C., Kilborn, J. R., Duclos, T., Tragert, C., & Morelli, T. L. (2018). Looking beyond wildlife: using remote cameras to evaluate accuracy of gridded snow data. *Remote Sensing in Ecology and Conservation*, *4*(4), 375–386. https://doi.org/10.1002/rse2.85

Skaugen, T., Stranden, H. B., & Saloranta, T. (2012). Trends in snow water equivalent in Norway (1931–2009). *Hydrology Research*, *43*(4), 489–499. https://doi.org/10.2166/nh.2012.109

Solomon, S., Intergovernmental Panel on Climate Change, & Intergovernmental Panel on Climate Change (Eds.). (2007). *Climate change 2007: the physical science basis: contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge ; New York: Cambridge University Press.

Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B. H., et al.

(2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, *152*, 159–177. https://doi.org/10.1016/j.agrformet.2011.09.009

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., et al. (2017). Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, *15*(1), 26–34. https://doi.org/10.1002/fee.1448

Sugiura, K., Nagai, S., Nakai, T., & Suzuki, R. (2013). Application of time-lapse digital imagery for ground-truth verification of satellite indices in the boreal forests of Alaska. *Polar Science*, *7*(2), 149–161. https://doi.org/10.1016/j.polar.2013.02.003

Sun, C., Beirne, C., Burgar, J. M., Howey, T., Fisher, J. T., & Burton, A. C. (2021). Simultaneous monitoring of vegetation dynamics and wildlife activity with camera traps to assess habitat change. *Remote Sensing in Ecology and Conservation*, *7*(4), 666–684. https://doi.org/10.1002/rse2.222

Thapa, S., Chhetri, P. K., & Klein, A. G. (2019). Cross-Comparison between MODIS and VIIRS Snow Cover Products for the 2016 Hydrological Year. *Climate*, *7*(4), 57. https://doi.org/10.3390/cli7040057

Tobler, M. W., Zúñiga Hartley, A., Carrillo-Percastegui, S. E., & Powell, G. V. N. (2015). Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology*, *52*(2), 413–421. https://doi.org/10.1111/1365-2664.12399

Townshend, J. (2016). Global Forest Cover Change (GFCC) Forest Cover Change Multi-Year Global 30 m V001 [Data set]. NASA EOSDIS Land Processes DAAC.

https://doi.org/10.5067/MEASURES/GFCC/GFCC30FCC.001

Urbanek, R. E., Ferreira, H. J., Olfenbuttel, C., Dukes, C. G., & Albers, G. (2019). See what you've been missing: An assessment of Reconyx® PC900 Hyperfire cameras. *Wildlife Society Bulletin*, *43*(4), 630–638. https://doi.org/10.1002/wsb.1015

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wuertz, D., Setz, T., Chalabi, Y., & Boshnakov, G. (2023). timeDate: Rmetrics - Chronological and Calendar Objects (Version R package version 4022.108). Retrieved from https://CRAN.R-project.org/package=timeDate

Xiao, X., He, T., Liang, S., Liu, X., Ma, Y., Liang, S., & Chen, X. (2022). Estimating fractional snow cover in vegetated environments using MODIS surface reflectance data. *International Journal of Applied Earth Observation and Geoinformation*, *114*, 103030. https://doi.org/10.1016/j.jag.2022.103030

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

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zhang, H., Zhang, F., Zhang, G., Che, T., Yan, W., Ye, M., & Ma, N. (2019). Ground-based evaluation of MODIS snow cover product V6 across China: Implications for the selection of NDSI threshold. *Science of The Total Environment*, *651*, 2712–2726. https://doi.org/10.1016/j.scitotenv.2018.10.128

Zhang, T., Stamnes, K., & Bowling, S. A. (1996). Impact of Clouds on Surface Radiative Fluxes

and Snowmelt in the Arctic and Subarctic. *Journal of Climate*, *9*(9), 2110–2123. https://doi.org/10.1175/1520-0442(1996)009<2110:IOCOSR>2.0.CO;2

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems: *Data exploration*. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Chapter 3. SNOW DEPTH EXTRACTION FROM TIMELAPSE IMAGERY USING KEYPOINT DEEP LEARNING MODEL

Publication history: A version of this chapter was previously published in *Water Resources Research*. The full citation for this publication is:

Breen, C. M., Currier, W. R., Vuyovich, C., Miao, Z., & Prugh, L. R. (2024). Snow Depth Extraction From Time-Lapse Imagery Using a Keypoint Deep Learning Model. *Water Resources Research*, *60*(7), e2023WR036682. https://doi.org/10.1029/2023WR036682

Abstract: Snow pole time-lapse photography, in which a snow pole of a known height is installed in front of a camera and photographed repeatedly over the course of a snow season, allows a large network of sites to be established relatively quickly and affordably. However, current approaches for extracting snow depth from snow poles typically relies on time intensive manual photo processing. By integrating computer vision algorithms with snow pole photography, we present a method that uses a keypoint detection model to automatically observe snow height across a network of sites. At 20 snow pole locations from Grand Mesa, CO (n = 9,722 images), our model successfully predicts the top and bottom of the pole with a mean absolute error (MAE) of 1.30 cm. To assess model generalizability, we tested the model on 12 sites in Washington State (n = 1,770 images). When the Colorado trained model was fine-tuned using a subset of Washington images, the model predicted snow depth with a MAE of 4.0 cm. Best performance was achieved when both data sets were included during training, with a MAE

of 2.05 cm for Colorado images and a MAE of 1.14 cm for Washington images. We demonstrate that, especially when trained using a subset of site-specific data, a keypoint detection model can accelerate snow pole automation. This algorithm brings the hydrology community one step closer to a generalized snow pole detection model, and we call for a future model that integrates across time-lapse images from additional locations.

Keywords: snow depth, keypoint model, time-lapse imagery, automation

3.1 Introduction

Within-season snow depth observations are necessary for a variety of applications, including estimating water storage, understanding the physical processes that control snow accumulation and ablation, and identifying snow conditions and their influence on animal movement and behavior (Jonas et al., 2009; Martin et al., 2020). Snow depth, one of the simplest snow variables to measure, is closely linked with a suite of other snow properties, including snow-water equivalent and snow-melt timing (Sturm et al., 2010). Methods for measuring snow depth include less-expensive methods that have less coverage, such as manual snow probing in the field and acoustic snow depth sensors, and more expensive, advanced methods, such as airborne remote sensing (e.g., Light Infrared Detection and Ranging or LIDAR) (Deems et al., 2013; Pflug & Lundquist, 2020). Snow pole time-lapse photography, in which a pole is installed in front of a camera and photographed repeatedly over the course of a snow season, carries characteristics of both traditional and advanced methods for snow depth. The approach can scale to large areas, produce detailed time series, and is cheaper and easier to install than advanced methods (Bongio et al., 2021; Currier et al., 2016; Dwivedi et al., 2023; Floyd & Weiler, 2008; Parajka & Blöschl, 2008; Robles et al., 2021).

The combination of snow poles and time-lapse photography allows snow scientists to establish a network of sites across a given study area and easily measure snow depth (Broxton et al., 2019; Cao et al., 2018). However, these cameras typically generate thousands of images that require processing to extract snow depth. The labeling demand increases in year-to-year studies, and when networks are installed in large watersheds or habitat ranges where upwards of hundreds of cameras have been installed at a time (Genthon et al., 2016; Kohler et al., 2006; Schöner et al., 2009). Furthermore, there are dozens of research groups utilizing time-lapse cameras with snow poles, but snow pole designs vary in color and pattern depending on the study needs (Collados-Lara et al., 2020; Cosgrove et al., 2021; Currier et al., 2016; Hofmeester et al., 2019; McCreight et al., 2014). The result is an enormous amount of data and a need for a scalable method to improve automated labeling. To date, no automated method exists to retrieve depth from snow poles across different sites. Previous strategies include color thresholding to create a linear mask to represent the pole (Bongio et al., 2021; Currier et al., 2016; Floyd & Weiler, 2008), but researchers found that variable lighting conditions across the winter season compromised the accuracy of a single season-long threshold. Parajka et al. (2012) presented a "routine" to automate the process, although the method was not described in detail. Strickfaden et al. (2023) developed code to extract depth from virtual snow poles using reference objects, but snow depths must be read manually. Thus, these studies suggest tools for automation, but no study provides a scalable method at high precision. Studies with large networks of camera sites with snow poles still default to manual approaches (Hofmeester et al., 2019; Sirén et al., 2018).

Computer vision algorithms such as neural network models offer potential for a scalable method as they can learn complex patterns on large amounts of data (Xu & Liang, 2021). With time-lapse cameras, they have successfully identified objects of interest in images across

different lighting conditions. For example, object detectors, neural network models that identify instances of an object of interest (Schneider et al., 2018), have identified animals in camera images under various lighting conditions and against different environmental backgrounds (Norouzzadeh et al., 2021). Mask region-based convolutional neural network models successfully segmented snow poles from camera trap images (Kopp et al., 2019), but so far the method has only been tested on two sites. Application of neural networks to snow pole detection is still limited, and more work is needed to understand the ability of neural networks to detect snow poles across multiple sites and on sites not seen during training.

Due to the limited work on neural network models for snow pole detection, there is not yet consensus on the model architecture that best identifies snow poles for snow depth extraction, or the best way to apply the model for future use. Keypoint detection models are types of convolutional neural network models that define locations or points that stand out in an image. They have shown success in generalizing to new data when trained on large sample sizes (Mathis et al., 2021), but, to our knowledge, these models have not been tested on snow poles. Keypoint detection models are commonly used to detect features on human faces (Zhang et al., 2021) and body features on animals (Mathis et al., 2021), as the model learns points of interest by incorporating the information surrounding the point. For snow pole detection, one main advantage of keypoint detection models is that the number of points detected can be set, so that the model does not detect more objects than necessary. This approach differs from object detector models, in which the object of interest can be detected multiple times in the same image, leading to false positive detections that require post-processing to filter out (Beery, Morris, & Yang, 2019). Previous approaches to measure snow depth from snow poles often identify the pole's top and the snowline-pole interface (i.e., hereafter referred to as “bottom”) as a first step

(Bongio et al., 2021; Currier et al., 2016; Garvelmann et al., 2013). Because the top and bottom of a pole have similar information surrounding the point from image to image, keypoint detection models' ability to identify changes in the pole length throughout the winter season carries potential.

Using snow pole images from two different regions of the Western U.S., we present a trained keypoint detection model to automate snow depth extraction from snow poles. We first train the model on a random subset of images from sites in Colorado to demonstrate that the keypoint detection model can identify the correct snow pole lengths at varying snow depth. Second, we assess this model for errors as a result of site factors, such as lens distortion and proximity of the pole to the camera. Next, we test the accuracy of the model to predict snow depth on a novel data set in Washington, mimicking a work-flow for a future user. Finally, we show that by increasing the number of images from Washington presented during training, the model can better identify the top and bottom of the pole and calculate subsequent snow depth on the Washington images better than a model trained on Washington images alone. Thus, the pre-trained Coloado model can serve as a foundation model for future users to fine-tune the model using subsets on their own data set, and we call for integrated data sets and scalable models in future work. We present the model and code base for future use.

3.2 Methods

3.2.1 *Colorado Images*

To train the initial model, we used images from 20 sites from the NASA SnowEx 2020 field campaign (Figure 3.1; Breen et al., 2022). The NASA SnowEx project is a 5-year effort to develop remote sensing techniques for measuring snow and water availability in snowpacks

(Marshall et al., 2021). The 2020 field campaign occurred in Grand Mesa, Colorado, USA (latitude: c. 39.007° to 39.055°; longitude: c. −107.934° to 108.216°), a 470 km^2 arid plateau with elevations ranging from 3,440 m to 3,922 m a.s.l. along an east to west gradient, and precipitation averaging around 600 mm (De Lannoy et al., 2018; Appendix B, Figure B1). Forest composition includes shrubby cinquefoil (*Dasiphora fruticose*) steppe with isolated Engelmann spruce (*Picea engelmannii*) tree islands in the west. Farther to the east, there is dense, continuous Engelmann spruce and subalpine fir (*Abies lasiocarpa*) forests with lodgepole pine (*Pinus contorta var. latifolia*) and some aspen (*Populus tremuloides*) stands (Hojatimalekshah et al., 2021). At each site, a time-lapse camera was located 3–5 m away from a snow pole. Snow poles were constructed by driving a 1.524 m tall t-post into the ground and placing a PVC pipe (3.05 m tall, 2.54 cm in diameter) over it. Each PVC pipe was spray-painted red and had a 10 cm band of yellow duct tape at the top. The red pole design was selected due to previous studies finding that red contrasted with white snow, green vegetation, and blue sky and was advantageous for both human and automated processing (Currier et al., 2016; Floyd & Weiler, 2008). Brightly colored poles have disadvantages, including that they can bias animal movement and can draw attention from recreationists (Meek et al., 2019; Rowcliffe et al., 2008), but these impacts were not incorporated for this study because the study was not focused on animal activity and took place away from recreation trails. Poles in Grand Mesa, CO were installed between late September 2019 and early June 2020. Wingscape cameras (WCT-00126 TimeLapseCam Pro) were installed on t-posts 2 m above ground, using a wood board to hold the camera in place (Appendix B, Figure B2). Cameras were set to take either three images daily (11 a.m., 12 p.m., 1 p.m.) or twice daily (11 a.m. and 12 p.m.).

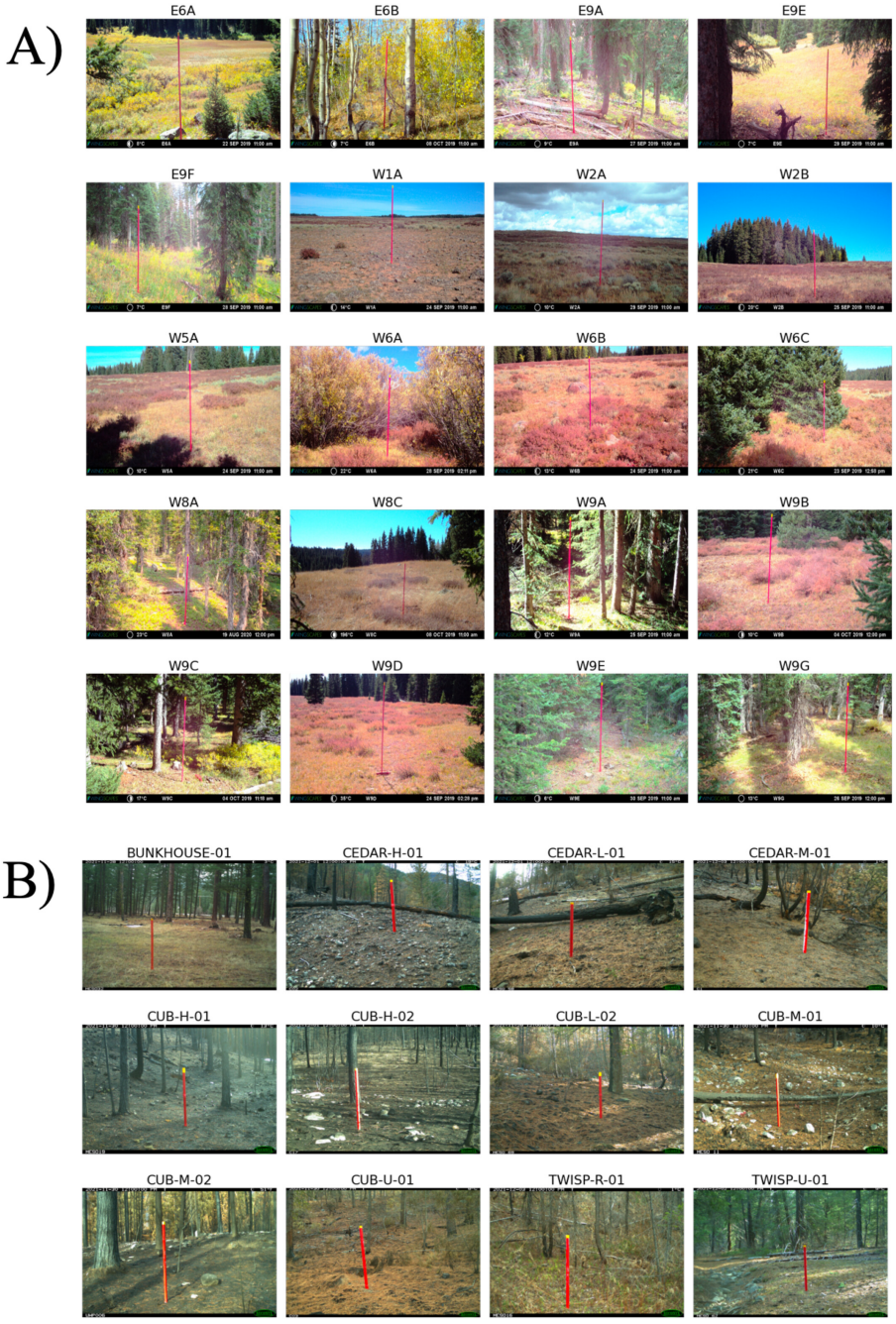


Figure 3.1 Example images from each camera show variation in lighting conditions, forest canopy cover, ecosystem characteristics, and pole placement from sites in panels (a) Grand Mesa, CO and (b) Okanogan County, WA. Example pictures from a site set-up for both Colorado and Washington are included in Appendix B, Figure B2.

3.2.2 *Washington Images*

To test the pre-trained Colorado model on a novel data set, we used images from 12 sites installed in Okanogan County, Washington (Figure 3.1). Poles were installed in late November 2021 and retrieved mid-May 2022. The images were part of a University of Washington (UW) study investigating snow depth characteristics in different burned areas of the Methow Valley, Washington (latitude: c. 48.050° to 49.150°; longitude: c. −120.900° to −119.700°). Unlike the drier, alpine climate of Grand Mesa, the 11,040 km^2 Okanogan study area is a temperate forest with elevations between 693 m and 1,051 m a.s.l, with mean annual precipitation of 1,000 mm. Forests consist of ponderosa pine (*Pinus ponderosa*), western hemlock (*Tsuga heterophylla*), and western red cedar (*Thuja plicata*) (Bassing et al., 2023). Following the protocol of the Colorado installation, sites were composed of a time-lapse camera (Reconyx model HC500 or HC600) mounted approximately 1 m above ground on a tree, with a snow pole installed 3–5 m away in its field of view. The pole was of PVC pipe spray-painted red with a 5–10 cm band of yellow duct tape at the top, and it was placed over a t-post driven into the ground. Washington poles were slightly shorter and wider than Colorado poles, ranging between 1 and 2 m tall depending on local snow conditions and 3.81 cm in diameter. Poles were cut into 50 cm long segments, then assembled using at least one pipe connector to ease transportation to the field. Cameras were set to take one picture every hour. We selected images taken at 12 p.m. to stay consistent with the daytime images in the Colorado data set, as previous studies have found that nighttime images have lower accuracy (Beery, Wu, et al., 2019). Additionally, selecting for one image per day allowed us to test for the model's performance on new pole locations while minimizing labeling efforts. As further expanded upon in Section 3.2.3, labeling was done manually by clicking the top and bottom of the pole. The precise height of the camera was determined by pole placement

and the ability to photograph the full pole without obstruction from vegetation or terrain. We summarize the attributes of the two data sets in Table 3.1.

Table 3.1. Differences Between Colorado and Washington Snow Pole and Camera Installations

	Colorado	**Washington**
Height	3.05 m	1-2 m
Diameter	2.54 cm	3.81 cm
Ecosystem	Dry, arid mesa	Temperate forest
Time-lapse image	11AM, 12PM, 1PM	12PM
Camera	Wingscape camera	Reconyx model
Distance between pole and camera	3-5 m	3-4 m

3.2.3 *Data Labeling*

To account for camera or pole tilt that sometimes occurred, we manually inspected both data sets and removed images that had tilts placing the top or bottom of the pole beyond the view of the camera. Next, we determined the height, in pixels, of each pole. A user inspected each image, and manually selected the pole top and bottom using the *ginput* function in Python's "matplotlib" package, which gave pixel coordinates of each point. Snow mounds can occasionally build up on the base of the pole in heavy snow conditions. In these instances, we selected the place in the image where the pole intercepted the snowpack in the absence of the snow mound. Average time to process photos was about four hours per 1,000 images, or about 35 hours of total labeling. After, we calculated the length of the pole in pixels by finding the Euclidean distance between the pixel coordinates using the *dist* function from Python's "math" package. We then calculated the pole length in centimeters by multiplying the length in pixels by a pixel-to-centimeter conversion that was unique to each pole based on pole placement. The centimeters per pixel (hereafter "cm/pixel") conversion was calculated by dividing the height of

the pole in centimeters by the height in pixels of the pole without snow, following the approach in Floyd and Weiler (2008) and Bongio et al. (2021) for converting from pixels to centimeters. The height of the pole in centimeters was recorded when each pole was installed in the field. The height of the pole in pixels without snow was identified by finding the pixel length of a pole in a snow-free image. We repeated this 10 times to account for user error and found the mean and standard deviation. We compared that conversion to using the top 10 cm yellow band instead of the full pole. However, we found the clicking precision for the yellow band more difficult when the poles were further away from the camera. Additionally, the standard error was larger for the top 10 cm yellow band compared to the conversion when using the full pole in all 32 cases (Appendix B, Table B1). Using the cm/pixel conversion, all pixel lengths were converted to a depth in centimeters by subtracting the length from the original length of the pole in centimeters (Equation 3.1).

$$snow\ depth = snow\ free\ pole\ (cm) - pole\ with\ snow\ (pixels) \times \frac{length\ (cm)}{pixel} \qquad \text{(Eqn. 3.1)}$$

3.2.4 *Model Training Using Colorado Images Only*

To first investigate a model's ability to predict various snow depths when trained on a dataset of interest, we randomly split the Colorado data set 90/10 for training and testing, comprising 8749 and 972 images, respectively (Schneider et al., 2020). We called this model the "CO-only" model, referencing that it was trained on Colorado images only. We downsized images to 224 × 224 pixels following machine learning best practices that suggests reducing memory load improves training time (Miao et al., 2018). We then predicted two key points, the top and bottom of the pole, using a ResNet-50 CNN model from Python's "pytorch" package (He et al., 2015). A ResNet-50 model is part of the "residual network" family of CNN models,

because they use the residual of the output and input to improve results over the course of training. They were developed to improve performance on tasks with fine-grain features, such as images, compared to previous CNN models (He et al., 2015). The model contains 48 convolutional layers and 2 MaxPool layers. Convolutional layers are a set of filters (or kernels) that perform matrix calculations along the input data and are learned throughout the training. MaxPool layers take the maximum value of the region of overlap between the filter and input data (He et al., 2015). We utilized a ResNet-50 model pre-trained on ImageNet, a database of over 1 million images that when used to initialize model parameters, reduces training time and can achieve improved results on feature detection tasks (Deng et al., 2009). Other neural networks trained on camera trap images have successfully used them for wildlife species tracking and identification (Norouzzadeh et al., 2021; Tuia et al., 2022). CNN models can be set up with different parameter constants that control learning, including batch size, the weight decay parameter, and the number of epochs. Batch size is the number of units (i.e., images) presented to the model before it is updated. The weight decay is a regularization parameter that controls the complexity of the model, and the number of epochs is the number of iterations the model makes through all the data (Smith, 2018). We used the maximum possible batch size that our Graphics Processing Unit (GPU) could process (n = 64), as consistent with findings that the maximum possible batch size speeds up run-time and improves accuracy (Kochura et al., 2020). We experimented with the learning rate, because learning rate regulates the weights of the neural networks and can affect model learning (Simões et al., 2023). After finding that 1e−4 learning performed best (MAE equal to 1.68 versus 1.92 with a learning rate of 1e-5), we continued with the model that utilized that learning rate for model evaluation. We also employed early stopping after the model validation loss no longer improved after 10 epochs. Early stopping prevents over-

fitting on the training set, which can make models more generalizable on test sets (Jabbar & Khan, 2014). All models were trained on an Azure NC6s v3 virtual machine (VM) with a Tesla V100-PCIE-16GB GPU with a Linux "Data Science Virtual Image." Model training took about 1 min per epoch, or 1 hr for 60+ epochs, and cost $0.90 USD/hour using pay-as-you-go pricing. Training is doable on a local machine with a smaller batch size and longer run time, and we include steps to do so on the GitHub repository.

After model training, snow depth was calculated from the predicted keypoints using the same approach to calculate depth as the manually derived snow depths (see Equation 3.1). All calculations were done after projecting the detected keypoints onto the image at native resolution to use the same conversion factor as manual estimates for snow depth. The reprojection step introduces uncertainty as the predicted point is the centroid of the grid of pixels that corresponds to the predicted point at the down-sampled (224 × 224) resolution. To quantify this uncertainty, we divided the original height and width by 224, then divided by two to return the upper and lower bounds of the uncertainty in pixels. The uncertainty was then converted to centimeters by multiplying by each site's cm/pixel conversion (from Table B1 in Appendix B). The upper and lower uncertainty ranged between 1.71 and 3.55 cm for all cameras.

3.2.5 *Model Evaluation*

We evaluated the model trained on Colorado images using the Colorado test set of the manually derived snow depths (n = 972 images). The model was evaluated by calculating the residual error, the mean absolute error (MAE), and the coefficient of determination (R^2) between the predicted and true values for snow depth in cm. MAE was calculated using the *mean_absolute_error* function from Python's "sklearn" package. The R^2 value represents the

goodness of fit of the model, and we calculated it using the *corrcoeff* function from Python's "numpy" package. These metrics have been used by other studies that evaluated accuracy of snow depth time series (Bongio et al., 2021; Garvelmann et al., 2013; Kelly et al., 2003).

3.2.5.1 Evaluating Impact of Lens Distortion and Other Impacts on Model Accuracy

Lens distortion and the distance between pole and camera could affect model accuracy. Lens distortion is caused by the curve of the lens, the focal distance, and the angle at which the photo was taken (Leorna et al., 2022). Our model did not account for lens distortion, so we compared results to the publicly available version of the Colorado data set that included a lens distortion adjustment (Breen et al., 2022). Objects photographed closer to the camera tend to be easier for neural networks to identify (Norouzzadeh et al., 2018), so we tested whether poles closer to the camera had higher accuracy. The number of pixels per centimeter is larger when the pole is closer to the camera. This value is equivalent to the inverse of the cm/pixel ratio, and we used pixel/cm rather than cm/pixel to calculate the gain in accuracy with increased pixel resolution. We then fitted a linear model to predict the absolute value of the difference between the manual and automated snow depth values in centimeters from the pixel/cm conversion using the "statsmodels" package in Python.

Neural networks are considered black box models, because it can be difficult to assess the conditions under which they make successful decisions (Liang et al., 2021). To better understand conditions that lead to poor model performance, we identified the top 30 outlier predictions using the interquartile range (IQR) method and inspected images for patterns of impacts from variations in lighting, storms, and camera malfunction, as suggested for data point inspection for interpretable machine learning in ecology (Lucas, 2020). Previous studies using neural networks

also document how lighting conditions and weather can impact model accuracy (Gomez Villa et al., 2017; Maile et al., 2023; Schneider et al., 2018). We therefore added categories to each image in our Colorado test set to represent conditions that could result in differences in lighting and weather as well as other categories that may influence accuracy. The categories included: (a) lighting (sunny vs. cloudy); (b) weather (precipitation vs. no precipitation); (c) canopy cover (open vs. closed); (d) month of year; and (e) snow depth (binned in 20 cm increments). We tested for differences in error when compared to the overall error mean. In cases of two-way comparisons, we conducted a t-test using the *ttest_ind* function from Python's "scipy" package. In cases of three or more comparisons, we conducted a one-way Analysis of Variance test using the *f_oneway* function also from Python's "scipy" package.

3.2.6 *Testing and Fine-Tuning Using Washington Images*

CNN models often perform well on data presented during training, and generalizing to new data is an important area of research (J. Wang et al., 2022). Furthermore, we wanted to test the model using a workflow that would mimic real-world application. Given that snow poles are currently installed under a variety of environmental conditions (Cosgrove et al., 2021; Sanmiguel-Vallelado et al., 2017; Schöner et al., 2009), we wanted to test our model in a new region. We tested the performance of our best-performing Colorado model from 3.1 on the Washington images (n = 1,770 images), using the same evaluation metrics defined in 3.2. The Washington images represent how the model might perform on a future user's network of snow poles without any additional model training.

We then tested whether "fine-tuning" with a small subset of the Washington images could improve Washington snow depth estimation without sacrificing accuracy in Colorado

snow depth estimation. “Fine-tuning,” or the process of retraining the model on data of interest to improve model performance, is recommended by researchers before using other keypoint detection models (Mathis et al., 2021). It involves using the weights of a pre-trained model as the starting weights, or in this case, the weights from the model trained on the Colorado data set. Using the model trained on Colorado images with the lowest MAE (as detailed in 3.2), we “unfroze” the top layer to modify model weights to continue training (Bajramović & Žunić, 2023). We then proceeded to train the model in multiple experiments, with each experiment increasing the number of images from each of the 12 Washington sites. The following subsets of *n* images per camera were used: 1, 2, 5, 10, 15, 20, 25, 30. In each case, the images were stratified across the winter season, such that for each camera, *n* images were evenly spaced apart starting with the earliest image in the camera set. By doing so, we increased the chances of different snow depth and winter conditions in our training (L. Cao & Shen, 2022). This procedure replicates a real-world pipeline in which a researcher labels a few images from each camera, fine-tunes this model, then deploys the model on the entire data set. Our Washington training subset ranged from 12 to 360 images depending on the experiment. We called these models the “fine-tuned models,” referencing the process of fine-tuning the pre-trained CO-only model with a subset of Washington images. We tested these models on the Colorado test set (*n* = 972 images) and the Washington data set without the images used from the fine-tuned training (*n* = 1,410 images), using the same evaluation metrics in 3.2. We kept the Washington test set the same size to be able to compare across experiments.

We selected the fine-tuned model with the least number of images to train the model while also having an accuracy below 5 cm (the approximate accuracy achieved in other methods: see Kopp et al., 2019). We compared this performance to a model trained on the same subset of

Washington images without the pre-trained weights from the Colorado model. This model was called the "WA-only" model, reflecting that it was trained on only Washington images. The fine-tuned and WA-only model were trained on a CPU (i.e., Apple Macbook Pro with M1 processor) to demonstrate the feasibility of this experiment without expensive GPU resources. We amended the batch size to 4 to avoid overloading the computer memory and left all other parameters the same. We include a workflow diagram in Figure S3 in Supporting Information S1 summarizing the steps for model training, fine-tuning, and how these steps can be replicated by a user. We also combined both data sets and retrained the model, calling this model the "CO + WA model." Due to the data set size, the CO + WA model was trained on the GPU machine with the same parameters as Section 3.2.

3.3 Results

3.3.1 *Colorado Model Results*

The CO-only model, or the model trained on a random 90% subset of the Colorado images, was able to identify the correct pole length in centimeters with 2.21 ± 2.96 cm residual error, 1.30 cm MAE, and R^2 of 0.99 (Table 3.2). The model demonstrated a strong ability to track variations in snow depth along the pole at individual sites and in different lighting conditions (Figure 3.2a). When inspecting the performance at the Colorado sites (n = 20), the automated time series closely matched the manually derived time series for snow depth at all cameras (Figure 3.3). Overall, the maximum error at any camera was <5 cm (Appendix B, Table B2). The reprojection step added an uncertainty ranging between 1.71 and 3.55 cm depending on the distance between the camera and pole (Appendix B, Table B3). We detected 30 outliers using IQR method and identified the predicted values that were in the first and fourth quartile

(Chandola et al., 2009). Upon inspection, fog, high-contrast shadows, patchy snow, and tall grass may have created larger differences between the actual and predicted values (Figure 4). Camera W8C had the most outliers (n = 6) compared to a mean of 1.5 outliers at each camera. This site had the largest cm/pixel conversion, or the lowest resolution for pixel to centimeters, suggesting that any difference in pixels would lead to a larger difference in centimeters.

Table 3.2 Epochs, Train and Test Data Set Information, and Model Performance for Four Different Models

Model	**Epochs**	**Train size**	**Data region**	**Test size**	**RE (cm)**	**MAE (cm)**	R^2
CO-only	35	8749	Colorado	972	-2.21 ± 2.96	1.30	0.99
			Washington	1770	68.21 ± 28.85	68.21	0.42
WA-only	100	120	Washington	1410	-2.07 ± 6.08	6.42	0.92
			Colorado	1749	-60.43 ± 41.18	60.85	0.44
Fine-tuned	42	120	Washington	1410	-1.35 ± 5.08	3.98	0.96
			Colorado	972	-78.0 ± 39.76	78.14	0.52
CO+WA	72	8917	Washington	170	-0.70 ± 1.30	1.14	0.99
			Colorado	972	-1.55 ± 2.09	2.05	0.99

Note. The CO-only model refers to the model trained on a random 90% subset of the Colorado images. The WA-only model refers to a model trained on a subset of Washington images. The "Fine-tuned model" refers to the Colorado model, fine-tuned with Washington images. The CO + WA model was trained on a combined 90% subset of the Colorado and Washington images.

3.3.1.1 Impacts of Lens Distortion and Site Factors on Model Accuracy

When comparing the automated time series to the publicly available time series that accounted for lens distortion, the mean difference was −1.73 cm (σ = 2.93). The accuracy only slightly decreased at poles further from the camera: about a 1 cm loss in absolute residual error for every ∼0.5 pixel increase in the number of pixels that represented a centimeter, but the difference was not significant (slope = −0.45, $F_{1, 20}$ = 3.09, p = 0.09; Appendix B, Figure B4). Conditions, including lighting, weather, and canopy cover, showed no significant differences in error means (Appendix B, Figure B5). Month of year and snow depth, however, both

demonstrated conditions with above average error. Many of the shoulder winter months (i.e., September, November, December, May, and June) demonstrated higher error ($F = 12.8$, $df = 890$, $p < 0.005$, Appendix B, Figure B6), and snow depth bins below 80 cm and the highest bin (160–180 cm) had higher error ($F = 5.5$, $df = 890$, $p < 0.005$, Appendix B, Figure B6). Upon inspecting images, many of the images coincided with the images that were also identified as outliers above, including images taken during the early season and images with patchy snow conditions.

3.3.2 *Fine-Tuned Model Results*

The CO-only model performed poorly on the Washington images ($n = 1{,}770$ images), and could not detect the top and bottom of the pole (Table 3.2; Figure 3.2B). Figure 3.2B shows that the model incorrectly predicts that the pole is further to the left than the actual pole placement. However, we found that fine-tuning the Colorado model using a subset of images from the Washington data set improved Washington snow depth predictions. For example, with 10 images per camera, we found a residual error of -1.35 ± 5.08 cm, a MAE of 3.98 cm, and a R^2 value equal to 0.96. The performance was also stronger and reached the criteria for early stopping faster than a model that was not trained on Colorado images (Table 3.2). Figure 3.5 demonstrates that this fine-tuned model was able to track snow depth across sites and winter season, although there is larger error at higher snow depths. We found that when we added the maximum snow depth value from each camera into the training set, this problem was corrected (Appendix B, Figure B9). The model predicted negative snow depth values during the early and melt season for some poles (e.g., BUNKHOUSE-01, CEDAR-H-01, CUB-L-02, CUB-M-01, CUB-U-01, and TWISP-U-01), but overall matched the trend of snow depth at all sites. The reprojection step

added an uncertainty ranging between 0.96 and 1.99 cm depending on the distance between the camera and pole (Appendix B, Table B3). Snow mounds were common in the Washington data set, and the model was able to identify the point approximately where the pole would intercept the snowpack in the absence of the snow mound (Appendix B, Figure B7). Occasionally, the model overestimated snow depth in the early season, such as at CEDAR-M-02 and the CUB-M-02 sites. Upon inspection, the pole tilted slightly at CEDAR-M-02, and the model may have predicted the pole based off previous images without the tilt, before learning the new placement of the pole. This site was also on a slope, and the slope shifted the correct prediction lower in the image than previous. At CUB-M-02, the model mistook a shadow as being the base of the pole in both the early and melt season (Appendix B, Figure B7). Unsurprisingly, we found that accuracy increased when the model was fined-tuned with more images from the Washington data set (Figure 3.6; Appendix B, Table B4).

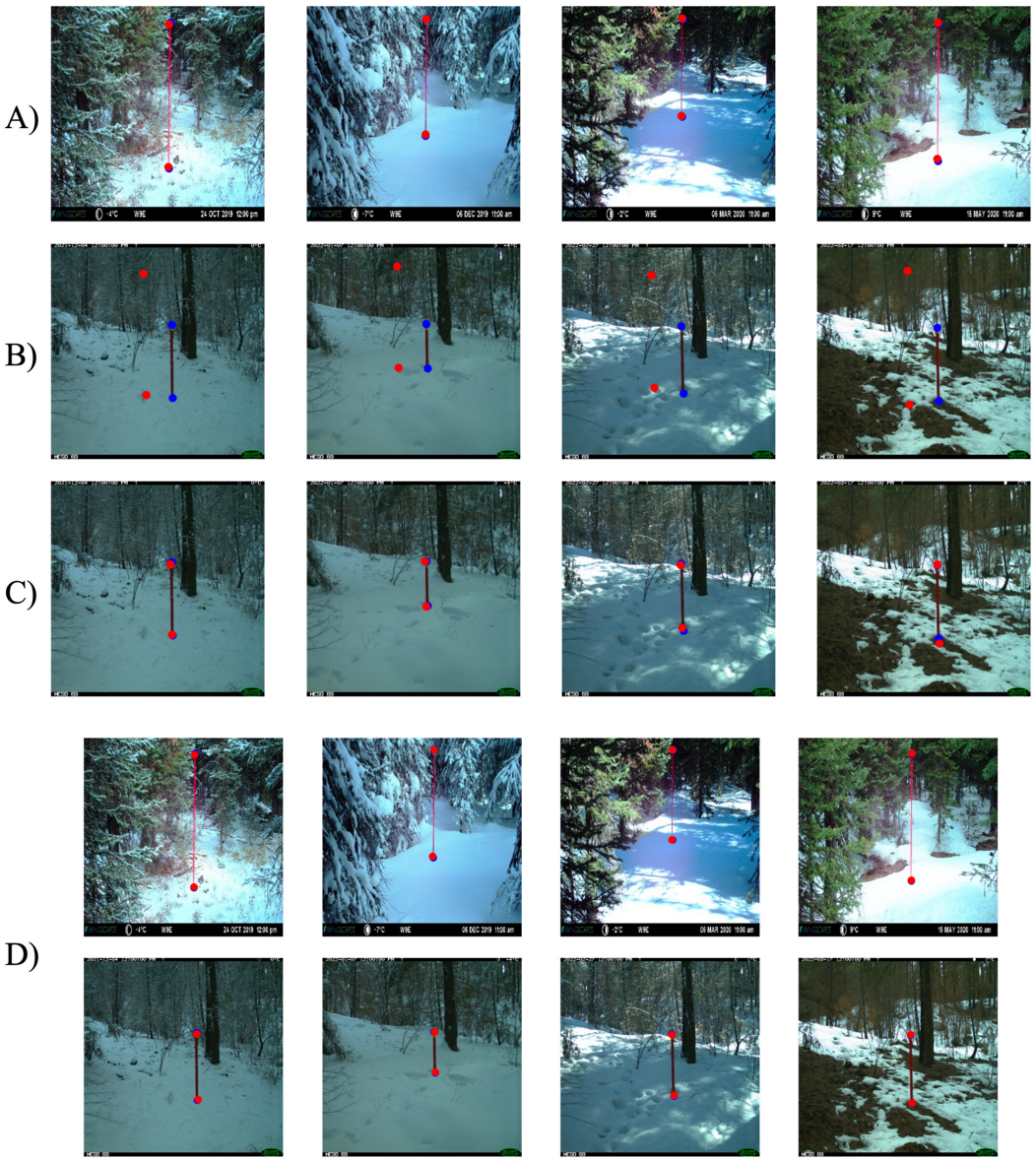


Figure 3.2 Blue dots represent actual top and bottom locations, the red dot represents the predicted top and bottom point. When there is overlap, blue dots are behind red dots. (a) Shows Colorado site W9E and (b) Washington site CUB-L-02 results when the model is trained only on a 90% random subset of Colorado data. The model (i.e., the CO-only model) successfully identifies the target points at various snow depths and in various lighting conditions when presented with example images during training. However, the middle row (b) suggests that the model inaccurately predicts points when it has not seen the sites during training. Row (c) shows

that when a subset of WA images is included in training (i.e., the fine-tuned model), the model improves prediction of top and bottom locations at CUB-L-02. Row (d) shows that when both data sets are combined during training (i.e., CO + WA model) the model can predict both top and bottom locations.

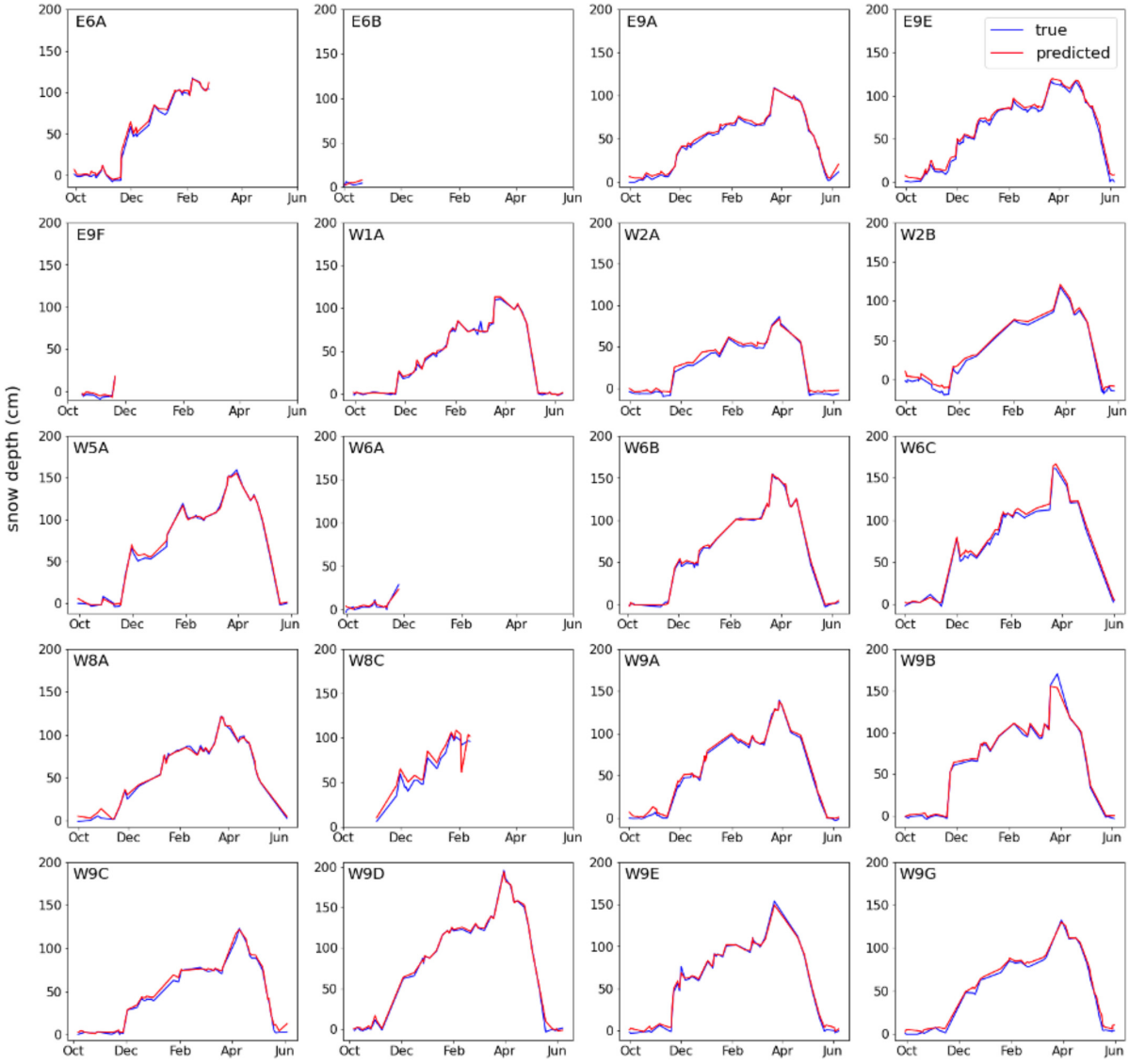


Figure 3.3 Comparison of snow depths (cm) from the validation data set at all Colorado sites. The true snow depth line was calculated by manually selecting the top and bottom of the pole. The predicted snow depth line was found using the keypoint detection model. Both true and

predicted depths were converted from pixels to centimeters by finding the pole length in pixels, multiplying by a cm/pixel conversion, and subtracting from the full length of the pole in centimeters. Missing data from E9F, W6A, W8C are the result of camera malfunctions on 30 November 2019, 28 November 2019, and 24 February 2020, respectively. Missing data from E6B are a result of the camera tilting such that the pole is outside the image.

Figure 3.4 These 30 images present the greatest difference between actual and predicted measurement in centimeters, sorted by most error (top-left corner) to least error (bottom-right corner). The blue circles represent the actual locations and the red circles represent the predictions generated by the keypoint detection model.

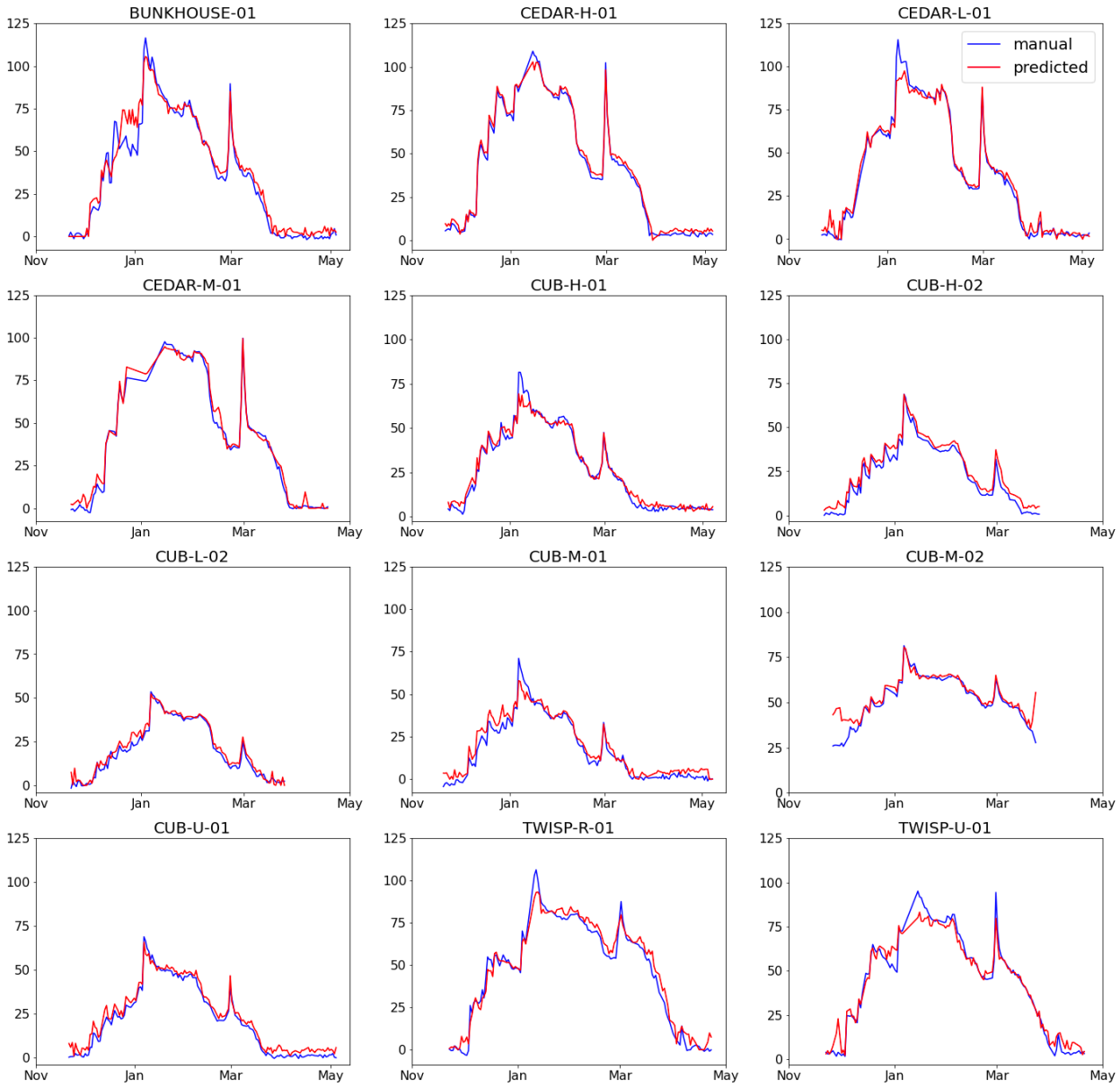


Figure 3.5 Comparison of snow depths (cm) at Washington sites. The manual snow depth line was calculated by manually selecting the top and bottom of the pole and converting to snow depth. The predicted snow depth line was found using the fine-tuned model with 10 images from each camera.

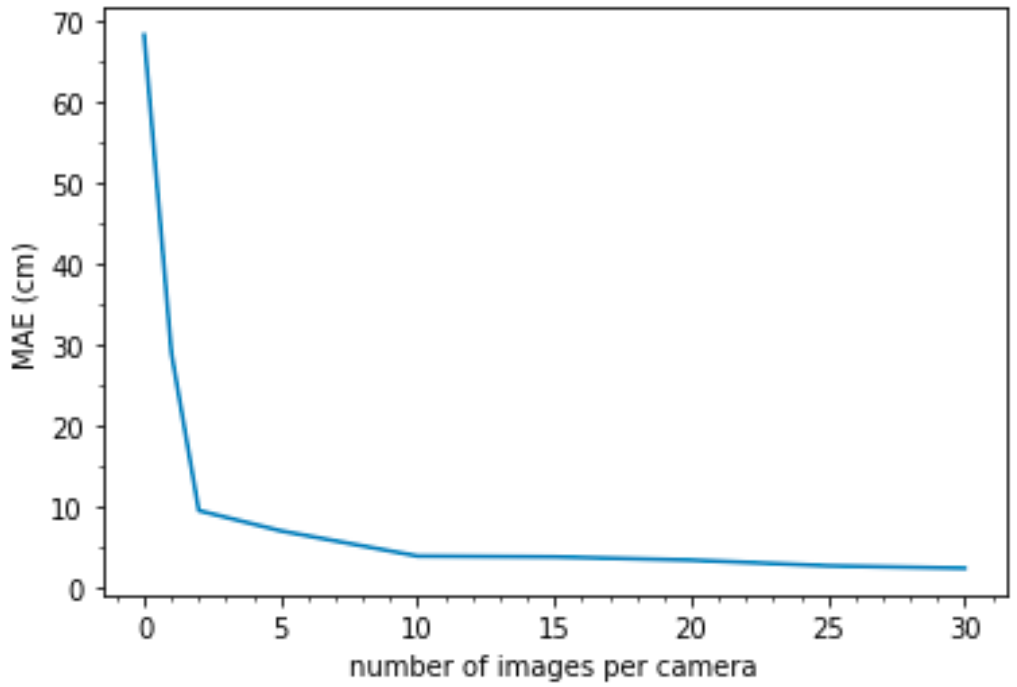


Figure 3.6 Performance of the fine-tuned model improved with more training data. The *x*-axis indicates how many images from each of the 12 Washington cameras was used during training. The *y*-axis shows the resultant mean absolute error (cm) when testing the model on the remainder of the data set.

The increase in accuracy of processing the Washington images, however, came with a decrease in accuracy of processing the Colorado images. Unlike the original Colorado model, the fine-tuned model predicted snow depth in the Colorado images with a MAE of 78.1 cm and a R^2 value equal to 0.52. When looking at the time series at each camera, the model was able to capture the trend of the snow series at each Colorado camera, but had a consistent offset with the actual snow depth and overestimation at high snow depth (Appendix B, Figure B8). After combining both the Washington and Colorado images in training, the "CO + WA model," the

model achieves accuracy within a MAE of 2.05 cm for the Colorado data set and 1.14 cm for the Washington data set (Table 3.2).

3.4 Discussion

We present a model trained on time-lapse camera images to detect the top and bottom of snow poles. Depth in pixels was subsequently converted to centimeters using each pole's pixel to centimeter conversion. While accuracy is low on pole sites that the model did not see during training, the model demonstrates the ability to successfully track snow depths using time-lapse camera images as long as the model was trained on pole sites presented during training (MAE value equal to 1.30 cm). The accuracy is low when using images from a novel data set at a new location, but an additional training step on a small fraction (6.7%) improved the model's performance to a MAE value equal to 4.0 cm, compared to a MAE equal to 6.42 cm using the non-pre-trained model. Neural network models thus show promise for snow pole detection, and keypoint detection models are advantageous methods for large-scale camera trap installations when a subset of the data can be labeled for model fine-tuning or model retraining.

3.4.1 *Model Performance and Impacts on Accuracy*

Our model had measurement errors that were comparable to those of other methods used to measure snow depth (lidar ~± 10 cm, manual snow probing ~1–2 cm) (Deems et al., 2013; Prokop et al., 2008; Sturm & Holmgren, 2018). Previous automated and semi-automated methods for obtaining snow depth from time-lapse imagery report accuracy within 4 and 10 cm (Bongio et al., 2021; Garvelmann et al., 2013; Strickfaden et al., 2023), and our model trained on the Colorado data set shows that this keypoint detection model performs with high accuracy and

across more poles when presented with the pole site during training. The performance corresponds to the success of other keypoint models in detecting face features and estimating poses (K. Wang et al., 2010).

Results from the Colorado model provides useful insights into the optimal snow pole placement. Model performance did not change significantly on poles further away, a contrast to other findings that suggest that objects further away from the camera are harder to detect (Norouzzadeh et al., 2018). During installations, crews were instructed to place poles as close to cameras as possible while allowing for pole and camera tilt. The distance was possibly not far enough to impact model performance. Additionally, camera installers were instructed to ensure no vegetation blocked the view between the camera and the pole. Vegetation and background clutter is one of the reasons models perform more poorly on objects further away (Borji et al., 2015). Given these reasons, future work should systemically explore if there is a distance after which performance drops. Even though we did not detect an overall effect on the distance of pole on model performance, we noticed that when the pole is further away, fewer pixels represent each centimeter, and error in model predictions are enlarged when converting from pixels to centimeters. This error could be avoided by having poles close to the camera to maximize accuracy and/or a standardized protocol to place a pole a certain distance away from the camera. While we did not investigate an exact distance, we noticed that including a buffer of 25% of the pole height above and below it in the image, enabled consistent observations even if there was pole tilt. Although a standardized approach may be difficult in variable or rocky terrain and in deep snowpacks (poles have to be placed further away in deeper snowpacks because snow will be higher), NASA SnowEx provides installation instructions for installing a time-lapse camera without securing to a tree (Breen et al., 2022), which allows for pole and camera

placement flexibility. Other impacts included month of year and snow depth, with shoulder winter months, lower snow depths (<80 cm), and higher snow depths (above 160 cm) demonstrating lower accuracy. The lower accuracy is likely a result of changes in lighting condition, vegetation, and novel snow conditions (i.e., patchy snow or maximum snow depth values) adding variability to the data with fewer examples in the training set. Although common in neural network applications, it is difficult to know the exact reason why performance drops in these conditions, and the inability to fully interpret model predictions is one drawback of this method (Liang et al., 2021; Lucas, 2020). However, given that the error is below 5 cm for all cameras, when snow depth is binned in 20 cm increments, and all months but September, the method demonstrates the ability to predict snow depth values in a variety of conditions and pole placements.

Other error may result from using imagery to derive snow depth and not caused by the model. For example, we made the pixel-to-centimeter conversion as accurate as possible by using the full length of the pole and repeating the process multiple times, but there could be unknown systematic differences due to human error or the pole tilting in a direction not discernible in a 2D image (i.e., away or toward the camera). Other studies have employed intermediate markings on the pole (Cosgrove et al., 2021), and intermediate markings could provide an additional source of validation. For this model, markings on the pole would likely not interfere with the model to identify the top and bottom of the pole, but we recommend presenting these instances to the model ahead of time for fine-tuning. We did not account for lens distortion in the model, and we found effects from lens distortion were likely minimal. Previous work on lens distortion in time-lapse cameras suggests that objects are less distorted when they are in the middle of image and close to the camera (Leorna et al., 2022). The lack of impact suggests that

for the camera used in this case (Wingscape) and orientation of the poles (i.e., vertical and in the middle of the image), lens distortion had minimal effects on the accuracy of the resulting snow depth. However, we recommend installing poles in the center of the camera frame although sometimes this it not possible if cameras are set up on game trails. These caveats should be considered when deciding pole design and camera type, and ultimately depend on the goal of the study.

3.4.2 *Current Model Generalizability to New Data Sets*

Model performance is strong when this model is trained on the data set of interest, but accuracy when generalizing to new sites without incorporating additional training data is low. Generalizing to a new data set is an active area of research when predicting information from camera trap images using neural networks, such as the recent work to develop models that can identify animal species across different habitats (Norman et al., 2023). We speculate that the low performance of our original Colorado model on Washington images is because the model likely fit to the placement of Colorado poles sites depending on the corresponding background conditions, as suggested by previous work that the accuracy of deep learning networks depends on the image backgrounds presented during training (Beery et al., 2018; Miao et al., 2019). Therefore, the model with the Colorado pre-trained weights is not advised "off-the-shelf," unless identifying snow depth from the Grand Mesa sites as presented here.

However, fine-tuning the Colorado model using a subset of images from Washington improved model performance. This approach is recommended by other keypoint detection models to improve performance on data of interest (Mathis et al., 2021). Overall accuracy is similar to other automated methods, and better than a model without the pre-trained Colorado

weights. The model using the pre-trained weights and fine-tuned with Washington images also reached the criteria for early stopping 50 epochs sooner than the model trained on the Washington images without the pre-trained Colorado weights. Some model error arose due to new conditions not seen frequently in the Colorado data set, such as snow mounds (Appendix B, Figure B7). Washington snow is considered wetter and more maritime relative to drier, inner-continental Colorado snow (Sturm & Liston, 2021). These conditions make snow "stickier" and more likely to form mounds on snow poles. The Washington poles were also wider, which could make mounds more apparent or more likely to form. While the model predicted the approximate place that the pole intercepted the snowpack underneath the snow mound, it did not always match the training data. It is possible that the training data is incorrect if the labeler incorrectly estimated snow depth due to the snow mound. Regardless, the effects of snow loading can be temporary, so we encourage the user to set multiple time-lapse images per day to reduce the effects of snow loading and weather, and stratify data labeling to present the model with a variety of conditions throughout the winter season.

We recommend that before using the Colorado pre-trained model, users add a training step to the Colorado model using a subset of their own data, as we demonstrated here. This step will result in a custom keypoint detection model for the user's snow pole network. Based on the results from this study, a series of images separated by 10 days from each camera would be sufficient for fine-tuning the model. More images per camera increased overall accuracy of the customized model, and users can label depending on their desired accuracy. Figure B9 in Appendix B also shows that adding an image at high snow depth along with 10 images per camera may increase the accuracy at high snow depth values. To maintain performance accuracy

across all sites, our results suggest it is better to combine both sets of images and re-train the model.

3.4.3 *Limitations of Neural Networks and Future Work*

Despite the high accuracy for data of interest, neural network models have limitations. Training this model required ~$500 USD of GPU resources, ~$500 USD of VM storage, and image preprocessing steps (i.e., downscaling) that introduced uncertainty. For users to apply the model, we decreased GPU need by fine-tuning on a CPU machine, but this step still requires labeling roughly ~10 images per camera, which would require ~1 hr of labeling for a network of 20–30 cameras. We provided a README and scripts in the GitHub repository, including labeling.py, train.py, and predict.py to facilitate use, but users of this model will also need basic GitHub and command line knowledge.

Despite these considerations, neural networks offer advantages to pre-existing manual methods when it comes to generalization and scaling. While customizing the pre-trained Colorado model using a dataset of interest sacrificed performance in the Colorado model, fitting with literature that fine-tuning can affect model performance of the pre-trained data sets (Kumar et al., 2022; Lee et al., 2023), this loss was avoided when Colorado and Washington images were combined in model training. In an ancillary test, we examined model performance when training the model on all but two poles from Washington and Colorado (i.e., one pole from Washington and one pole from Colorado in the test set). Model performance on each pole was low (MAE = 20.4 cm; Appendix B, Table B5), but higher than when the model had no regional information. We refrained from further testing this design due to low accuracy, but this finding underscores that the model may learn information from the surrounding ecosystem, and future

models may require less fine-tuning. A future model could thus integrate images across multiple experiments to create a model applicable to various domains. MegaDetector (Beery, Morris, & Yang, 2019), a publicly available object detector model to identify animals in camera trap images, is updated regularly to include new user data according to the codebase GitHub page. In future iterations of this model, the model could be trained on large databases of publicly available time-lapse snow pole data sets to achieve accuracy across multiple domains. Many studies reference snow pole data sets (Collados-Lara et al., 2020; Cosgrove et al., 2021; Currier et al., 2016; Hofmeester et al., 2019; McCreight et al., 2014), but currently no standardized labeling framework or comprehensive database exists to consolidate across sources as in other domains (e.g., ImageNet; Deng et al., 2009). Future work could explore publishing a database of multiple datasets or a standardized labeling framework, lowering the threshold for future model training and testing. Additionally, other neural network designs should be considered for generalization. Segmentation models have been used in other cases for snow pole detection (Kopp et al., 2019). Segmentation works to classify each pixel and has also been used to identify objects of interest in new domains (Beery et al., 2018). Our data labels can be easily converted to segmentation labels by converting the space between the top and bottom of the pole to positive pixels and all other surrounding pixels to negative pixels, in order to explore such comparisons between other model architectures. A consolidated database of snow poles from various studies would further ease comparison across model architectures.

Future work could also explore adding corrections to the model to improve real-world application. We allowed the model to predict negative snow depth values, and one simple improvement is to add a correction that requires the model to replace the negative value with 0 cm. Other machine learning models for ecology applications apply environmental context to

improve predictions. For example, computer vision models incorporate geographic information on species abundance to control species-ID predictions (Durso et al., 2021). Moreover, we did not crop the images during data preparation. This choice prevented additional work by the user if a pole or camera tilted during the season, but future work could explore whether cropping images improves model accuracy. Going forward, this model could be combined with an object detection model, where the model first identifies the pole (or poles) in an image, returns the crop, then the keypoint detection model predicts the snow depth. This pipeline would mimic other machine learning models for ecology where simpler classification models are developed first and then iterated on by more advanced models (Beery, Wu, et al., 2019). Furthermore, we cleaned both data sets for cameras or poles that tilted beyond the view of the pole, but future work could explore extracting snow depth using the bottom detected point even if the top is not visible by identifying the location of the bottom point relative to the location of the bottom point in a snow-free image. Additionally, while we determined our conversion by first manually inspecting 10 images and determining the number of pixels that represented the whole pole, this step, too, could be automated. The user could input the length of the full pole along with an image of a snow-free stake, and the model could calculate the conversion using the predicted length of the snow-free pole in pixels.

Finally, when generalizing, it will be important to account for circumstances not shown in this study. We did not test for the impact of nighttime, or low-light, on the model performance. Many snow monitoring locations are in the polar north, where long nights are common, and we predict that without nighttime examples in the training data set, performance will be low. We encourage future work to incorporate both daytime and nighttime images into the training set and testing for accuracy depending on the time of day. Kopp et al. (2019)

demonstrates that when night images are presented to the model during training, nighttime MAE error is slightly higher but under 5 cm. In the meantime, the model is applicable to study designs that can set cameras to daily midday time-lapse images, such as 12p.m., to maximize daytime conditions. Similarly, we used red poles for model training, but more work is needed to determine whether the model could learn other snow pole designs. White poles, for example, require less user preparation (i.e., PVC poles are usually white), and they can more easily blend into snowy backdrops (Cosgrove et al., 2021). Striped poles are also advantageous because they provide intermediate depth markings to compare to the depth in pixels (Garvelmann et al., 2013; Hofmeester et al., 2019). Future iterations of the model could even incorporate virtual snow pole approaches, such as in Strickfaden et al., 2023, and superimpose the virtual snow pole onto the image of interest and select the top and bottom of the virtual pole for the training data. We call for future work to integrate multiple data sets from various ecosystems, multiple years, and pole set-ups into one model, thereby leveraging a larger sample size to improve accuracy on new data sets and increasing applicability to currently existing data sets. We present this model as the first step into what we hope is an expansion and integration of snow pole data across multiple platforms into a single model.

3.5 Conclusion

Snow is an important component of hydrological and ecological systems, and improved automation will accelerate our ability to monitor snow. Future users can fine-tune this model using a subset of their own data, creating a custom keypoint detection model for their snow pole network, or future versions of the model can integrate multiple data sets during training for improved performance across multiple data sets. With the increasing use of snow poles to

monitor snow depth, the ability of a keypoint model to automate snow poles provides a foundation for a future generalizable model.

3.6 Data Availability Statement

The images used for the model in the study are available at National Snow and Ice Data Center via https://doi.org/10.5067/P3D1QRH7O8X5; The snow depth values used for the model in the study are available at the National Snow and Ice Data Center via https://doi.org/10.5067/14EU7OL5051V; Code used for the model is preserved at https://github.com/catherine-m-breen/snowpoles via the MIT license.

3.7 References

Bajramović, M., & Žunić, E. (2023). Image classification with transfer learning using a convolutional neural network. In *2023 22nd international symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1–5). IEEE. https://doi.org/10.1109/INFOTEH57020.2023.10094197

Bassing, S. B., DeVivo, M., Ganz, T. R., Kertson, B. N., Prugh, L. R., Roussin, T., et al. (2023). Are we telling the same story? Comparing inferences made from camera trap and telemetry data for wildlife monitoring. *Ecological Applications, 33*(1). https://doi.org/10.1002/eap.2745

Beery, S., Morris, D., & Yang, S. (2019). Efficient pipeline for camera trap image review. *arXiv*. Retrieved from http://arxiv.org/abs/1907.06772

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita. In V. Ferrari, M.

Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (Vol. 11220, pp. 472–489). Springer International Publishing. https://doi.org/10.1007/978-3-030-01270-0_28

Beery, S., Wu, G., Rathod, V., Votel, R., & Huang, J. (2019). Long term temporal context for per-camera object detection. *arXiv:1912.03538 [Cs,Eess, q-Bio]*. Retrieved from http://arxiv.org/abs/1912.03538

Bongio, M., Arslan, A. N., Tanis, C. M., & De Michele, C. (2021). Snow depth time series retrieval by time-lapse photography: Finnish and Italian case studies. *The Cryosphere, 15*(1), 369–387. https://doi.org/10.5194/tc-15-369-2021

Borji, A., Cheng, M.-M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE Transactions on Image Processing, 24*(12), 5706–5722. https://doi.org/10.1109/TIP.2015.2487833

Breen, C. M., Vuyovich, C., & Hiemstra, C. (2022). SnowEx20 grand mesa snow depth from snow pole time-lapse imagery, version 1 [Dataset]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/14EU7OLF051V

Broxton, P. D., Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving snow water equivalent maps with machine learning of snow survey and lidar measurements. *Water Resources Research, 55*(5), 3739–3757. https://doi.org/10.1029/2018WR024146

Cao, L., & Shen, H. (2022). CSS: Handling imbalanced data by improved clustering with stratified sampling. *Concurrency and Computation: Practice and Experience, 34*(2), e6071. https://doi.org/10.1002/cpe.6071

Cao, Q., Painter, T. H., Currier, W. R., Lundquist, J. D., & Lettenmaier, D. P. (2018). Estimation

of precipitation over the OLYMPEX domain during winter 2015/16. *Journal of Hydrometeorology, 19*(1), 143–160. https://doi.org/10.1175/JHM-D-17-0076.1

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58. https://doi.org/10.1145/1541880.1541882

Collados-Lara, A., Pardo-Igúzquiza, E., & Pulido-Velazquez, D. (2020). Optimal design of snow stake networks to estimate snow depth in an alpine mountain range. *Hydrological Processes, 34*(1), 82–95. https://doi.org/10.1002/hyp.13574

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

Currier, W. R., Lundquist, J., & Njissen, B. (2016). *An independent evaluation of frozen precipitation from the WRF model and PRISM in the Olympic Mountains for WY 2015 and 2016.* University of Washington. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/38604/Currier_washington_0250O_16777.pdf?sequence=1

Deems, J. S., Painter, T. H., & Finnegan, D. C. (2013). Lidar measurement of snow depth: A review. *Journal of Glaciology, 59*(215), 467–479. http://doi.org/10.3189/2013JoG12J154

De Lannoy, G., Vanrykel, A., Lievens, H., Kim, E., & Brucker, L. (2018). Snow estimation under a vegetation gradient using satellite remote sensing data and land surface modeling during snowex 2017. In IGARSS 2018 - 2018 IEEE international geoscience and remote sensing symposium (pp. 6294–6297). IEEE. https://doi.org/10.1109/IGARSS.2018.8517893

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale

hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848

Durso, A. M., Moorthy, G. K., Mohanty, S. P., Bolon, I., Salath., M., & Ruiz De Castañeda, R. (2021). Supervised learning computer vision benchmark for snake species identification from photographs: Implications for herpetology and global health. *Frontiers in Artificial Intelligence, 4*, 582110. https://doi.org/10.3389/frai.2021.582110

Dwivedi, R., Biederman, J. A., Broxton, P. D., Lee, K., & Van Leeuwen, W. J. D. (2023). Snowtography quantifies effects of forest cover on net water input to soil at sites with ephemeral or stable seasonal snowpack in Arizona, USA. *Ecohydrology, 16*(2). https://doi.org/10.1002/eco.2494

Floyd, W., & Weiler, M. (2008). Measuring snow accumulation and ablation dynamics during rain-on-snow events: Innovative measurement techniques. *Hydrological Processes, 22*(24), 4805–4812. https://doi.org/10.1002/hyp.7142

Garvelmann, J., Pohl, S., & Weiler, M. (2013). From observation to the quantification of snow processes with a time-lapse camera network. *Hydrology and Earth System Sciences, 17*(4), 1415–1429. https://doi.org/10.5194/hess-17-1415-2013

Genthon, C., Six, D., Scarchilli, C., Ciardini, V., & Frezzotti, M. (2016). Meteorological and snow accumulation gradients across Dome C, East Antarctic plateau: Meteorological and snow accumulation gradients at Dome C. *International Journal of Climatology, 36*(1), 455–466. https://doi.org/10.1002/joc.4362

Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring:

Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics, 41*, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv*. Retrieved from http://arxiv.org/abs/1512.03385

Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., Andrén, H., et al. (2019). Using by-catch data from wildlife surveys to quantify climatic parameters and the timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation, 6*(2), 129–140. https://doi.org/10.1002/rse2.136

Hojatimalekshah, A., Uhlmann, Z., Glenn, N. F., Hiemstra, C. A., Tennant, C. J., Graham, J. D., et al. (2021). Tree canopy and snow depth relationships at fine scales with terrestrial laser scanning. *The Cryosphere, 15*(5), 2187–2209. https://doi.org/10.5194/tc-15-2187-2021

Jabbar, H. K., & Khan, R. Z. (2014). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). In *Computer science, communication and instrumentation devices* (pp. 163–172). Research Publishing Services. https://doi.org/10.3850/978-981-09-5247-1_017

Jonas, T., Marty, C., & Magnusson, J. (2009). Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology, 378*(1–2), 161–167. https://doi.org/10.1016/j.jhydrol.2009.09.021

Kelly, R. E., Chang, A. T., Tsang, L., & Foster, J. L. (2003). A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 41*(2), 230–242. https://doi.org/10.1109/TGRS.2003.809118

Kochura, Y., Gordienko, Y., Taran, V., Gordienko, N., Rokovyi, A., Alienin, O., & Stirenko, S.

(2020). Batch size influence on performance of graphic and tensor processing units during training and inference phases. In Z. Hu, S. Petoukhov, I. Dychka, & M. He (Eds.), *Advances in computer science for engineering and education II* (Vol. 938, pp. 658–668). Springer International Publishing. https://doi.org/10.1007/978-3-030-16621-2_61

Kohler, J., Brandt, O., Johansson, M., & Callaghan, T. (2006). A long-term Arctic snow depth record from Abisko, northern Sweden, 1913–2004. *Polar Research, 25*(2), 91–113. https://doi.org/10.3402/polar.v25i2.6240

Kopp, M., Tuo, Y., & Disse, M. (2019). Fully automated snow depth measurements from time-lapse images applying a convolutional neural network. *Science of the Total Environment, 697*, 134213. https://doi.org/10.1016/j.scitotenv.2019.134213

Kumar, A., Raghunathan, A., Jones, R., Ma, T., & Liang, P. (2022). Fine-tuning can distort pretrained features and underperform out-of distribution. *arXiv*. Retrieved from http://arxiv.org/abs/2202.10054

Lee, Y., Chen, A. S., Tajwar, F., Kumar, A., Yao, H., Liang, P., & Finn, C. (2023). Surgical fine-tuning improves adaptation to distribution shifts. *arXiv*. Retrieved from http://arxiv.org/abs/2210.11466

Leorna, S., Brinkman, T., & Fullman, T. (2022). Estimating animal size or distance in camera trap images: Photogrammetry using the pinhole camera model. *Methods in Ecology and Evolution, 13*(8), 1707–1718. https://doi.org/10.1111/2041-210X.13880

Liang, Y., Li, S., Yan, C., Li, M., & Jiang, C. (2021). Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing, 419*, 168–182. https://doi.org/10.1016/j.neucom.2020.08.011

Lucas, T. C. D. (2020). A translucent box: Interpretable machine learning in ecology. *Ecological*

Monographs, 90(4). https://doi.org/10.1002/ecm.1422

Maile, R. E., Duggan, M. T., & Mousseau, T. A. (2023). The successes and pitfalls: Deep-learning effectiveness in a Chernobyl field camera trap application. *Ecology and Evolution*, 13(9), e10454. https://doi.org/10.1002/ece3.10454

Marshall, H. P., Deeb, E., Forster, R., Vuyovich, C., Elder, K., Hiemstra, C., & Lund, J. (2021). L-band InSAR depth retrieval during the NASA SnowEx 2020 campaign: Grand Mesa, Colorado. In *2021 IEEE international geoscience and remote sensing symposium IGARSS* (pp. 625–627). IEEE. https://doi.org/10.1109/IGARSS47720.2021.9553852

Martin, M. E., Moriarty, K. M., & Pauli, J. N. (2020). Forest structure and snow depth alter the movement patterns and subsequent expenditures of a forest carnivore, the Pacific marten. *Oikos*, 129(3), 356–366. https://doi.org/10.1111/oik.06513

Mathis, A., Biasi, T., Schneider, S., Yuksekgonul, M., Rogers, B., Bethge, M., & Mathis, M. W. (2021). Pretraining boosts out-of-domain robustness for pose estimation. In *2021 IEEE winter conference on applications of computer vision (WACV)* (pp. 1858–1867). IEEE. https://doi.org/10.1109/WACV48630.2021.00190

McCreight, J. L., Small, E. E., & Larson, K. M. (2014). Snow depth, density, and SWE estimates derived from GPS reflection data: Validation in the western U. S. *Water Resources Research, 50*(8), 6892–6909. https://doi.org/10.1002/2014WR015561

Meek, P. D., Ballard, G. A., Sparkes, J., Robinson, M., Nesbitt, B., & Fleming, P. J. S. (2019). Camera trap theft and vandalism: Occurrence, cost, prevention and implications for wildlife research and management. *Remote Sensing in Ecology and Conservation, 5*(2), 160–168. https://doi.org/10.1002/rse2.96

Miao, Z., Gaynor, K. M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M. S., et al. (2018).

A comparison of visual features used by humans and machines to classify wildlife (preprint). *Ecology*. https://doi.org/10.1101/450189

Miao, Z., Gaynor, K. M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M. S., et al. (2019). Insights and approaches using deep learning to classify wildlife. *Scientific Reports, 9*(1), 8137. https://doi.org/10.1038/s41598-019-44565-w

Norman, D. L., Bischoff, P. H., Wearn, O. R., Ewers, R. M., Rowcliffe, J. M., Evans, B., et al. (2023). Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods in Ecology and Evolution, 14*(1), 242–251. https://doi.org/10.1111/2041-210X.14031

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution, 12*(1), 150–161. https://doi.org/10.1111/2041-210X.13504

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences, 115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images - Potential for snow cover mapping: Spatio-temporal combination of MODIS images. *Water Resources Research, 44*(3). https://doi.org/10.1029/2007WR006204

Parajka, J., Holko, L., Kostka, Z., & Blöschl, G. (2012). MODIS snow cover mapping accuracy in a small mountain catchment – Comparison between open and forest sites. *Hydrology and Earth System Sciences, 16*(7), 2365–2377. https://doi.org/10.5194/hess-16-2365-2012

Pflug, J. M., & Lundquist, J. D. (2020). Inferring distributed snow depth by leveraging snow pattern repeatability: Investigation using 47 lidar observations in the Tuolumne watershed, Sierra Nevada, California. *Water Resources Research, 56*(9). https://doi.org/10.1029/2020WR027243

Prokop, A., Schirmer, M., Rub, M., Lehning, M., & Stocker, M. (2008). A comparison of measurement methods: Terrestrial laser scanning, tachymetry and snow probing for the determination of the spatial snow-depth distribution on slopes. *Annals of Glaciology, 49*, 210–216. https://doi.org/10.3189/172756408787814726

Robles, M., Payton, E., & Biederman, J. (2021). Snowtography: Snowpack & soil moisture monitoring handbook. https://doi.org/10.25810/R9S7-4T28

Rowcliffe, J. M., Field, J., Turvey, S. T., & Carbone, C. (2008). Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology, 45*(4), 1228–1236. https://doi.org/10.1111/j.1365-2664.2008.01473.x

Sanmiguel-Vallelado, A., Morán-Tejeda, E., Alonso-González, E., & López-Moreno, J. I. (2017). Effect of snow on mountain river regimes: An example from the Pyrenees. *Frontiers of Earth Science, 11*(3), 515–530. https://doi.org/10.1007/s11707-016-0630-z

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution, 10*(7), 3503–3517. https://doi.org/10.1002/ece3.6147

Schneider, S., Taylor, G. W., & Kremer, S. C. (2018). Deep learning object detection methods for ecological camera trap data. arXiv:1803.10842 [Cs]. http://arxiv.org/abs/1803.10842

Schöner, W., Auer, I., & B.hm, R. (2009). Long term trend of snow depth at Sonnblick (Austrian

Alps) and its relation to climate change. *Hydrological Processes, 23*(7), 1052–1063. https://doi.org/10.1002/hyp.7209

Simões, F., Bouveyron, C., & Precioso, F. (2023). DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning. *Ecological Informatics, 75*, 102095. https://doi.org/10.1016/j.ecoinf.2023.102095

Sirén, A. P. K., Somos-Valenzuela, M., Callahan, C., Kilborn, J. R., Duclos, T., Tragert, C., & Morelli, T. L. (2018). Looking beyond wildlife: Using remote cameras to evaluate accuracy of gridded snow data. *Remote Sensing in Ecology and Conservation, 4*(4), 375–386. https://doi.org/10.1002/rse2.85

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay. *arXiv*. Retrieved from http://arxiv.org/abs/1803.09820

Strickfaden, K. M., Behan, M. L., Marshall, A. M., Svancara, L. K., Ausband, D. E., & Link, T.E. (2023). Virtual snow stakes: A new method for snow depth measurement at remote camera stations. *Wildlife Society Bulletin, e1481*(3). https://doi.org/10.1002/wsb.1481

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.* https://doi.org/10.1175/JHM-D-21-0070.1

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., & Lea, J. (2010). Estimating snow

water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology, 11*(6), 1380–1394. https://doi.org/10.1175/2010JHM1202.1

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications, 13*(1), 792. https://doi.org/10.1038/s41467-022-27980-y

Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., et al. (2022). Generalizing to unseen domains: A survey on domain generalization. *arXiv*,1. https://doi.org/10.1109/tkde.2022.3178128

Wang, K., Franklin, S. E., Guo, X., & Cattet, M. (2010). Remote sensing of ecology, biodiversity and conservation: A review from the perspective of remote sensing specialists. *Sensors, 10*(11), 9647–9667. https://doi.org/10.3390/s101109647

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *WIREs Water, 8*(5), e1533. https://doi.org/10.1002/wat2.1533

Zhang, J., Chen, Z., & Tao, D. (2021). Towards high performance human keypoint detection. *International Journal of Computer Vision, 129*(9), 2639–2662. https://doi.org/10.1007/s11263-021-01482-8

Chapter 4. DIEL AND SEASONAL CYCLES OF SNOW HARDNESS PREDICT CHANGES IN ACTIVITY PATTERNS FOR ROE DEER BUT NOT MOUNTAIN HARES

Publication history: This study was co-authored with John Odden, John Linnell, Neri Thorsen, Inger Maren Rivrud and Laura Prugh. At the time this dissertation was submitted, no version of this manuscript was in review, accepted, or published elsewhere.

Abstract: The strength of the snow surface, or snow hardness, is an important factor for wildlife movement, and there is evidence that snow hardness displays dramatic changes over the course of the day throughout the course of the winter season. However, it is unknown how snow diel cycles change in timing and magnitude throughout the winter season, and how animals respond to temporal changes in snow hardness. Using ~250 hourly surface hardness measurements along a latitudinal gradient in Norway, we observe and model diel cycles of snow hardness in relation to environmental variables. Time relative to sunrise and sunset, season, canopy cover, and snow density were the best predictors of surface hardness (R^2 = 0.63). Snow hardness peaked 8 hours after sunset in the spring and fluctuated most dramatically during spring and in open habitat. We matched diel cycles of predicted surface hardness to diel activity in roe deer (*Capreolus capreolus*) and mountain hares (*Lepus timidus*) between 2017 - 2023 using data from the large-scale Scandcam camera-trap network, located in Norway (n = 698 cameras). Roe deer demonstrated increased morning activity when the snow was harder and deeper and preferred harder snow at higher snow depths. Hares did not show diel shifts in relation to snow hardness and preferred softer snow at higher snow depths. As climate change influences snow conditions,

this study highlights the importance of morphological differences between species for on-snow movement, and diel activity adaptations to environmental conditions.

Keywords: snow, diel activity, snow hardness, roe deer, mountain hare

4.1 Introduction

Snow conditions play a critical role in shaping wildlife behavior (Penczykowski et al. 2017), with strong evidence for the influence of snow depth on movement (Mysterud et al. 1997, Pedersen et al. 2021). Much less studied is the influence of the snow surface strength, or snow hardness, on wildlife movement (Mahoney et al. 2018, Boelman et al. 2019). Previous work has demonstrated the impact of snow hardness on the timing of deer migration in the spring (Jakopak et al. 2021), foraging abilities (Collins and Smith 1991, Hansen et al. 2019), hunting success (Stenseth et al. 2004, Sullender et al. 2023), and locomotive abilities in both small and large mammals (Sivy et al. 2018, Poirier et al. 2021, Sullender et al. 2023). As warming winter temperatures increase the frequency of mid-winter freeze-thaw days (Cohen et al. 2015, Mooney and Li 2021), an improved understanding of snow hardness, and its impact on wildlife, becomes increasingly important.

Snow hardness varies substantially across the landscape, and the heterogeneity can strongly affect wildlife movements in winter (Murray and Boutin 1991a, Ossi et al. 2015, Pozzanghera et al. 2016, Mahoney et al. 2018). However, snow hardness can also change markedly on an hourly time scale, especially during the late winter-spring season (U.S. Army Corps of Engineers 1956, Ozeki and Akitaya 1998, Cagnati et al. 2004, Brogioni et al. 2009). Incoming shortwave solar radiation and above-zero temperatures during the day increase liquid water content in the snowpack. On cold nights, this liquid water subsequently refreezes into an icy crust, sometimes

referred to as the “nocturnal snow crust” (U.S. Army Corps of Engineers 1956), providing a hard and supportive surface until it melts again the next day. However, empirical studies of snow hardness over diel cycles are lacking. De Michele et al. (2013) and Helfricht et al. (2018) measured hourly changes in snow density (mass per volume) and found that it varied with air temperature, humidity, wind, and surface temperature, but neither study measured snow hardness directly. Density can be poorly related to hardness, especially in spring when slushy snow can have high water content (i.e., high density) but low strength. Canopy cover may also impact snow hardness as longwave radiation from trees will add to the energy in the system (Varhola et al. 2010). Given that wildlife can shift their diel activity patterns in response to external factors such as human activity (Gaynor et al. 2018), measurement of snow hardness at hourly scales may reveal new insights into the complex drivers of animal activity patterns.

In this study, we present observations of hourly snow surface hardness in a boreal forest ecosystem. We first quantified how hardness changed over the 24-hour diel cycle, examining the influence of forest canopy cover and season. We expected to find a lag between sunset and peak snow hardness because time is required for the energy exchange between the air and the snowpack surface. We also expected a larger amplitude of change in hardness in open habitats and during the spring season compared to winter due to the increased daytime solar radiation and temperatures. We examined patterns in activity over the diel cycle using data from a Norwegian wildlife camera trap network (Scandcam, viltkamera.nina.no), comprising detections from 698 cameras with deployment dates from 2017 - 2023. We focused analysis on two common herbivore species, roe deer (*Capreolus capreolus*) and the mountain hare (*Lepus timidus*). These species have overlapping diets and are both prey for the Eurasian lynx (*Lynx lynx*) but have contrasting morphologies in relation to snow. Roe deer, which are relatively small ungulates,

have high foot loading (the ratio of body mass to foot surface area) ranging from 300-600 g/cm^2 (Kelsall 1969). In contrast, mountain hares have foot loading estimates of 15-20 g/cm^2 (Hackländer and Zachos 2020). We predicted that animals would be most active when the snow was hardest and thus easiest to walk on. We expected this effect to be stronger for roe deer than hares due to differences in foot-loading, and we also expected snow hardness to affect movements most strongly at sites with deeper snow (Mahoney et al. 2018).

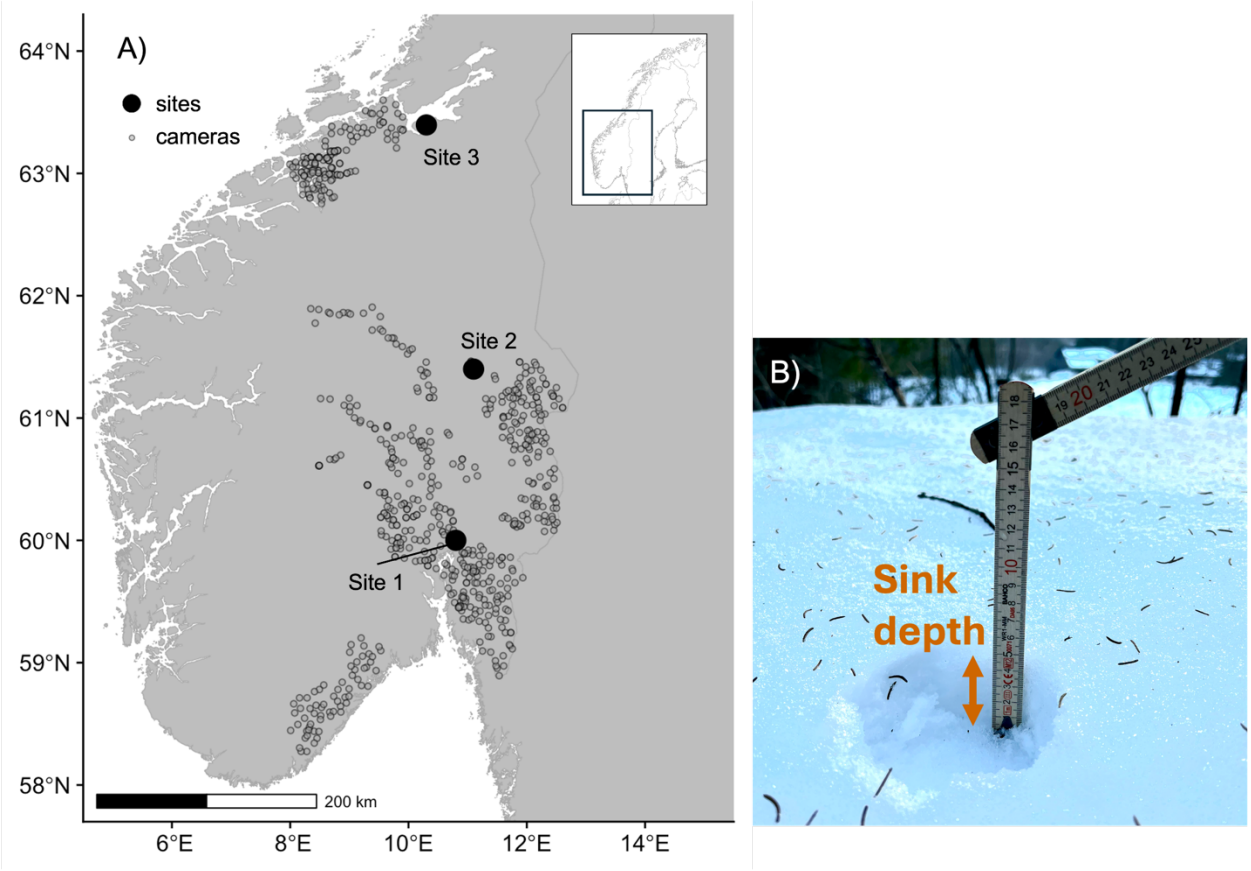


Figure 4.1. A) Sites for snow hardness measurements. We conducted 24-hour observation periods across the late winter and spring season (30 January to 15 May 2023) at three sites in Norway along a latitudinal gradient (black dots). The sites were located within the Scandcam camera trap network region (viltkamera.nina.no; grey dots). B) Example of snow hardness

measurement taken by measuring the sinking depth of a 200 g cylindrical penetrometer (i.e., tuna can) dropped from 50 cm.

4.2 Methods

4.2.1 *Study Area*

Snow measurements were carried out at three sites along a 3.5° latitudinal gradient in southern Norway (Fig. 4.1; Table 4.1). These sites fell within the southern portion of a large-scale camera trap network, Scandcam. Scandcam is a long-term, year-round study established in 2010 by the Norwegian Institute for Nature Research to monitor recovering Eurasian lynx (Carricondo-Sanchez et al. 2017). The southern portion of the network totaled 698 cameras. Our snow measurement sites captured the variation in physical factors influencing snowpack conditions at the surrounding cameras, including differences in latitude, elevation, aspect, and climate. Because the cameras were deployed for lynx monitoring, all cameras were in lynx habitat, mostly boreal forests. Boreal forests in Norway are characterized by Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) and subject to clear-cut forestry practices, so they can vary dramatically in the percentage of canopy cover. Elevations of the cameras in the south (58°N – 60.5°N) ranged between 0 - 400 m ($\bar{x}$ = 244.9 m, n = 306 cameras). Further north and inland (60.6°N – 61.5°N), cameras ranged between 200 and 850 m ($\bar{x}$ = 443.8 m, n = 204 cameras), and cameras furthest north (61.6°N - 64°N) ranged between 0 and 400 m ($\bar{x}$ = 155.4 m, n = 188 cameras). The snow measurement sites also reflected a maritime-continental gradient, with sites 1 and 3 representing rainier winters with temperatures often fluctuating around freezing, and site 2 experiencing a more continental climate with winter temperatures typically more stable below freezing (Table 4.1). Roe deer and hares are present in all areas. Other large

mammals present include Eurasian lynx, moose *(Alces alces*), red deer (*Cervus elaphus*), and red foxes (*Vulpes vulpes*). Wolves (*Canis lupus*), wolverines (*Gulo gulo*), wild boars (*Sus scrofa*), and brown bears (*Ursus arctos*) are present in parts of the study area.

Table 4.1. Latitude, elevation, slope, aspect, average winter daily temperature, and average winter daily precipitation information for the three snow measurement sites in southern Norway. Average temperature and average precipitation were obtained from Norway's daily snow model (senorge.no; Saloranta et al. 2012).

Site	Latitude (°)	Elevation (m)	Slope (°)	Aspect (°)	Avg. temp. (°C)	Avg. precip. (cm)
1	60.0	203	8	91	-0.5	4.5
2	61.4	318	16	300	-4.2	2.7
3	63.5	250	6	171	-0.6	4.2

4.2.2 *Data Collection*

Snow data collection at each site occurred between 30 January and 15 May 2023, after snow reached peak depth and until snow melted (Saloranta et al. 2012). During that period, we visited each snow measurement site every two to three weeks. We limited sampling to days when the forecast predicted no precipitation and when daily temperatures were expected to include a freeze-thaw event (i.e., crossed 0°C, hereafter "freeze-thaw day"). We set these parameters because precipitation would interfere with changes in snow hardness, and because we only expected snow hardness to have a diel cycle on days that had a freeze-thaw event.

During each visit, we measured snow hardness every hour in both open and closed canopy areas to capture relationships between canopy and snow hardness (Ellis et al. 2013, Horstkotte and Roturier 2013, Lundquist et al. 2013). Canopy cover was categorized as either "closed" or "open" by field observer estimation. The closed and open canopy sampling occurred within 15-20 m of each other to avoid forest edge effects and to accurately represent the same area. Snow hardness was measured in centimeters using the average of two sinking depth measurements of a 200 g cylindrical penetrometer (7 oz tuna can, Chicken of the Sea Chunk Light) dropped from 50 cm (Sullender et al. 2023). Harder snow corresponded to lower sinking depth values, while softer snow corresponded to higher sinking depth values. Other studies have used similar inexpensive, rapid methods to measure surface snow hardness in wildlife studies (Murray and Boutin 1991a, Pozzanghera et al. 2016, Droghini and Boutin 2018, Peers et al. 2020, Sullender et al. 2023).

Along with each tuna can sinking depth measurement, we recorded date, time of day, air temperature, and snow depth. Air temperature was recorded at 2 m above the ground using a dial stem thermometer (Snow Metrics Dial Stem Thermometer), taken in the shade to block impacts of solar radiation. Snow depth was recorded using a snow probe (Ortovox Snow Probe). We set 10 cm as the minimum snow depth for measurements to avoid effects of latent ground warming, which can affect the 10 centimeter of snow closest to the ground depending on climate conditions (Storebakken 2022). Additionally, at the beginning of the season, we set up two wildlife cameras (Reconynx Hyperfire HC 500) at each site to record camera temperature information at hourly intervals. We positioned one camera facing north and one camera facing south to account for solar radiation that might impact the camera temperature reading (Buchholz et al. 2021). Observations occurred for 24-36 hours to include an overnight observation. The

only site where we did not conduct overnight observations was in the south (site 1: Oslo) due to lack of available overnight lodging accommodations. On subsequent sampling days for all sites, we returned to the same locations both in and out of canopy.

4.2.3 *Snow Hardness Modeling*

We estimated snow hardness using a generalized linear modeling framework for all field observation days that coincided with a freeze-thaw day (i.e., crossed 0ºC). Specifically, we constructed a model that regressed sinking depth from the tuna can penetrometer, our measurement for snow hardness, as a combination of several fixed effects (see below for details) that represented sources of longwave radiation, shortwave radiation, and snow history (Kozak et al. 2003, Lundquist et al. 2013). All fixed effects included interaction terms for canopy cover and season, as previous work suggests that the dominant forces in the snow melt-refreeze cycle can change depending on whether the system is in open versus closed canopy and timing in the snow-covered season (Lundquist et al. 2013). We used Akaike information criterion (AIC) to identify the most parsimonious model for the combination of variables in the following full model:

$$Snow\ hardness \sim \sin(sun\ time) * canopy\ cover * season + \cos(sun\ time) * canopy\ cover * season + snow\ density\ * canopy\ cover * season \quad \text{(Eqn. 4.1)}$$

The variable ‘sun time’ represented time of day relative to sunrise and sunset to account for variable daylengths, transformed from clock time using R’s *overlap* package (Ridout and Linkie 2009). Because of the intrinsic circularity of the time variable, we used sine and cosine transformations to ensure appropriate circular behavior (Chen et al. 2020). We used a modeled

daily snow density value for the fixed effect variable ‘snow density’ from Norway’s daily snow model, which produces predictive outputs at 1-km resolution based on interpolations and downscaling from existing weather stations (senorge.no; Saloranta 2012). Canopy cover was defined as either open or closed using the observations from data collection. Similarly, season was defined as either winter (February and March), or spring (April and May). Both variables were compared *a priori* to continuous alternatives, specifically canopy cover ranging from 0 – 100 and season as both a continuous day of year and month, but binary variables performed better as univariate predictors of sink depth. Binary categorical variables also enabled the model to capture non-linear changes between seasons and canopy cover, whereas using continuous variables would assume linear relationships. We scaled all variables using the *scale* function in R, which subtracts the mean and divides by the standard deviation for each variable, to ensure variables were on similar scales, reduce chances of model convergence issues, and ease interpretation of model summaries (Zuur et al. 2010).

All variables were checked for correlation prior to inclusion in the model and found to be below $r = 0.5$ (Appendix C, Fig. C1). During model development, we explored including multiple sources of hourly temperature (i.e., air temperature, north-facing camera temperature, and south-facing camera temperature) and daily temperature values by fitting univariate linear models. Daily temperatures were derived from the daily 1-km grid covering Norway (from senorge.no) and included minimum, maximum, and average daily temperature. Average daily temperature was the best singular predictor among all temperature variables ($R^2 = 0.06$, $p < 0.01$), but it was strongly correlated with season ($r = 0.6$, $p < 0.01$), so we deferred to the season variable because it was a better predictor of sink depth in univariate tests. Predictor and response variables for the model are summarized in Table 4.2.

All models were fit using the *gam* function from R's 'mgcv' package with a Tweedie distribution (Foster and Bravington 2013). The *gam* function is a flexible model call that allows for models with both linear and non-linear fits. Importantly, the *gam* function allowed model fitting with a Tweedie distribution in the *dredge* function from R's 'MuMIn' package to select the model with the lowest AIC value (Barton 2009). A Tweedie distribution is a compound Poisson and Gamma distribution that accounts for zero-inflated, continuous, non-negative data (Foster and Bravington 2013; Appendix C, Fig. C2). We used only linear fits in the model, because we had too few observations for non-linear splines (Wood 2017). We inspected all models with a $\Delta AIC < 4$ and selected the model with the lowest AIC value (Burnham et al. 2011). We inspected model residuals for heteroskedasticity using the Breush Pagan Test implemented through the *bptest* function in the 'lmtest' package in R (Breusch and Pagan 1979; Zeileis et al. 2022). After identifying heteroskedasticity ($p < 0.05$), we used the *vcovHC* function in the 'sandwich' package in R to adjust coefficients for inflated p-values and standard error (Zeileis 2020). Using the best model (i.e., lowest AIC value), we calculated the lags for the maximum and minimum snow hardness values from sunrise and sunset. We identified the difference in timing from sunrise or sunset in hours by converting from radians to hours and multiplying by the day length, or the reverse of the *sunTime* function in R's 'overlap' package (Ridout and Linkie 2009).

Table 4.2. Variables used to model snow hardness. Temperature, snow density, and cloud cover were scaled by subtracting the mean and dividing by the standard deviation before inclusion in the model.

Variable	Source	Values
Snow hardness (penetrometer sinking depth)	Field data (tuna can)	0 – 8 cm (lower values = firmer snow)
Sun time	Derived from time	0 – 2*pi
Canopy Cover	Field data	Open or closed
Season	Field data	Winter (February or March) or Spring (April or May)
Density	Norway's snow model (senorge.no), daily	0 – 700 kg/m^3

4.2.4 *Analyzing Wildlife Diel Activity Cycles*

Roe deer and hare detections were extracted from the Scandcam camera trap database to determine the times of day each species was active at each camera station. Images were classified to species by a software using artificial intelligence, and the classifications were later controlled by human observers (Iannino et al. 2024). Sequences of images of the same animal(s) within a 5-minute period were included as a single detection. Date and time of day were extracted from the image, or the first image if taken from a series. Canopy cover at each camera station was derived using the 10-m resolution Sentinel-2A/2B derived tree cover product for 2015 (European Environment Agency 2020), with an additional adjustment for any subsequent yearly tree loss using the 30-m resolution NASA Global Forest Change product subsampled to 10-m resolution (Sexton et al. 2013; NASA 2020). Daily maximum and minimum air

temperature, snow depth, and snow density were extracted from Norway's 1-km daily climate grid for all detections (Saloranta et al. 2012).

For each species, we tested for differences in the timing of activity between days that we expected a snow hardness diel cycle and days without a snow hardness diel cycle across the study period. To do so, we separated roe deer and hare detections into freeze-thaw days and non-freeze-thaw days. We defined a freeze-thaw day as a day when the minimum air temperature was below 0˚C and the maximum air temperature was above 0˚C. Non-freeze-thaw days were all other days. We constrained the comparison to all detections with snow depth greater than or equal to 10 cm, as this is the height after which roe deer begin to experience locomotive difficulties on snow, and consistent with the depth threshold we used for snow hardness measurements (Andersen 1998, Janiszewski et al. 2011). We also conducted separate analyses to compare roe deer and hare diel cycles for changes relative to season (winter and spring) and canopy (open and closed), to reflect the categorical variables used in our snow hardness model. We generated activity curves for each set of comparisons using the *overlap* R package (Ridout and Linkie 2009), which uses a kernel density estimator to smooth frequency of activity over the 24-hour diel cycle. Diel activity overlap can range from 0 (no overlap) to 1 (complete overlap), with lower values indicating a stronger shift in the timing of activity in relation to snow hardness. We used the non-parametric Mardia-Watson-Wheeler test to determine if the activity shift was significant using the *circular* package in R (Agostinelli & Lund 2023; Landler et al. 2021).

While the overlap analysis provides a useful visual and a simple test for an activity shift, it does not include snow hardness values. To include hardness as a continuous predictor of animal activity, we used generalized additive models (GAMs) to model hourly activity for both

species in relation to snow hardness. We then summed detections for each hour (1-24) across all days for each camera, with a value equal to 0 for any hour the animal was not detected. We converted clock time to sun time (Vazquez et al. 2019) and aggregated our detection data from the hourly scale to the week-hour scale, summing all the detections in each hour over the course of a week, using the *week* function in R's 'lubridate' package (Grolemund and Wickham 2011). We subsequently calculated the snow hardness predictions from our snow hardness model at each hour of the day. The result was a unique value for each camera, week, and hour. We limited the data to days that met freeze-thaw critieria. We also constrained observations to when snow depth < 150 cm, as we had few points of observations for hares and roe deer above 150 cm (0.5% for roe deer and 0.3% for hares), and GAMs that included the full snow depth range showed sporadic relationships at high snow depth values likely due to sparsity of data.

We constructed a GAM to model how the predicted snow hardness impacted hourly animal activity. Snow depth is recognized as a strong driver in animal behavior (Mysterud et al. 1997, Stenseth et al. 2004, Jansson and Pehrson 2007, Melin et al. 2023), so we built a model to quantify the impacts of snow hardness on animal detections, incorporating snow depth as both a predictor and as an interaction with snow hardness and sun time (Eqn. 4.2):

$$\text{animal detections} \sim s(\text{sun time}) + s(\text{snow hardness}) + s(\text{snow depth}) + s(\text{sun time} \times \text{snow hardness depth} \times \text{snow depth}) \quad \text{(Eqn. 4.2)}$$

We fit sun time with a circular cubic function, snow depth and snow hardness with simple isotropic smoothing splines, and the interaction term was constructed using a tensor product. We plotted partial effects of each predictor and interaction term, using plots to assess roe deer and

hare activity changes in response to snow conditions. We assessed for model fit using the proportion of deviance explained (Wood 2023), and we evaluated the statistical significance of each factor using p-values (alpha < 0.05). Model were constructed using the R package *mgcv* (Wood 2023).

4.3 Results

4.3.1 *Data Collection Summary*

We conducted seven 24-hour snow hardness surveys and four 8-hour surveys for a total of 251 hourly hardness observations. We confirmed that all measurement days contained freeze-thaw conditions (i.e., temperatures crossed 0°C) using the on-the-ground hourly air temperature measurements. Six measurements from site 3 followed a large morning snowstorm and were removed from analysis. Sinking depth of the tuna can penetrometer ranged from 0-8 cm ($\bar{x}$ = 2.89 ± 2.55 cm). Average sink depth in open canopy indicated a softer and more variable snow surface than in closed canopy forest (4.25 ± 2.34 cm versus 1.40 ± 1.35 cm). Visual inspection of sinking depth from the penetrometer revealed hourly changes in both open and closed canopy (Appendix C, Fig. C3).

4.3.2 *Snow Hardness Modeling*

All supported models (ΔAIC < 4; n = 13 models) included fixed effects for sun time, canopy cover, season, and snow density as predictors of surface snow hardness. All supported models also included significant interactions between canopy and snow density, season and sine of sun time, and season and cosine of sun time. The model with the lowest AIC value included a fourth interaction for canopy and the cosine of sun time. All variables in the model with the

lowest AIC were significant, and the model performed with an overall R^2 of 0.63 (Appendix C, Table C1 and C2).

Surface hardness (i.e., sinking depth from tuna can penetrometer) predictions ranged between 0 and 7.3 cm at all cameras and all hours, with an average of 2.52 ± 1.98 cm. Overall, softer snow occurred in the afternoon (higher sinking depth from penetrometer values), and harder snow occurred overnight and into the morning (lower sinking depth from penetrometer values). In the closed canopy, snow hardness diel cycles remained relatively consistent between winter and spring, with oscillations between ~1 cm and ~ 2 cm in winter and 0 cm and ~2 cm in spring. However, in the open canopy, hardness diel cycles differed seasonally in both timing and magnitude, oscillating between ~3 cm to ~6 cm in winter and 0 cm to ~5 cm in spring (Figure 4.2A-B). The hardest snow (i.e., nocturnal snow crust) in the diurnal cycle was just before sunrise in both winter and spring. After adjusting to account for Norway's variable day lengths, this was a 14-hour lag in winter and 8-hour lag in spring. The shorter lag time in the spring reflected the faster energy exchange between the snow surface and air as a result of the larger difference in

The relationship between the daily Norway snow model's snow density and measured snow hardness changed depending on canopy cover and time of day. During the winter, the relationship was relatively consistent at all times of day, but in the spring in the open canopy, the relationship between snow hardness and density differed depending on the time of day, with density values at sunset corresponding to up to 6-fold softer snow than values between midnight and noon (Appendix C, Figure C4A-D).

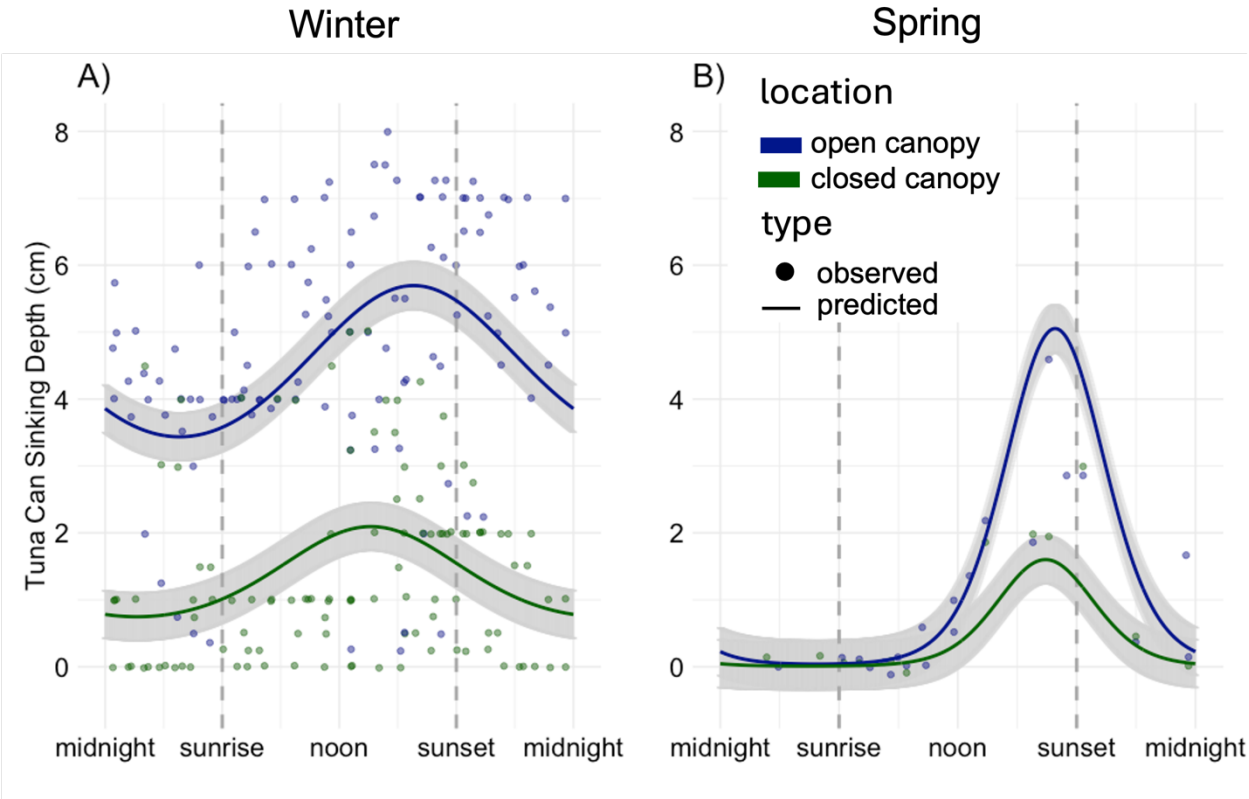


Figure 4.2 Plot of snow hardness observations from the tuna can penetrometer separated for closed and open canopy for A) winter (9 overnights; n = 192 observations) and B) spring (2 overnights; n = 55 observations). The x-axis represents hour of day when converted to radians relative to sunrise and sunset. Dashed lines indicate sunrise and set. Dark green and dark blue lines represent predicted values with corresponding 95% confidence intervals from the top model predicting snow hardness when snow density is held constant at 300 g/cm^3 (μ= 296.1 ± 34.4 g/cm^3). In both panels, lower sinking depth values correspond to harder snow, and higher values correspond to softer snow. We applied a jitter to show data more easily.

4.3.4 *Overlap Analysis*

Roe deer showed increased morning activity and less activity after sunset on freeze-thaw days compared to non-freeze thaw days (Figure 4.3A; Δ=0.89, 95% CI = 0.86 – 0.93; $\mathcal{X}^2$ = 8.51,

df = 2, $p < 0.05$). In contrast, hares showed no significant differences in activity patterns in relation to snow hardness diel cycles, and instead showed strong nocturnal behavior (Figure 4.3B; Δ=0.91, 95% CI = 0.89 – 0.95; $\mathcal{X}^2 = 4.78$, df = 2, $p = 0.09$). Roe deer diel cycles showed no differences depending on canopy cover, but there were significant differences depending on timing in the melt season (Figures 4.3C; Δ=0.78, 95% CI = 0.70 – 0.83; $\mathcal{X}^2 = 14.17$, df = 2, $p < 0.001$).Deer were most active mid-day in winter, and, their activity shifted toward mornings in the spring. Hares showed nocturnal behavior and no effects of canopy or season on activity level (Figures 4.3D; Δ=0.90, 95% CI = 0.86 – 0.94; $\mathcal{X}^2 = 3.66$, df = 2, $p = 0.161$).

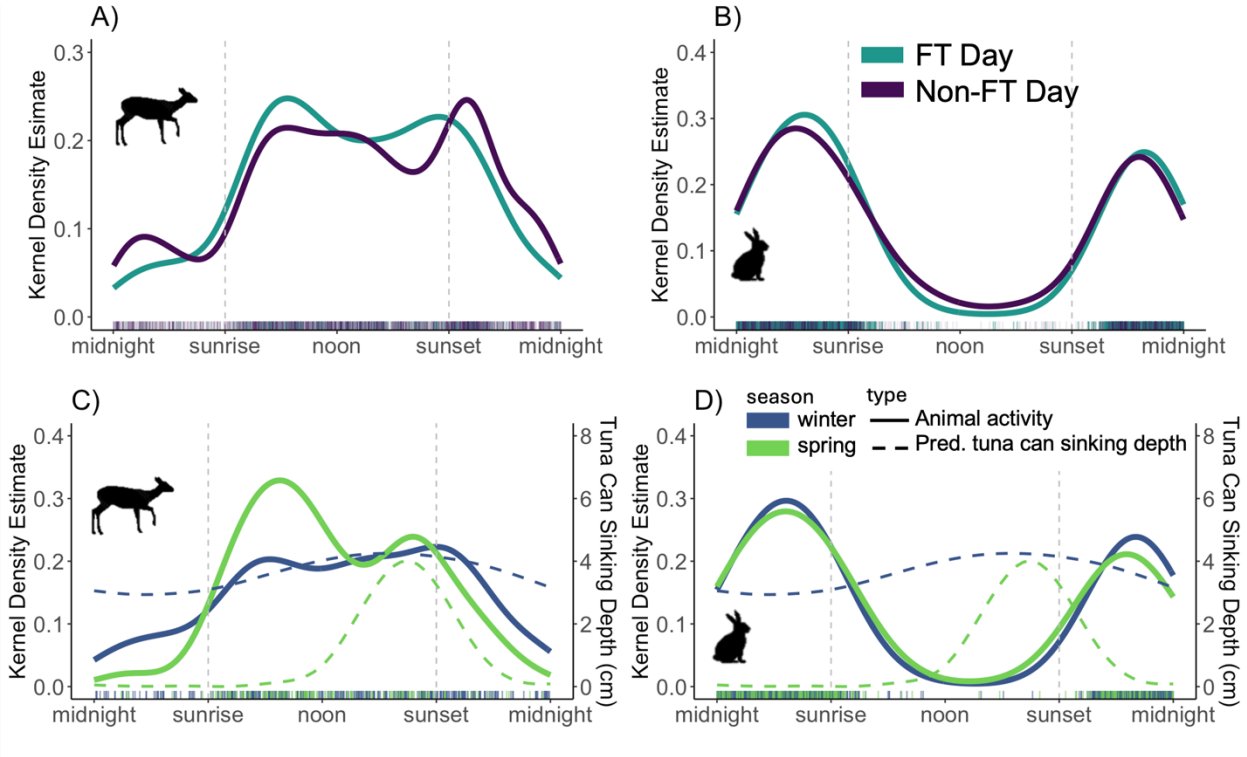


Figure 4.3 Comparison of diel cycles for A) roe deer and B) hare for freeze-thaw (FT) days and non-FT days across the study period. Roe deer showed significant differences between activity on FT days and non-FT days ($p < 0.001$). Hares showed no significant differences between these conditions ($p = 0.06$). C) and D) show comparisons for winter and spring activity for roe deer and hare, respectively. Roe deer showed significant differences depending on season ($p < 0.001$) but hares did not. For reference, predicted snow hardness values from our snow hardness model for both winter and spring is on the dotted line. Lower values indicate harder snow, and higher values indicate softer snow.

4.3.5 *Animal Activity Model*

Consistent with the overlap analysis, the GAM for roe deer activity showed overall crepuscular behavior for deer, with activity peaking just after sunrise and just before sunset (Figure 4.4A). Roe deer were more active when the snow surface was harder snow (i.e.,

predicted sinking depth values < 2.5 cm; $p < 0.05$*;* Fig. 4.4B). Deer were less active when snow depth exceeded 55 cm, a depth that is close to the threshold predicted to seriously impede movement based on their chest height ($p < 0.05$; Fig. 4.4C; Kelsall 1969). The interaction among sun time, snow hardness, and snow depth was significant, with roe deer showing increased morning activity when snow was deep (> 80) and hard (tuna can sinking depth < 3 cm) ($p < 0.05$; Appendix C, Figure C4 and Table C3). Overall, this model explained 12.2% of the deviance in the timing of roe deer activity.

The GAM analysis for hares confirmed their nocturnal behavior, with activity peaking after sunset and before midnight (Fig. 4.4D). Snow hardness had a negative effect on activity at harder snow values up to 2.5 cm, then a positive effect on activity at soft snow values > 2.5 cm (Fig. 4.4E). Snow depth had little effect up to 25 cm, and a negative effect at values > 25 cm (Fig. 4.4F). Like roe deer, the interaction among sun time, snow hardness, and snow depth was significant ($p < 0.05$; Appendix C, Figure C4 and Table C3). However, using the GAM model to predict hare activity showed that most hare activity remained between sunset and midnight at soft snow values ranging between 2 and 5 cm (Fig. 4.5B). This model explained 13.7% deviance (Appendix C, Table C3).

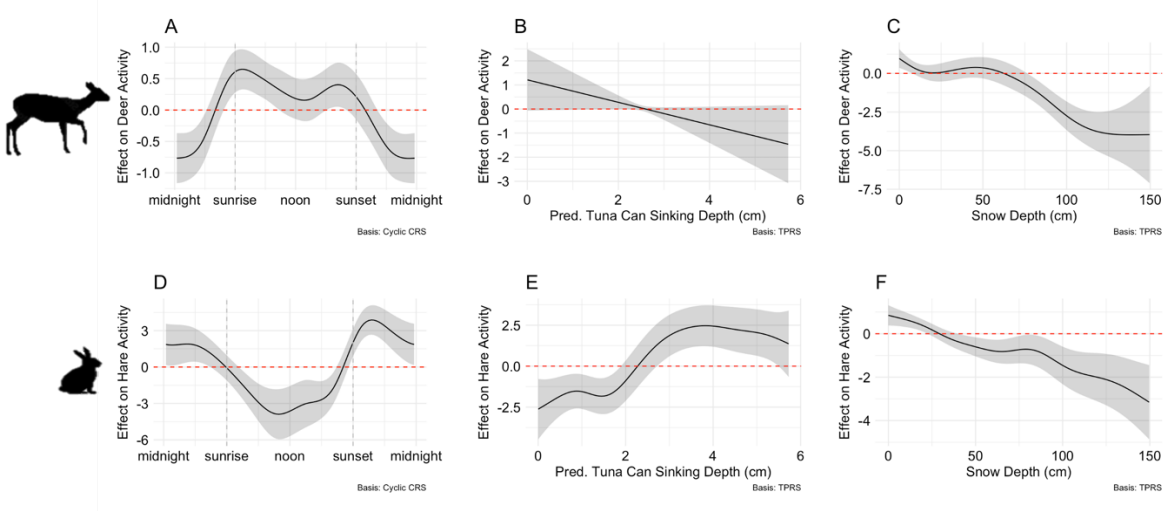


Figure 4.4. Regression splines for the fixed effects from the generalized additive model for roe deer (A-C) and hare (D-F), which regresses animal activity with sun time, estimated snow hardness, snow depth, and an interaction for sun time, snow hardness, and snow depth. Values above zero (the dashed red line) show the predictor variable had a positive effect on animal activity, and values below zero are associated with a negative effect. Methods of fitting the fixed effects in the regression, either cyclic cubic regression spline (CRS) or thin-plate regression spline (TPRS), are located in the bottom right-hand corner. The predicted tuna can sinking depth was derived from our snow hardness model, where higher values indicate softer snow.

4.4 Discussion

We quantified hourly changes in snow hardness in boreal Norway, demonstrating shifts in magnitude and timing depending on season, canopy cover, and snow density. Our model suggests the 'nocturnal snow crust' (maximum snow hardness) lags by approximately 14 hours after sunset in winter and 8 hours after sunset in spring, with the hardest snow typically occurring shortly before sunrise. The snow was softest in the afternoon, usually before sunset. When modeled with wildlife diel cycles, the timing of roe deer activity was affected by snow

hardness whereas hare activity was less sensitive, likely due to morphological differences. Snow conditions thus present a complex landscape for less snow-adapted animals, and understanding physical processes at the snow surface at a fine temporal scale should help to predict optimal movement for animals sensitive to snow conditions.

Our results emphasize the importance of canopy cover and seasonal changes in solar radiation as key factors driving changes in surface snow hardness. Snow hardness has largely been measured in the context of avalanche prediction studies (Kozak et al. 2003, Kinar and Pomeroy 2015), but our findings highlight the importance of canopy cover in the energy exchange at the snow surface level (Horstkotte and Roturier 2013, Lundquist et al. 2013). Snow density was correlated with, but not a sole predictor for, surface hardness, reinforcing the differences between snow density and hardness (Pielmeier and Schneebeli 2003, Geldsetzer and Jamieson 2001) and the importance of direct snow surface strength measurements for wildlife application (Penczykowski et al. 2017). For example, at a given density value in the spring, surface hardness could vary 6-fold between sunset and midnight in open canopy habitats, whereas hardness varied less than 0.5-fold at a given density in closed canopy sites (Figure 4.2 and Appendix C, Figure C4). Although prior studies have found that surface density is strongly correlated to surface hardness during witner in Alaska and Washington (Sullender et al. 2023, Sivy et al. 2018), density may be an especially poor proxy for hardness during the spring melt season in open habitats. Our model may be useful in these cases, improving snow hardness correlation with daily density by incorporating hour of day, canopy cover, and time of year information.

Our findings also underscore the sensitivity of roe deer to snow hardness, and the differences in sensitivity to snow hardness depending on species. For roe deer, we found

increased morning activity in the spring compared to winter, consistent with a recent study in Wyoming showing that ungulates are more active in the morning in the spring compared to fall (Jakopak et al. 2022). Additionally, while the Jakopak et al. study speculated about the role of harder snow in the morning in explaining this finding, our analysis provided evidence of this connection by identifying increased morning activity for roe deer on freeze-thaw days relative to non-freeze-thaw-days. Furthermore, results from our GAM for roe deer activity explicitly quantified the influence of snow hardness on activity. Harder snow had a positive effect on morning activity, and predictions from the GAM model revealed more activity in the morning on harder snow at higher snow depths. Ungulates prefer supportive snow spatially (Ossi et al. 2015, Mahoney et al. 2018), and this work demonstrates roe deer select for supportive snow temporally as well.

We found that roe deer increased activity on hard snow when it was deep, and that roe deer may prefer to travel on soft snow when it is shallow. This finding matched other studies suggesting an interaction between snow depth and snow strength (Mahoney et al. 2018). According to our model, the snow depth with which the roe deer selected for soft snow or hard snow switched at 50 cm. This result is surprising given that chest height is approximately 40 cm on roe deer depending on age and sex (Janiszewski et al. 2011), and the exponential increase in subsequent locomotive difficulties at snow depths above this height (Kelsall 1969). However, there may be some instances when softer snow is advantageous. Softer snow may offer easier access to forage, as harder snow restricts grazing access to small bushes (Ossi et al. 2015). Softer snow was also consistently in open canopy, and previous studies have shown that roe deer will select for areas with stronger heat radiation from the sun for thermoregulation (Ratikainen et al. 2007). Snow hardness is thus not the only factor roe deer must account for during winter activity,

and this study highlights the trade-offs in movement animals may make to balance for foraging opportunities and temperature regulation.

In contrast, hares maintained nocturnal behavior, showing no change in response to freeze-thaw days or changes in season. Results from our GAM model showed that hare activity increased at softer snow at higher snow depths. Softer snow is advantageous for predator avoidance in hares, who have a lower foot load than their predators, the lynx (Murray and Boutin 1991b). Hares are also highly nocturnal (Mech et al. 1966, Shiratsuru et al. 2023), and they likely leverage their low foot loading to tolerate softer snow.

Our sample size of empirical snow hardness measurements was relatively low, and we extrapolated to several hundred camera sites using information collected from only three areas. Our model had strong predictive ability (a reasonably high R^2) despite these limitations, representing an important first step in quantifying diel cycles of hardness. We recommend future studies conduct a larger study to include more overnight observations covering a larger regional area with elevational and coast-inland gradients. A larger study may be able to include other meteorological variables identified to affect snow hardness, such as temperature and cloud cover (Cagnati et al. 2004, Bormann et al. 2013). We only collected data on days when temperatures crossed 0, and a larger dataset may also be able to capture other conditions snow hardness diel cycles occur. We also had disproportionately more measurements in the winter (n = 192) than spring (n = 47), and future work should continue to focus on spring measurements when snow diel cycles overlap with wildlife migration and reproduction. We used modeled snow density and snow depth at 1-km resolution in our analysis to allow for extrapolating to the broader Scandcam sampling area, but recent work has emphasized the importance of spatial and temporal scale when monitoring various snow conditions (Reinking et al. 2022). Given the spatial variability of

density and depth over small areas, future work should explore increased resolutions from model outputs, as they may offer better snow hardness modeling at fine scales.

Improved understanding of snow hardness and animal behavior invites new questions for large-scale animal movement and landscape permeability. We demonstrate that activity timing is different for roe deer depending on snow hardness, suggesting new hypotheses for temporal barriers as well as potential avenues for movement. We recommend continued research on snow hardness in the context of landscape permeability, particularly for migratory animals like many ungulate species (Ramanzin et al. 2007, Panzacchi et al. 2016). Wildlife snow tracking may benefit from insight on snow diel cyles in the future too. Wildlife ecologists measure the sinking depth of animal tracks to assess locomotive abilities and predator-prey interactions (Pozzanghera et al. 2016, Sivy et al. 2018, Sullender et al. 2023), and improved information on the relation between sinking depth and time of day may provide more precise record on when the track was made as well as the snow landscape when the animal made the track.

4.6 Conclusion

Our empirical data reveal that snow surface hardness changes up to 6-fold on an hourly basis depending on canopy cover and timing during the melt season. Animals with increased locomotive difficulties on snow, such as roe deer, displayed stronger shifts in activity in response to harder snow than mammals with morphological advantages for snow, like mountain hares. While most research on snow hardness has focused on avalanches and Arctic tundra locations (e.g., Svalbard), this study emphasizes the influence of snow hardness on animal movement in the world's largest ecosystem, the boreal forest. Many arctic and boreal regions are experiencing increased rain-on-snow events (Cohen et al. 2015, Mooney and Li 2021) and winter days with

large temperature oscillations (Wilson et al. 2013, Berteaux et al. 2017), both of which should cause more changes in snow hardness. Our study highlights snow hardness as a landscape feature that shapes animal movement decisions temporally as well as spatially.

4.7 References

Agostinelli, C. and Lund, U. (2023). R package 'circular': Circular Statistics (version 0.5-0). URL https://CRAN.R-project.org/package=circular

Andersen, R. (Ed.). (1998). *The European roe deer: the biology of success*. Oslo: Scandinavian University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Berteaux, D., Gauthier, G., Domine, F., Ims, R. A., Lamoureux, S. F., Lévesque, E., & Yoccoz, N. (2017). Effects of changing permafrost and snow conditions on tundra wildlife: critical places and times. *Arctic Science*, *3*(2), 65–90. https://doi.org/10.1139/as-2016-0023

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., et al. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, *14*(1), 010401. https://doi.org/10.1088/1748-9326/aaeec1

Bormann, K. J., Westra, S., Evans, J. P., & McCabe, M. F. (2013). Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, *484*, 63–73. https://doi.org/10.1016/j.jhydrol.2013.01.032

Brogioni, M., Macelloni, G., Palchetti, E., Paloscia, S., Pampaloni, P., Pettinato, S., et al. (2009).

Monitoring Snow Characteristics With Ground-Based Multifrequency Microwave Radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(11), 3643–3655. https://doi.org/10.1109/TGRS.2009.2030791

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. https://doi.org/10.1007/s00265-010-1029-6

Cagnati, A., Crepaz, A., Macelloni, G., Pampaloni, P., Ranzi, R., Tedesco, M., et al. (2004). Study of the snow melt–freeze cycle using multi-sensor data and snow modeling. *Journal of Glaciology*, *50*(170), 419–426. https://doi.org/10.3189/172756504781830006

Carricondo-Sanchez, D., Odden, M., Linnell, J. D. C., & Odden, J. (2017). The range of the mange: Spatiotemporal patterns of sarcoptic mange in red foxes (Vulpes vulpes) as revealed by camera trapping. *PLOS ONE*, *12*(4), e0176200. https://doi.org/10.1371/journal.pone.0176200

Cohen, J., Ye, H., & Jones, J. (2015). Trends and variability in rain-on-snow events. *Geophysical Research Letters*, *42*(17), 7115–7122. https://doi.org/10.1002/2015GL065320

Collins, W. B., & Smith, T. S. (1991). Effects of Wind-Hardened Snow on Foraging by Reindeer (*Rangifer Tarandus*). *ARCTIC*, *44*(3), 217–222. https://doi.org/10.14430/arctic1541

David Mech, L., Heezen, K. L., & Siniff, D. B. (1966). Onset and cessation of activity in cottontail rabbits and snowshoe hares in relation to sunset and sunrise. *Animal Behaviour*, *14*(4), 410–413. https://doi.org/10.1016/S0003-3472(66)80038-6

Droghini, A., & Boutin, S. (2018). The calm during the storm: Snowfall events decrease the

movement rates of grey wolves (Canis lupus). *PLOS ONE*, *13*(10), e0205742. https://doi.org/10.1371/journal.pone.0205742

Ellis, C. R., Pomeroy, J. W., & Link, T. E. (2013). Modeling increases in snowmelt yield and desynchronization resulting from forest gap-thinning treatments in a northern mountain headwater basin: FOREST COVER IMPACTS ON MOUNTAIN SNOWMELT. *Water Resources Research*, *49*(2), 936–949. https://doi.org/10.1002/wrcr.20089

European Environment Agency. (2020). Tree Cover Density 2018 (raster 10 m), Europe, 3-yearly, Sep. 2020 (Version 02.00) [GeoTIFF]. European Environment Agency. https://doi.org/10.2909/486F77DA-D605-423E-93A9-680760AB6791

Foster, S. D., & Bravington, M. V. (2013). A Poisson–Gamma model for analysis of ecological non-negative continuous data. *Environmental and Ecological Statistics*, *20*(4), 533–552. https://doi.org/10.1007/s10651-012-0233-0

Gaynor, K. M., Hojnowski, C. E., Carter, N. H., & Brashares, J. S. (2018). The influence of human disturbance on wildlife nocturnality. *Science*, *360*(6394), 1232–1235. https://doi.org/10.1126/science.aar7121

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with **lubridate**. *Journal of Statistical Software*, *40*(3). https://doi.org/10.18637/jss.v040.i03

Hackländer, K., & Zachos, F. E. (Eds.). (2020). *Handbook of the Mammals of Europe*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65038-8

Hansen, B. B., Lorentzen, J. R., Welker, J. M., Varpe, Ø., Aanes, R., Beumer, L. T., & Pedersen, Å. Ø. (2019). Reindeer turning maritime: Ice-locked tundra triggers changes in dietary niche utilization. *Ecosphere*, *10*(4), e02672. https://doi.org/10.1002/ecs2.2672

Horstkotte, T., & Roturier, S. (2013). Does forest stand structure impact the dynamics of snow

on winter grazing grounds of reindeer (Rangifer t. tarandus)? *Forest Ecology and Management*, *291*, 162–171. https://doi.org/10.1016/j.foreco.2012.10.044

Janiszewski, P., Gugołek, A., & Hanzal, V. (2011). Zoometric characteristics of the european roe deer (Capreolus capreolus) hunter-harvested in north-eastern Poland. *Annales UMCS, Zootechnica*, *29*(2). https://doi.org/10.2478/v10083-011-0008-3

Jansson, G., & Pehrson, Å. (2007). The recent expansion of the brown hare (Lepus europaeus) in Sweden with possible implications to the mountain hare (L. timidus). *European Journal of Wildlife Research*, *53*(2), 125–130. https://doi.org/10.1007/s10344-007-0086-2

Kelsall, J. P. (1969). Structural Adaptations of Moose and Deer for Snow. *Journal of Mammalogy*, *50*(2), 302. https://doi.org/10.2307/1378347

Kinar, N. J., & Pomeroy, J. W. (2015). Measurement of the physical properties of the snowpack. *Reviews of Geophysics*, *53*(2), 481–544. https://doi.org/10.1002/2015RG000481

Kozak, M. C., Elder, K., Birkeland, K., & Chapman, P. (2003). Variability of snow layer hardness by aspect and prediction using meteorological factors. *Cold Regions Science and Technology*, *37*(3), 357–371. https://doi.org/10.1016/S0165-232X(03)00076-4

Landler, L., Ruxton, G. D., & Malkemper, E. P. (2021). Advice on comparing two independent samples of circular data in biology. *Scientific Reports*, *11*(1), 20337. https://doi.org/10.1038/s41598-021-99299-5

Lundquist, J. D., Dickerson-Lange, S. E., Lutz, J. A., & Cristea, N. C. (2013). Lower forest density enhances snow retention in regions with warmer winters: A global framework developed from plot-scale observations and modeling: Forests and Snow Retention. *Water Resources Research*, *49*(10), 6356–6370. https://doi.org/10.1002/wrcr.20504

Mahoney, P. J., Liston, G. E., LaPoint, S., Gurarie, E., Mangipane, B., Wells, A. G., et al.

(2018). Navigating snowscapes: scale-dependent responses of mountain sheep to snowpack properties. *Ecological Applications*, *28*(7), 1715–1729. https://doi.org/10.1002/eap.1773

Melin, M., Matala, J., Mehtätalo, L., Pusenius, J., & Packalen, T. (2023). The effect of snow depth on movement rates of GPS-collared moose. *European Journal of Wildlife Research*, *69*(2), 21. https://doi.org/10.1007/s10344-023-01650-w

Mooney, P. A., & Li, L. (2021). Near future changes to rain-on-snow events in Norway. *Environmental Research Letters*, *16*(6), 064039. https://doi.org/10.1088/1748-9326/abfdeb

Murray, D. L., & Boutin, S. (1991a). The influence of snow on lynx and coyote movements: does morphology affect behavior? *Oecologia*, *88*(4), 463–469. https://doi.org/10.1007/BF00317707

Murray, D. L., & Boutin, S. (1991b). The influence of snow on lynx and coyote movements: does morphology affect behavior? *Oecologia*, *88*(4), 463–469. https://doi.org/10.1007/BF00317707

Mysterud, A., Bjørnsen, B. H., & Østbye, E. (1997). Effects of snow depth on food and habitat selection by roe deer *Capreolus capreolus* along an altitudinal gradient in south-central Norway. *Wildlife Biology*, *3*(1), 27–33. https://doi.org/10.2981/wlb.1997.004

NASA. (2020). *NASA Global Forest Change, 2015 data*. Retrieved from https://lpdaac.usgs.gov/products/gfcc30tcv003/

Ossi, F., Gaillard, J.-M., Hebblewhite, M., & Cagnacci, F. (2015). Snow sinking depth and forest

canopy drive winter resource selection more than supplemental feeding in an alpine population of roe deer. *European Journal of Wildlife Research*, *61*(1), 111–124. https://doi.org/10.1007/s10344-014-0879-z

Ozeki, T., & Akitaya, E. (1998). Energy balance and formation of sun crust in snow. *Annals of Glaciology*, *26*, 35–38. https://doi.org/10.3189/1998AoG26-1-35-38

Panzacchi, M., Van Moorter, B., Strand, O., Saerens, M., Kivimäki, I., St. Clair, C. C., et al. (2016). Predicting the *continuum* between corridors and barriers to animal movements using Step Selection Functions and Randomized Shortest Paths. *Journal of Animal Ecology*, *85*(1), 32–42. https://doi.org/10.1111/1365-2656.12386

Pedersen, S. H., Bentzen, T. W., Reinking, A. K., Liston, G. E., Elder, K., Lenart, E. A., et al. (2021). Quantifying effects of snow depth on caribou winter range selection and movement in Arctic Alaska. *Movement Ecology*, *9*(1), 48. https://doi.org/10.1186/s40462-021-00276-4

Peers, M. J. L., Majchrzak, Y. N., Menzies, A. K., Studd, E. K., Bastille-Rousseau, G., Boonstra, R., et al. (2020). Climate change increases predation risk for a keystone species of the boreal forest. *Nature Climate Change*, *10*(12), 1149–1153. https://doi.org/10.1038/s41558-020-00908-4

Penczykowski, R. M., Connolly, B. M., & Barton, B. T. (2017). Winter is changing: Trophic interactions under altered snow regimes. *Food Webs*, *13*, 80–91. https://doi.org/10.1016/j.fooweb.2017.02.006

Poirier, M., Fauteux, D., Gauthier, G., Domine, F., & Lamarre, J. (2021). Snow hardness impacts intranivean locomotion of arctic small mammals. *Ecosphere*, *12*(11), e03835. https://doi.org/10.1002/ecs2.3835

Pozzanghera, C. B., Sivy, K. J., Lindberg, M. S., & Prugh, L. R. (2016). Variable effects of snow conditions across boreal mesocarnivore species. *Canadian Journal of Zoology*, *94*(10), 697–705. https://doi.org/10.1139/cjz-2016-0050

Ramanzin, M., Sturaro, E., & Zanon, D. (2007). Seasonal migration and home range of roe deer (*Capreolus capreolus*) in the Italian eastern Alps. *Canadian Journal of Zoology*, *85*(2), 280–289. https://doi.org/10.1139/Z06-210

Ratikainen, I. I., Panzacchi, M., Mysterud, A., Odden, J., Linnell, J., & Andersen, R. (2007). Use of winter habitat by roe deer at a northern latitude where Eurasian lynx are present. *Journal of Zoology*, *273*(2), 192–199. https://doi.org/10.1111/j.1469-7998.2007.00314.x

Reinking, A. K., Højlund Pedersen, S., Elder, K., Boelman, N. T., Glass, T. W., Oates, B. A., et al. (2022). Collaborative wildlife–snow science: Integrating wildlife and snow expertise to improve research and management. *Ecosphere*, *13*(6). https://doi.org/10.1002/ecs2.4094

Ridout, M. S., & Linkie, M. (2009). Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural, Biological, and Environmental Statistics*, *14*(3), 322–337. https://doi.org/10.1198/jabes.2009.08038

Saloranta, T. M. (2012). Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model. *The Cryosphere*, *6*(6), 1323–1337. https://doi.org/10.5194/tc-6-1323-2012

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., et al. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, *6*(5), 427–448. https://doi.org/10.1080/17538947.2013.786146

Shiratsuru, S., Studd, E. K., Boutin, S., Peers, M. J. L., Majchrzak, Y. N., Menzies, A. K., et al. (2023). When death comes: linking predator–prey activity patterns to timing of mortality to understand predation risk. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1999), 20230661. https://doi.org/10.1098/rspb.2023.0661

Sivy, K. J., Nolin, A. W., Cosgrove, C. L., & Prugh, L. R. (2018). Critical snow density threshold for Dall's sheep (*Ovis dalli dalli*). *Canadian Journal of Zoology*, *96*(10), 1170–1177. https://doi.org/10.1139/cjz-2017-0259

Stenseth, N. Chr., Shabbar, A., Chan, K.-S., Boutin, S., Rueness, E. K., Ehrich, D., et al. (2004). Snow conditions may create an invisible barrier for lynx. *Proceedings of the National Academy of Sciences*, *101*(29), 10632–10634. https://doi.org/10.1073/pnas.0308674101

Storebakken, B. (2022). Influence of forest on snow properties in Oslo, Norway.

Sullender, B. K., Cunningham, C. X., Lundquist, J. D., & Prugh, L. R. (2023). Defining the danger zone: critical snow properties for predator–prey interactions. *Oikos*, *2023*(10), e09925. https://doi.org/10.1111/oik.09925

Torres, R. T., Virgós, E., Panzacchi, M., Linnell, J. D. C., & Fonseca, C. (2012). Life at the edge: Roe deer occurrence at the opposite ends of their geographical distribution, Norway and Portugal. *Mammalian Biology*, *77*(2), 140–146. https://doi.org/10.1016/j.mambio.2011.11.002

United States, Army, Corps of Engineers. (1956). *Snow hydrology: Summary report of the snow investigations*. North Pacific Division, Corps of Engineers, U.S. Army.

Varhola, A., Coops, N. C., Weiler, M., & Moore, R. D. (2010). Forest canopy effects on snow accumulation and ablation: An integrative review of empirical results. *Journal of Hydrology*, *392*(3–4), 219–233. https://doi.org/10.1016/j.jhydrol.2010.08.009

Wilson, R. R., Bartsch, A., Joly, K., Reynolds, J. H., Orlando, A., & Loya, W. M. (2013). Frequency, timing, extent, and size of winter thaw-refreeze events in Alaska 2001–2008 detected by remotely sensed microwave backscatter data. *Polar Biology*, *36*(3), 419–426. https://doi.org/10.1007/s00300-012-1272-6

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems: *Data exploration*. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

Chapter 5. RAIN-ON-SNOW DETECTION FROM PASSIVE MICROWAVE REMOTE SENSING, WEATHER STATIONS, AND WILDLIFE CAMERAS

Publication history: This study was co-authored with Caleb Pan, Jinyang Du, Jessica Lundquist, John Odden, and Laura Prugh. At the time this dissertation was submitted, no version of this manuscript was in review, accepted or published elsewhere.

Abstract: Mid-winter rain events are increasingly important for ecology and hydrology monitoring in northern regions, but identifying rain and rain-on-snow (ROS) events is difficult in remote and complex terrain. Passive microwave information, from sources such as the Advanced Microwave Scanning Radiometer 2 (AMSR2), has been used with success in high arctic snowpacks, but its performance has not been evaluated in other snow-covered regions. We compared ROS events detected by AMSR2 to ROS events detected at weather stations in Norway, finding AMSR2 performed with lower accuracy in areas near water and with high tree cover. Using 253 cameras from a Norwegian wildlife camera network (Scandcam) managed by the Norwegian Institute for Nature Research, we trained a neural network model to classify rain in images, with an f1-score equal to 0.76. We then deployed the model on the remaining 103 cameras, combining the camera information with AMSR2 retrievals to improve accuracy with weather stations. We found an increase of 1% in accuracy for ROS detection when combining camera information with AMSR2 information, but that current methods for using AMSR2 to

detect ROS events in maritime, montane forest conditions are not robust. Remotely detecting rain on snow remains a wicked problem for both hydrologists and ecologists.

5.1 Introduction

Mid-winter rain events, specifically rain-on-snow (ROS) events, are becoming more common in high latitude regions as the climate warms (Cohen et al., 2015). These ROS events have drastic impacts on hydrology and ecology. Major rain events in mid and late winter can result in widespread flooding due to rapid snowpack melt (Musselman et al., 2018). Winter rain followed by refreezing can create an icy layer and restrict vegetation access for grazing animals, leading to catastrophic die-offs in some case studies on reindeer (*Rangifer tarandus platyrhynchus*), caribou (*Rangifer tarandus pearyi*), and muskox (*Ovibos moschatus*) (Hansen et al., 2019a, 2019b; Langlois et al., 2017; Putkonen et al., 2009). Subnivean animals also suffer shelter access when intense rains collapse soft snow layers used for tunnels beneath the snow surface (Mölle et al., 2022). The increasing frequency of ROS events, combined with the impacts on hydrology and ecology, underlines the urgency to improve ROS monitoring and detection.

Current methods to monitor ROS events present trade-offs in scalability and accuracy. Weather stations are the primary mode for monitoring winter weather, but stations are not always located in habitats of interest for wildlife studies or in rugged terrain for hydrology monitoring (Boelman et al., 2019). Additionally, weather stations often monitor precipitation without differentiating between rain and snow, which can switch as often as hourly within the same storm (Lundquist et al. 2008). Researchers use temperature or humidity thresholds to identify rain at weather stations (Lundquist et al., 2008; Pan et al., 2018; Wayand et al., 2017),

but there is uncertainty in mixed precipitation conditions and close to threshold boundaries. Specific precipitation detection devices, called “disdrometers,” differentiate between rain and snow (Wayand et al., 2016), but these devices are expensive and large, and therefore difficult to scale up.

Passive microwave satellites can detect rain-on-snow (ROS) events due to the dielectric effects of moisture on passive microwave signals (Grenfell & Putkonen, 2008). These sensors offer near-daily global retrievals and are insensitive to polar darkness and cloud cover (Kim et al. 2017). One passive microwave sensor is the Advanced Microwave Scanning Radiometer 2 (AMSR2), flown on board the JAXA GCOM-W satellite (Du et al. 2014). Detection of ROS events by the AMSR2 is possible by exploiting the differences in the polarization of the 18.5 GHz and 36.5 GHz temperature brightness (T*b*) frequencies for wet snow (Dolant et al., 2016; Langlois et al., 2017; Pan et al., 2018). These retrievals have been particularly useful wildlife studies in Arctic and boreal regions. For example, van de Kirk et al. (2020) found that freeze-thaw events derived from passive microwave remote sensing was the best predictor for adult Dall sheep (*Ovis dalli dalli*) survival in Alaska, and a separate study in Alaska linked a large ROS event derived from AMSR2 to caribou populations shifts (Prugh et al. 2024). In the Canadian Arctic Archipelago, Langlois et al. (2017) identified winters with more ROS events with a decrease in Peary caribou populations (*Rangifer tarandus pearyi*) using the AMSR-E sensor, the precursor to AMSR2. Passive microwave retrievals from sensors such as the AMSR2 offer the ability to capture the spatial extent and frequency of large-scale freeze-thaw and ROS events where disdrometers and weather stations would be unfeasible, but retrievals remain sensitive to on-the-ground conditions, including vegetation, snow wetness, and soil moisture (Gan et al., 2021). Especially in maritime and montane forest snowpacks, where wet

snow and dense vegetation is common, more work is needed to assess the utility of passive microwave sensing for ROS detections (Langlois et al., 2017).

Remote camera networks deployed for wildlife monitoring may be useful for validating and supplementing ROS detection, by offering abilities to scale up to networks in areas of interest (Burton et al. 2015) and providing ground-based precipitation detection where passive microwave observations may be less accurate. Meteorological variables, including overcast, sunshine, snow, hail, and undifferentiated precipitation have been manually extracted from camera images (Alison et al., 2024, Dertien et al, 2017). Although automated algorithms have been developed for detecting rain in traffic cameras, in-vehicle cameras, and movie scenes (Barnum et al., 2010; Bossu et al., 2011; Khan & Ahmed, 2019), success depends on camera parameters (Garg & Nayar, 2007). Algorithms have not been developed for processing wildlife camera images for rain, and camera trap images present a new domain for rain detection. Camera traps often take photos with infrared flash, resulting in low-light, grey scale images, making information extraction more difficult (Beery et al., 2019; Breen et al., 2023). Additionally, in forests and varied terrain, obstacles such as rocks and trees can prevent visual cues from reaching the camera, leading to incorrect weather detection.

In this study, we evaluated the utility of camera traps to supplement passive microwave retrievals from AMSR2 for detecting ROS events in boreal forests of Norway, a region that includes both maritime and montane forest snow conditions (Sturm & Liston, 2021). Norway represents one of the regions of the world experiencing the fastest changes in ROS frequency (Cohen et al., 2015; Pall et al., 2019), with increasing winter rain at lower elevations and forested regions (Pall et al., 2019). We first identified the accuracy of AMSR2 retrievals for ROS detection in Norway by validating with weather stations. Second, using images from a

large-scale wildlife camera network (Scandcam) managed by the Norwegian Institute for Nature Research, we trained a deep learning model to identify rain in camera images. Third, we combined AMSR2-detected ROS observations with rain detected at cameras to improve overall AMSR2 accuracy, identifying improvement in accuracy combining AMSR2 with on-the-ground observations.

5.2 Methods

5.2.1 *Study Area*

Mainland Norway spans approximately 12°of latitude (57°N – 70°N) and 25° of longitude (5°E to 32°E), encompassing a variety of forest, alpine, and high arctic climates. The Scandcam wildlife camera network, managed by the Norwegian Institute for Nature Research (NINA), was set up to monitor the recovering Eurasian lynx (*Lynx lynx*). Cameras were placed in lynx habitat, which is primarily boreal forest, spanning a latitude gradient of 57°N – 67°N and a maritime-continental climate gradient (Fig. 5.1). Boreal forests are mostly Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) and subject to clear-cut forestry practices, which can strongly affect tree canopy cover. Elevations in these regions range from 0 to 1500 m (Carricondo-Sanchez et al., 2017; Iannino et al., 2024). About 5% of mainland Norway is water, including many small lakes and rivers (Schartau et al. 2010).

5.2.2 *Weather Station Data*

We downloaded the locations of all weather stations in Norway (n = 1011) using the Norwegian Meteorological Institute's online weather station archive (https://frost.met.no). We then downloaded data from all stations with hourly temperature and precipitation

measurements located within 30 km of a ScandCam camera (n = 183 weather stations; Fig. 5.1). We opted for hourly rather than daily temperature and precipitation observations as this allowed for matching with camera timestamps, and weather stations with daily temperature and precipitation observations aggregated hourly observations. We selected 30 km to balance trade-offs between sample size and overlap with weather events. The weather stations were within 30 km of 253 cameras (see Fig. 5.1).

We pulled all weather station data for our study period, January 2018 to March 2020, matching the availability of our camera imagery data from the Scandcam network. We consolidated the hourly observations from weather stations to daily rain events using the temperature and precipitation criteria for a ROS event from Mooney et al. (2019) and Lundquist et al. (2008): average temperature $\geq 3°C$, precipitation ≥ 5 mm/day. Rain events were then filtered to those occurring on snow using the MODIS/Terra Cloud-Gap Filled Snow Cover Daily L3 Global product at 500-m resolution (Hall & Riggs, 2020) and a snow cover index value equal to 40 for the threshold for snow/no snow pixels (Breen et al. 2023). All buffer analyses and geoprocessing were conducted using python's *rasterio* and *geopandas* packages.

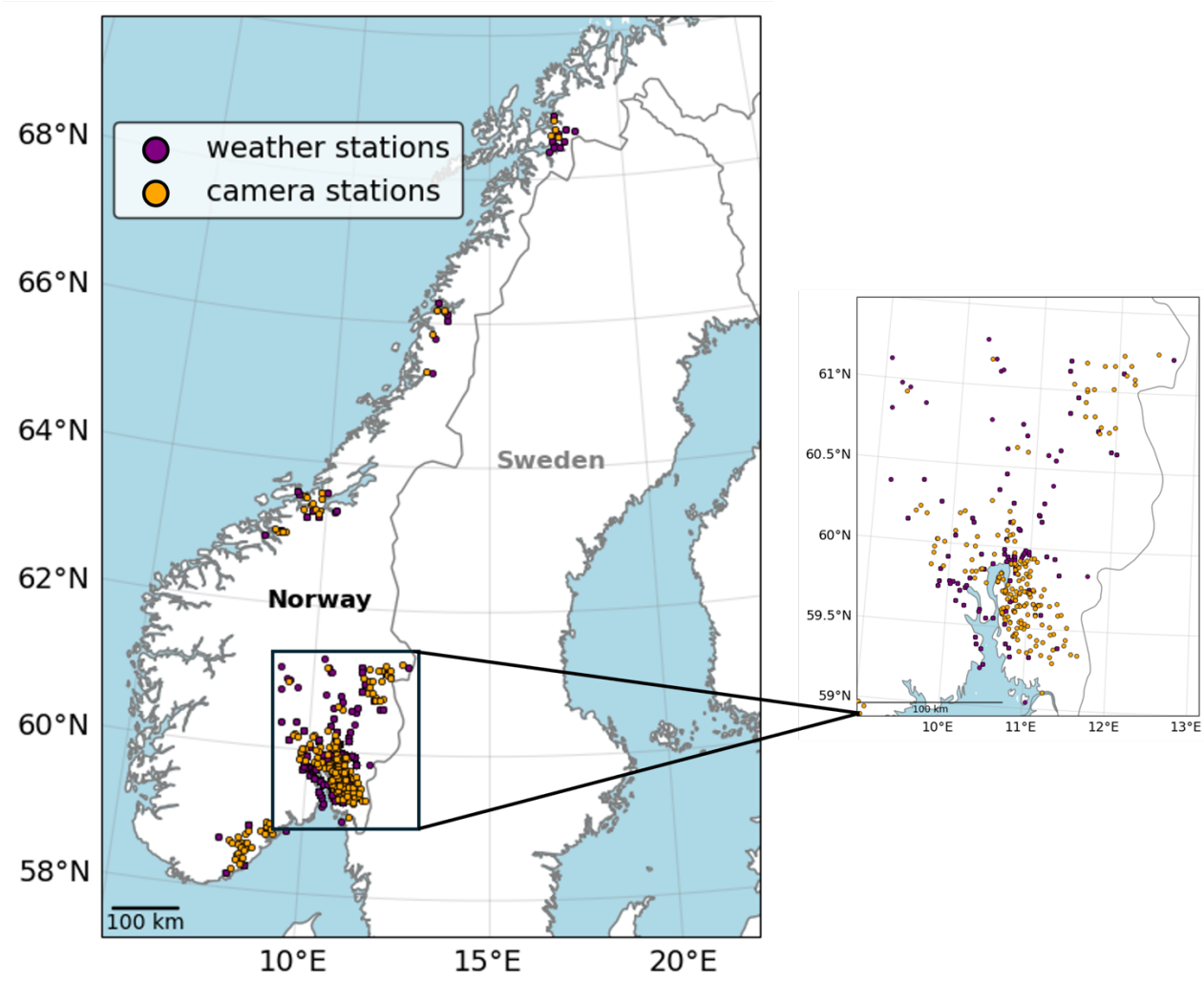


Figure 5.1. Weather stations used for AMSR2 validation (n = 183, purple dots). We selected weather stations within 30 km of camera traps (n = 253, orange dots) from the Scandcam wildlife camera network. We added an inset map to facilitate visualization in southern Norway, the region with the highest density of camera traps near weather stations.

5.2.3 *Passive Microwave Satellite Data*

We used the Calibrated, Enhanced-Resolution Brightness Temperature product obtained from the JAXA GCOM-W satellite's AMSR2 sensor (Du et al. 2014). Rain-on-snow events can be identified by examining the Gradient Ratio Polarization (GRP), or the ratio between the

vertical polarizations and the horizontal polarizations of the temperature brightness (T*b*) retrievals at frequencies 18.7 GHz and 36.5 GHz (Dolant et al, 2016, Du et al. 2016, Pan et al 2018, Prugh et al. 2024). The temperature brightness values change as snow wetness changes (Vuyovich et al. 2017), and the GRP exploits the differences in the vertical and horizontal polarizations during a ROS event, whereby horizontal polarizations return negative values and vertical polarizations return positive values (Langlois et al. 2017, Dolant et al. 2016). We followed processing steps described in Pan et al. (2018) and resampled the ~12.5 km AMSR2 T*b* observations to a 3 km resolution in the polar EASE-Grid (version 2) geographic projection using bilinear interpolation. Pan et al. (2018) used a 6 km grid, but more recent studies have used resampling methods at 3.125 km resolution (Prugh et al. 2024). After resampling, we identified the gradient ratio between 18.7 GHz and 36.5 GHz frequencies for both the vertical and horizontal polarizations, then calculated the GRP. We classified rain as GRP values < 1 at elevations ≤900 m and < -5 for elevations greater than 900 m (Pan et al. 2018) using the elevation layer from the Copernicus Digital Elevation Model at 30-m resolution (European Space Agency & Airbus, 2022). As with the weather stations, we masked for snow or no snow pixels using the MODIS/Terra Cloud-Gap Filled Snow Cover Daily L3 Global product at 500-m resolution (Hall & Riggs, 2020) and a snow cover index value equal to 40 for the threshold for snow/no snow pixels (Breen et al. 2023). The result was a time-series of ROS events for all days across our study period.

5.2.4 *Comparing Weather Stations and Passive Microwave Satellites for ROS Events*

We compared ROS events detected at weather stations to ROS events derived from the AMSR2 product, considering it a match if the AMSR2 record detected a ROS event the same

day as the weather station or the day after (Pan et al. 2018). We used accuracy, the true positive rate, and the true negative rate to evaluate overall AMSR2 performance. Accuracy provides the overall performance, while the true positive rate and true negative rate provide an indication for likelihood of true positives and true negatives when equal weight is assigned to both classes. These metrics have been used in previous studies for validating remote sensing maps (Comber et al., 2012). Accuracy was defined as the total number of correct predictions divided by the total number of predictions, ranging from 0% (weak performance) to 100% (perfect performance). The true positive rate, sometimes referred to as sensitivity, was defined as the number of true ROS detections while ignoring false positive detections. The true negative rate, sometimes referred to as the specificity, was defined as the sensor's ability to predict true negatives, reflecting the number of false positives. The perfect model would have 100% accuracy, and a true positive rate and true negative rate equal to 1 (Tabak et al., 2019). Accuracy was calculated using the *accuracy_score* function in python's 'sklearn' package. The true positive and true negative rates were calculated using the equations for sensitivity and specificity, respectively (eqns. 1 and 2).

$$True\ positive\ rate\ (sensitivity) = \frac{True\ Positives}{(True\ Positives + False\ Negatives)} \quad \text{(Eqn. 5.1)}$$

$$True\ negative\ rate\ (specificity) = \frac{True\ Negatives}{(True\ Negatives + False\ Positives)} \quad \text{(Eqn. 5.2)}$$

We constrained our analysis to Norwegian winter months, October to April (and March for year 2020) for each year.

5.2.5 *Threshold Adjustment*

ROS detection from AMSR2 is sensitive to the GRP threshold, with recommendation that thresholds might need to be adjusted based on environmental factors or snowpack type (Langlois et al. 2018). We repeated the accuracy metrics (accuracy, true positive rate, and true negative) after identifying the optimal GRP threshold using a receiver operating curve (ROC) analysis, whereby we counted the number of true positives and true negatives from the weather stations at every GRP threshold, and selected the threshold that optimized for both ROS days and non-ROS days (Robin et al., 2011). This approach represents the best threshold for this region if the threshold is region dependent. We did not adjust the threshold for higher elevations (GRP < -5) because only one of our weather stations had an elevation greater than 900 m. Other studies have added a criterion that the snow had to be dry the day before a ROS event (Du et al. 2016), which can avoid inflated false positive detections during intermittent melting events. For both the GRP threshold < 1 and the GRP threshold identified by the ROC analysis, we tested whether a "dry snow check" on the day before a ROS event improved performance. In addition to investigating overall accuracy, the true positive rate, and the true negatives rate after changes to the GRP, we also investigated the top three locations where we had the greatest number of ROS events detected at the weather station, inspecting a timeseries of GRP value throughout the winter season.

5.2.6 *Factors that Impact AMSR2 Accuracy when Compared to Weather Stations*

Passive microwave retrievals are sensitive to tree cover, elevation, and soil moisture content (Derksen et al., 2012; Derksen, 2008; Du et al., 2016). We evaluated the effect of percent tree cover, elevation, precipitation, air temperature, and proximity to water body on the

ability for the AMSR2 retrievals to detect a true positive event using a logistic regression. We used the best-performing threshold from 5.2.5 (highest overall accuracy), and classified all detections when AMSR2 retrievals indicated a ROS event (i.e., positive detections only) as true positives (1) or false positives (0). We then fit a logistic model to determine the effect of these factors on the true positive probability, ensuring correlations were below 0.4 (Dormann et al., 2013) and that variables were normalized prior to inclusion (eqn. 5.3):

$$p(\text{True Positive}) \sim \text{precipitation} + \text{air temperature} + \text{tree cover} + \text{elevation} + \text{proximity to water} \quad \text{(eqn. 5.3)}$$

Precipitation represented the daily sum of precipitation from the weather stations. Air temperature represented the daily maximum air temperature. Tree cover represented the percent tree cover at the weather station and was derived using the MODIS/Terra Vegetation Continuous Fields Yearly L3 Global at 250 m resolution (DiMiceli et al., 2015). Elevation represented the elevation of the site above sea level derived from the Copernicus Digital Surface Model at 30 m resolution (European Space Agency & Airbus, 2022). Proximity to water was defined as 1 (< 500 m from water), 0.5 (500 m < x < 4 km from water), or 0 (> 4km from water), matching previous approaches for flagging pixels near water (Pan et al. 2018). Water information was derived from the European Space Agency's GlobCover: Global Land Cover Map at 300 m resolution (ESA 2010 and UCLouvain). We fit the model using the *Logit* model class from the statsmodels package in python.

5.2.7 *Adding camera station information to passive microwave ROS record*

5.2.7.1 Machine Learning to Label Images

We used cameras images from the Scandcam camera network, managed by the Norwegian Institute for Nature Research (Carricondo-Sanchez et al., 2017; Iannino et al., 2024) to assess improvement in the AMSR2 retrievals with on-the-ground information. Cameras were optimized for lynx detection, but prior studies have indicated success for remote sensing validation (Breen et al. 2023). We obtained camera images from years 2018, 2019, and 2020 (2020 only had images from January to March due to data limitations as result of the 2020 covid pandemic).

To detect images of rain, we trained a machine learning model (hereafter, "ML model") using the Scandcam images labeled as "rain" or "not rain." We divided the Scandcam images into three distinct sets: training, validating, and testing. The training and validating set were used during model development, allowing the model to fit and adjust parameters over the course of several iterations through the data. The test set, however, was not used for model fitting, but instead was held separate to evaluate the model's performance on data that it did not see during model fitting. A three-way data split with a test set allows for consistent comparison across models if the training protocol is adjusted, and this approach provides insight into how the model may perform on novel datasets given that no information in the test set is provided during model development (Bothmann et al., 2023). For the training dataset, we first attempted to manually identify rain in Scandcam images across the network, but we found manual labeling was too time intensive and difficult to detect visually. We achieved some labeling using images that had a clear visual signature for rain or no rain (e.g., sunny) across 370 cameras, but there was a 70:1 class imbalance of detections and non-detections for rain images

(n_{rain} = 131 images, n_{none} = 9,382 images). We included these images in the training set, then added images from a random subset of 75 of the 253 cameras near weather stations to improve class balance in the training set. We labeled these images using the hourly precipitation weather station information. Rain was defined as any event where hourly temperature was > 3°C and precipitation was 0.5 mm/hr (Lundquist et al. 2008). In places when a camera was near to multiple weather stations within 30 km, we used the nearest station. For the validation set, we used another random 75 cameras labeled by the weather stations. The result was 445 individual cameras sites in training and 75 cameras in validation ($n_{training}$ = 12,350 images and $n_{validation}$ = 3,745 images; Table 5.1). This breakdown for training and validation roughly mimics a 80/20 split for both independent cameras sites and images, which is a common breakdown for traditional machine learning data splits (Gholamy et al., 2018).

Using the training and validation datasets, we trained a model using a ResNet-50 model architecture (He et al., 2015). During training, we applied a model weight of 2 to the "rain" class to correct for the 2:1 ratio of "no rain" to "rain" images across our dataset (Kellenberger et al., 2018). We also added a data augmentation step to vary the lighting, as changing lighting conditions has proven an effective strategy to improve models on unseen data in machine learning for ecological studies (James & Bradshaw, 2020). Lighting adjustments in the data augmentation step was done by changing the brightness parameter in the *ColorJitter* function of the 'torch' transforms class (Paszke et al., 2017). The model was trained using a batch size of 64, learning rate of 0.001, and 2000 epochs with the option for early stopping. Model training was done using the *torch* package in python on an Azure NC6s v3 virtual machine (VM) with a Tesla V100-PCIE-16GB GPU.

We used the remaining 103 of the 253 cameras near weather stations to operate as a "test set" or hold out set, which the model did not see during model fitting. We deployed the trained model on these images and compared the predicted ML-labels to the weather station-assigned labels. We evaluated the model on the ability to perform on the test set, using accuracy, precision, recall, and f1-score metrics. Accuracy was defined as the number of overall correct predictions out of all predictions. Precision was defined as the proportion of identified rain images out of all rain predictions. Recall was defined as the model's ability to identify rain events out of all rain occurrences. The f1-score is the harmonic mean between precision and recall, ranging from 0 to 1, providing a balanced insight to overall model performance (Alshahrani et al., 2021; Sidumo et al., 2022). An f1-score of around 0.80 is considered acceptable for real-world data, with f1-scores as low as 0.60 considered acceptable for image processing for ecological analysis (James & Bradshaw, 2020; Kellenberger et al., 2018). We calculated the metrics using the *accuracy_score*, *precision_score*, *recall_score*, and *f1_score* functions in python's 'sklearn' package. We deferred to the machine learning labels for all images in the test set to assess how machine learning can facilitate image labeling.

Table 5.1 Scandcam cameras used for ML model training, validation, and testing for labeling rain in wildlife camera trap imagery. “Far from weather stations” corresponds to > 30 km; “near weather stations” corresponds to ≤30 km from weather stations.

Dataset	**Use**	**# of Non-Rain Images**	**# of Rain images**
Images far from weather stations (*n* = 370 cameras)	Training	9,382	131
Images near weather stations (*n* = 75 cameras)	Training	1,448	1,389
Images near weather stations (*n* = 75 cameras)	Validation	2,273	1,472
Images near weather stations (*n* = 103 cameras)	Testing	3,589	4,838
	Total	16,692	7,830

5.2.7.2 Supplementing AMSR2 with camera station information

For each day of each weather station observation, we aggregated all ML-labeled images from cameras within 30 km. On days and locations where we had multiple images captured from cameras ≤ 30 km from the weather station, we averaged the labels assigned by the model and rounded to the nearest integer. We then updated the AMSR2 record based on the averaged label for rain or no rain. We continued to screen for snow/no-snow using the daily MODIS CGF snow layer after applying the update from the camera record. We evaluated the improvement in accuracy, true positive rate, and true negative rate, across all weather stations. We also inspected improvement in accuracy at the top three stations with the most supplementation from the camera traps.

5.3 Results

5.3.1 *Comparing weather stations to AMSR2*

We found that when using a GRP threshold of < 1 to indicate rain-on-snow (ROS), AMSR2 overestimated the frequency of ROS events, performing with a near-perfect true positive rate of 99%, but a low true negative rate (27%; Table 5.2). When aggregating all weather stations together, AMSR2 detected a ROS event on 80% of the days of the winter season, 68% of which were false positives (n = 46,133 false positives). The mean GRP for days that corresponded to ROS at the weather station was below 1, but the standard deviation suggested that it could oscillate far above and below 1 ($\mu = 0.02 \pm 3.35$). Upon inspection of the three stations with the most weather station-detected ROS events, AMSR2 detected ROS almost every day of the winter season across all three water years (Appendix D, Figure D1). The threshold derived by the ROC analysis (GRP < 0.02), calculated to optimize the number of true positives for both classes, increased the accuracy to 58%. It also reduced the likelihood of a false positive, increasing the true negative rate to 60%. However, it increased the likelihood of missing a ROS event, dropping the true positive rate to 30% (Table 5.2). Including a criterion that the day before had to meet conditions for dry snow increased accuracy for both the GRP < 1 and the GRP < 0.02 thresholds, with the GRP < 0.02 threshold plus a dry snow check performing with the best accuracy of 88%. It reduced the number of false positive detections from 40% (GRP < 0.02) to 7% (GRP < 0.02 plus a dry snow check), but some caution should be used, as the dry snow check also reduced the true positive rate to 10% (Table 5.2). Inspection of the three weather stations with the most weather-station detected events shows that even at stations with low to intermediate tree cover, the GRP threshold of < 0.02 plus a dry snow check detects some, but not all, of the ROS events (Fig. 5.2).

Table 5.2 Performance of AMSR2 for detecting rain on snow events in Norway when compared to weather stations (n = 183 weather stations) using the gradient ratio polarization (GRP) from thresholds set in Pan et al. (2018), as well as a custom GRP identified using a receiver operating curve (ROC) analysis (GRP < 0.02). We evaluated accuracy for both thresholds after a conducting a check for dry snow the day before. We report overall accuracy, as well as the true negative rate and true positive rate for all thresholds.

	GRP < 1 (Pan et al. 2018)		GRP < 0.02 (ROC analysis)	
	All days	Day before dry snow	All days	Day before dry snow
Accuracy (%)	31	86	58	88
True Positive Rate	0.99	0.08	0.30	0.10
True Negative Rate	0.27	0.92	0.60	0.94

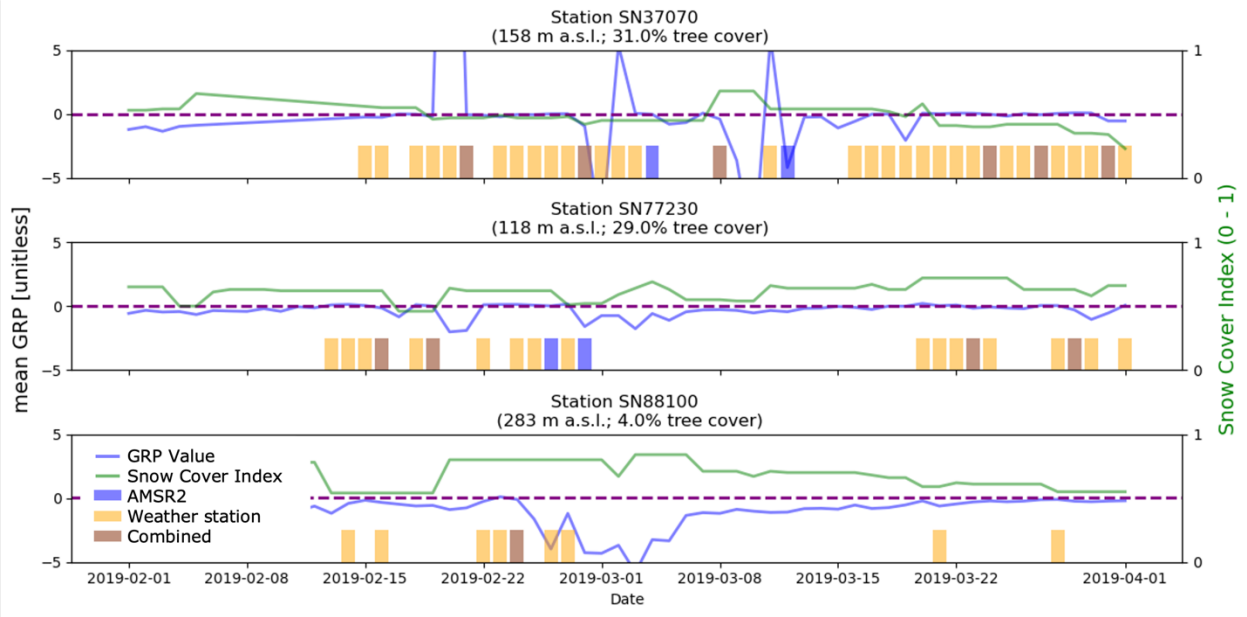


Figure 5.2. GRP values for three sample stations with the highest number of ROS events recorded from the weather station identified using a GRP threshold of < 0.02 (purple dashed line) plus a dry snow check. Stations from top to bottom include, SN37070 (58.9°N 8.6°E; n = 66 events), SN37070 (65.8°N 13.2°E; n = 56 events), and SN88100 (68.6°N 18.2°E; n = 50 events). We focus on 2 months from our study period, February 1, 2019, to April 1, 2019, to facilitate visualization. Blue bars indicate AMSR2 detected using a GRP < 0.02 with a check for dry snow, cream bars indicate a weather station detected a ROS event, and brown indicates both detected an ROS event. The blue line shows the daily GRP value, and the green line shows the snow cover index derived from the daily MODIS cloud-gap-filled product.

5.3.2 *Factors Impacting AMSR2 Accuracy*

We selected the GRP threshold of < 0.02 plus a dry snow check as the best threshold for detecting ROS events from AMSR2 because it reported with the highest accuracy. Using a logistic regression to predict the probability of a true positive detection using that threshold suggested that, unsurprisingly, higher precipitation and temperature were less likely to lead to a

false positive event (i.e., AMSR accurately detects severe ROS events). However, water and canopy cover also affected accuracy. Higher canopy cover and proximity to water both decreased the true positive likelihood (and subsequently increasing the false positive likelihood). Elevation had no significant effect on the true positive probability (Table 5.3).

Table 5.3. Coefficient estimates, standard error (SE), t-values, and *p*-values from a logistic model assessing factors that affect AMSR2 and weather station agreement (n = 26,097 observations) for three winter seasons: 1) January – April 2018, 2) October 2019 – April 2019, and 3) October 2019 – March 2020. We used a GRP threshold of < 0.02 plus a dry snow check to identify ROS from AMSR2. All variables were normalized and checked for correlation prior to inclusion in the model.

Parameter	Estimate	SE	t-value	*p*-value
Intercept	-3.01	0.078	-38.79	<0.001
Precipitation	0.74	0.125	5.87	<0.001
Temperature	0.67	0.063	10.54	<0.001
% Tree Cover	-0.14	0.073	-1.97	<0.05
Elevation	-0.11	0.078	-1.49	0.14
Proximity to Water	-0.24	0.069	-3.39	<0.01

5.3.3 *Supplementing AMSR2 with Camera Information*

5.3.3.1 Machine Learning Model

We compared the ML-predicted labels with the weather station labels for 103 cameras in the test set (n = 8,427 images) and found an overall accuracy of 70% for rain or no rain, with precision and recall for rain equal to 0.64 and 0.94, respectively. The f1-score, which is the weighted accuracy between the precision and recall, was equal to 0.76. Upon inspection, the

model made reasonable assessments for rain, with some instances demonstrating that it out-performed weather station labels likely due to the distance mismatch between the sources (Fig. 5.3). We deferred to the machine learning labels for all images.

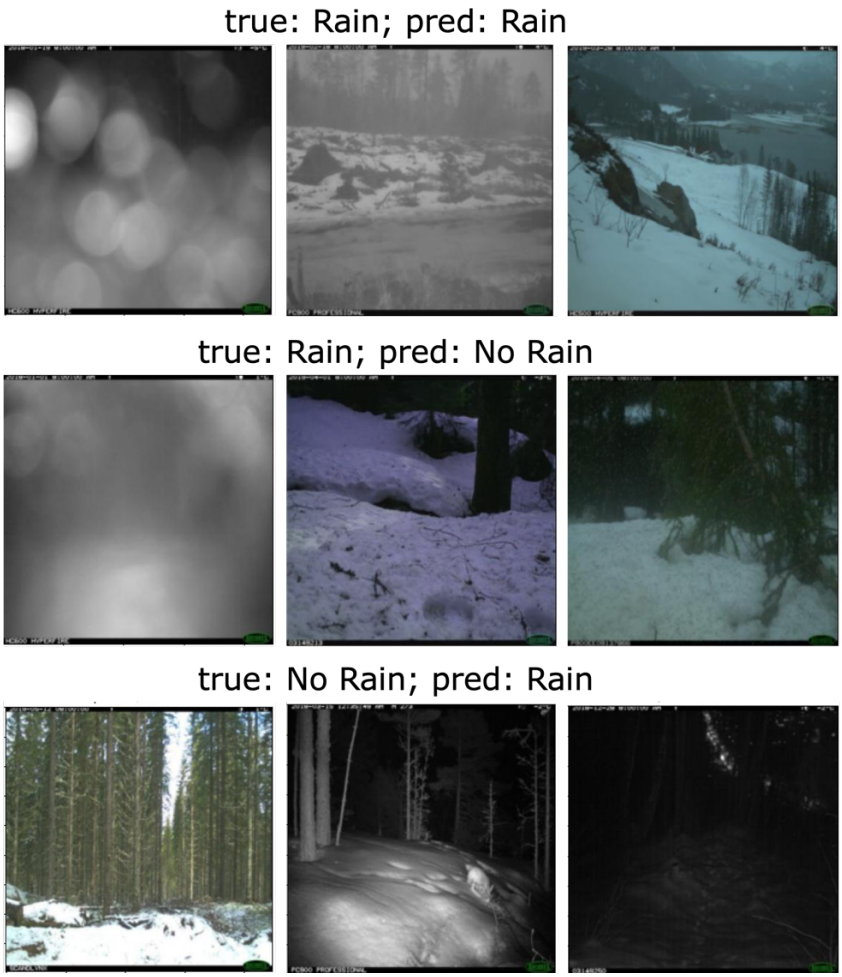


Figure 5.3 We used a deep learning neural network (f1-score = 0.76) to label images cameras from a wildlife camera network (Scandcam) for either “rain” or “not rain”. Above each row, we indicate the labeled value from nearby (< 30 km) weather stations (“true”) and the ML-model predicted value (“pred”). The top row represents images classified as rain from the model. The middle row represents instances when the model missed images that the nearby weather station recorded as rain. The bottom row represents instances when the model detected rain, but the nearby weather station did not.

5.3.3.2 Improvement in AMSR2 from camera supplementation

When the AMSR2 ROS retrievals were updated using the labels from the 103 cameras labeled by the machine learning model, we found 61 weather stations had at least one observation day updated by a camera. In total, we updated 1,909 observations (3% of total dataset), with 37% of the updates for a rain detection. When inspecting for changes in accuracy using the best threshold from 3.1 (GRP < 0.02 with a dry snow check), we found that it increased the accuracy from 88% to 89%, the true positive rate from 10% to 11% and the true negative rate stayed the same (93%). The supplementation increased the likelihood of a true positive detection by the updated record. The results were consistent for the other thresholds presented in 3.1 (Appendix D, Table D1)

When focusing on the top three stations with the most retrievals updated by camera labels (i.e., the most supplementation), we saw a 1-2% decrease in accuracy depending on the station. These stations had roughly one third of their observation days supplemented by camera images (Table 5.4). Across all three stations, it reduced the likelihood of a false positive (0.1 – 2% decrease in false positives across the three stations). In the case of stations SN3720 and SN17550, the added cameras resulted in 2 more ROS events detected (0.5% increase in true positives). Figure 5.4 shows that camera supplementation for AMSR2 increased true positive detection for ROS events such as on December 30, 2018, and March 2, 2019.

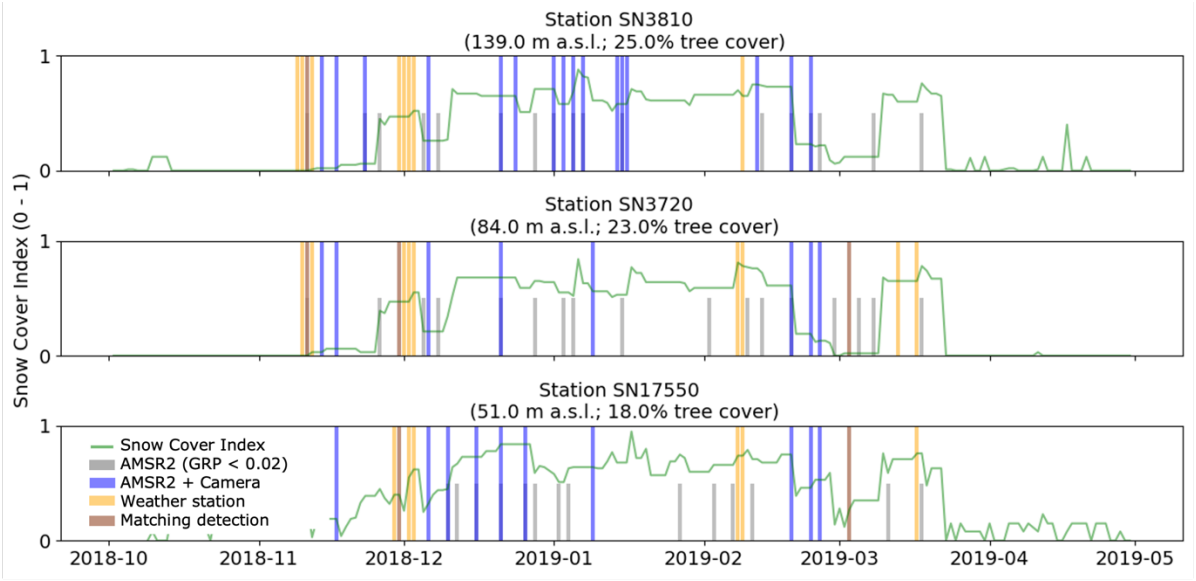


Figure 5.4. Timeseries plots for water year 2018-2019 for three weather stations with the most supplementation from camera trap images, station SN3810 (30% of observations supplemented), station SN4460 (29% of observations supplemented), and station SN3720 (35% of observations supplemented). AMSR2 accuracy improved with camera supplementation, by reducing some false positive detections (0.1-2%) and increasing some true positive detections (i.e., 2 events for stations SN4460 and SN3720). Grey bars indicate AMSR2-detected ROS using a GRP < 0.02 plus a dry snow check, blue indicate the record after updates with available camera values. Orange bars indicate ROS events detected at a weather station.

Table 5.4. The top three stations with the most supplementation of camera trap labels to the AMSR2 record for rain on snow events, stations SN3810 ((*n* = 164), SN3720 (*n* = 160), SN17550 (*n* = 143). We evaluated performance using accuracy, the true positive rate (TPR), and the true negative rate (TNR). "Before" indicates the performance without supplementation from camera traps, and "after" indicates supplementation from camera traps.

	SN3810		SN3720		SN17550	
	Before	After	Before	After	Before	After
Accuracy	0.88	0.89	0.88	0.90	0.91	0.92
TPR	0.15	0.15	0.26	0.26	0.05	0.15
TNR	0.92	0.93	0.93	0.95	0.95	0.96

5.4 Discussion

Detecting ROS events is important for calibrating hydrology models and understanding the impacts of changing winter conditions on wildlife. Weather stations provide some of the most accurate record of ROS events, but weather stations networks are often not at the appropriate scale for monitoring weather conditions in mountain watersheds and wildlife habitats. Passive microwave remote sensing, such as from the AMSR2 sensor, offers insight to ROS events by measuring on-the-ground moisture and temperature conditions at broad spatial scales (> 5 km) and up to daily observations (Askne & Westwater, 1986). AMSR2 has had success for ROS detections in the Arctic (Pan et al., 2018), but our results suggested that with current processing protocol, AMSR2 reports many false positives in maritime climates unless a check for dry snow conditions the day before is implemented. Machine learning offered capabilities to scale up ground-observations for data supplementation, and supplementing AMSR2 retrievals with information detected from cameras images provided a 1% improvement in accuracy by increasing true positive detection and reducing false positives. Snowpack in maritime, montane

forest regions present a major challenge for ROS monitoring, and this study advances understanding of the methods to integrate remote sensing and on-the-ground approaches for improved ROS monitoring.

Changes to the GRP threshold used to identify rain substantially improved ROS detection. Norway has complex topography, and southern regions have wet winters and dense forests (Pall et al., 2019; Rizzi et al., 2018). The environmental conditions in Norway present new opportunities to evaluate region-specific protocols for AMSR2 processing, and we found that the threshold used for low elevations in the Arctic (GRP < 1) performed poorly in this maritime, montane forest system. A region-specific threshold, identified for this region as GRP < 0.02, increased accuracy from 26% to 55%, supporting findings from Langlois et al. (2017) that ROS detection accuracy is higher when the GRP threshold is set to lower values for regions with denser snow. The assumption is that the signal must be more saturated (more negative) to trigger a positive ROS detection. In the boreal, maritime regions of Norway, average snow is denser than in Alaska due to increased air moisture, intermittent wetting events, and wind compaction (Bruland et al., 2015). A dry snow check for the region-specific threshold increased accuracy even more, to 88%, although it resulted in missed rain events as well (i.e., the true positive rate lowered from 30% to 10%). These findings suggest that other checks may be important to identify rain-on-snow events in this region. For example, we found that tree cover and proximity to water decreased the chances for a true positive AMSR2 ROS detection. Hence, tree cover or vegetation-based threshold adjustments may improve the accuracy of AMSR2 retrievals, similar to elevation-based adjustments in Alaska (Pan et al. 2018). During analysis, we assessed accuracy improvement with a tree cover unique threshold, using the average GRP at low tree cover (< 50%; μ_{GRP} = -0.02) and the average GRP at high

tree cover (>50%; μ_{GRP} = -0.06) as the thresholds for areas of low tree cover and high tree cover, respectively. Accuracy was similar to the GRP < 0.02 (accuracy = 63%, true positive rate= 0.18, true negative rate = 0.67), but not better than the dry snow check, so we did not present the results in the formal analysis. The low performance across the various thresholds demonstrates the complexity of detecting ROS events in a maritime climate, and the need for future work to further assess regional adjustments based on snowpack properties and environmental conditions.

Supplementing the AMSR2 record with rain detected from camera traps provided some improvement. We were able to supplement almost 1,200 AMSR2 observations, increasing accuracy depending on the threshold process by 1%. Our work demonstrates that networks installed for wildlife application present an additional source of environmental information for remote sensing applications, especially where on-the-ground sources are limited. However, one drawback to utilizing camera networks is that there is bias in the sampling depending on animal activity and when the camera is triggered. We used images from both the network's daily timelapse image (1 image daily at each camera at 8AM or 12PM) as well as the images that were motion triggered. The result is that we found we had more camera images that corresponded to rain-free days (n = 1077 images) than days with rain (n = 100 images), likely due to the relative infrequency of rain in this climate. This imbalance limited our ability to assess how cameras can supplement AMSR2 spatially, because days when we had the most detections from cameras coincided with days with smaller extent of ROS (see example in Appendix D2). Multiple daily timelapse images would substantially broaden the weather events sampled.

By leveraging machine learning to label the camera images, we present a novel model to detect precipitation from camera trap imagery, establishing a foundation for future work to build upon. The accuracy of our held-out test set (f1-score = 0.76) suggests that the model works reasonably well on Scandcam wildlife cameras it did not see during training, and future work should continue to add to the training data and test it on wildlife camera images from other networks. We found manually labeling for precipitation difficult due to the transparency of rain, and we recommend that future studies set up cameras near weather stations and disdrometers to build a robust dataset of validated images for model training. Additionally, other tools, such as passive acoustic recording devices typically used for recording bird calls and ambient sounds in an ecosystem (Hill et al., 2018), can identify rain (Bedoya et al., 2017), suggesting a promising source of supplementary data for identifying rain if deployed in the vicinity of the camera trap.

5.5 Conclusion

We present the first attempt at connecting wildlife cameras with AMSR2 to supplement ROS detections, using machine learning to extract the environmental information from the camera image. The current processing protocol for AMSR2 was not sufficient for ROS detection in a montane forest with maritime snow conditions, but we found that regional adjustments and checks for dry snow increased accuracy to an acceptable level (88%), although there were many missed ROS events. Further work on ways to account for the maritime and boreal impacts on AMSR2 temperature brightness signals will continue to improve ROS monitoring and highlights the urgent need to improve ROS monitoring across the multiple snow classifications identified globally (Sturm et al. 2021). Rain and snow events are expected

to increase in high latitude regions across the world, and improving remote sensing techniques to detect and monitor events remains a crucial problem.

5.6 References

Alison, J., Payne, S., Alexander, J. M., Bjorkman, A. D., Clark, V. R., Gwate, O., et al. (2024). Deep learning to extract the meteorological by-catch of wildlife cameras. *Global Change Biology*, *30*(1), e17078. https://doi.org/10.1111/gcb.17078

Alshahrani, H. M., Al-Wesabi, F. N., Al Duhayyim, M., Nemri, N., Kadry, S., & Alqaralleh, B. A. Y. (2021). An automated deep learning based satellite imagery analysis for ecology management. *Ecological Informatics*, *66*, 101452. https://doi.org/10.1016/j.ecoinf.2021.101452

Askne, J., & Westwater, E. (1986). A Review of Ground-Based Remote Sensing of Temperature and Moisture by Passive Microwave Radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-24*(3), 340–352. https://doi.org/10.1109/TGRS.1986.289591

Barnum, P. C., Narasimhan, S., & Kanade, T. (2010). Analysis of Rain and Snow in Frequency Space. *International Journal of Computer Vision*, *86*(2–3), 256–274. https://doi.org/10.1007/s11263-008-0200-2

Bedoya, C., Isaza, C., Daza, J. M., & López, J. D. (2017). Automatic identification of rainfall in acoustic recordings. *Ecological Indicators*, *75*, 95–100. https://doi.org/10.1016/j.ecolind.2016.12.018

Beery, S., Wu, G., Rathod, V., Votel, R., & Huang, J. (2019). Long Term Temporal Context for Per-Camera Object Detection. *arXiv:1912.03538 [Cs, Eess, q-Bio]*. Retrieved from http://arxiv.org/abs/1912.03538

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., et al. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, *14*(1), 010401. https://doi.org/10.1088/1748-9326/aaeec1

Bossu, J., Hautière, N., & Tarel, J.-P. (2011). Rain or Snow Detection in Image Sequences Through Use of a Histogram of Orientation of Streaks. *International Journal of Computer Vision*, *93*(3), 348–367. https://doi.org/10.1007/s11263-011-0421-7

Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., et al. (2023). Automated wildlife image classification: An active learning tool for ecological applications. *Ecological Informatics*, *77*, 102231. https://doi.org/10.1016/j.ecoinf.2023.102231

Breen, C., Vuyovich, C., Odden, J., Hall, D., & Prugh, L. (2023). Evaluating MODIS snow products using an extensive wildlife camera network. *Remote Sensing of Environment*, *295*, 113648. https://doi.org/10.1016/j.rse.2023.113648

Bruland, O., Færevåg, Å., Steinsland, I., Liston, G. E., & Sand, K. (2015). Weather SDM: estimating snow density with high precision using snow depth and local climate. *Hydrology Research*, *46*(4), 494–506. https://doi.org/10.2166/nh.2015.059

Carricondo-Sanchez, D., Odden, M., Linnell, J. D. C., & Odden, J. (2017). The range of the mange: Spatiotemporal patterns of sarcoptic mange in red foxes (Vulpes vulpes) as revealed by camera trapping. *PLOS ONE*, *12*(4), e0176200. https://doi.org/10.1371/journal.pone.0176200

Cohen, J., Ye, H., & Jones, J. (2015). Trends and variability in rain-on-snow events. *Geophysical Research Letters*, *42*(17), 7115–7122. https://doi.org/10.1002/2015GL065320

Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, *127*, 237–246. https://doi.org/10.1016/j.rse.2012.09.005

Derksen, C., Toose, P., Lemmetyinen, J., Pulliainen, J., Langlois, A., Rutter, N., & Fuller, M. C. (2012). Evaluation of passive microwave brightness temperature simulations and snow water equivalent retrievals through a winter season. *Remote Sensing of Environment*, *117*, 236–248. https://doi.org/10.1016/j.rse.2011.09.021

Derksen, Chris. (2008). The contribution of AMSR-E 18.7 and 10.7 GHz measurements to improved boreal forest snow water equivalent retrievals. *Remote Sensing of Environment*, *112*(5), 2701–2710. https://doi.org/10.1016/j.rse.2008.01.001

DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D.-H., Kelly, M., & Townshend, J. (2015). MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD44B.006

Dolant, C., Montpetit, B., Langlois, A., Brucker, L., Zolina, O., Johnson, C. A., et al. (2016). Assessment of the Barren Ground Caribou Die-off During Winter 2015–2016 Using Passive Microwave Observations. *Geophysical Research Letters*, 10.

Domine, F., Gauthier, G., Vionnet, V., Fauteux, D., Dumont, M., & Barrere, M. (2018). Snow physical properties may be a significant determinant of lemming population dynamics in the high Arctic. *Arctic Science*, *4*(4), 813–826. https://doi.org/10.1139/as-2018-0008

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-

0587.2012.07348.x

Droghini, A., & Boutin, S. (2018). The calm during the storm: Snowfall events decrease the movement rates of grey wolves (Canis lupus). *PLOS ONE*, *13*(10), e0205742. https://doi.org/10.1371/journal.pone.0205742

Du, J., Kimball, J. S., & Jones, L. A. (2016). Passive Microwave Remote Sensing of Soil Moisture Based on Dynamic Vegetation Scattering Properties for AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(1), 597–608. https://doi.org/10.1109/TGRS.2015.2462758

European Space Agency & Airbus. (2022). Copernicus DEM [Data set]. https://doi.org/10.5270/ESA-c5d3d65

European Space Agency (ESA) & UCLouvain. (2010). GlobCover: Global Land Cover Map at 300 m resolution. European Space Agency and Université catholique de Louvain. http://due.esrin.esa.int/page_globcover.php

Gan, Y., Zhang, Y., Kongoli, C., Grassotti, C., Liu, Y., Lee, Y.-K., & Seo, D.-J. (2021). Evaluation and blending of ATMS and AMSR2 snow water equivalent retrievals over the conterminous United States. *Remote Sensing of Environment*, *254*, 112280. https://doi.org/10.1016/j.rse.2020.112280

Garg, K., & Nayar, S. K. (2007). Vision and Rain. *International Journal of Computer Vision*, *75*(1), 3–27. https://doi.org/10.1007/s11263-006-0028-6

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation.

Grenfell, T. C., & Putkonen, J. (2008). A method for the detection of the severe rain-on-snow

event on Banks Island, October 2003, using passive microwave remote sensing: MICROWAVE DETECTION OF RAIN ON SNOW. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007WR005929

Hall, D., & Riggs, G. (2020). MODIS/Terra CGF Snow Cover Daily L3 Global 500m SIN Grid, Version 61 [Data set]. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD10A1F.061

Hansen, B. B., Lorentzen, J. R., Welker, J. M., Varpe, Ø., Aanes, R., Beumer, L. T., & Pedersen, Å. Ø. (2019a). Reindeer turning maritime: Ice-locked tundra triggers changes in dietary niche utilization. *Ecosphere*, *10*(4), e02672. https://doi.org/10.1002/ecs2.2672

Hansen, B. B., Lorentzen, J. R., Welker, J. M., Varpe, Ø., Aanes, R., Beumer, L. T., & Pedersen, Å. Ø. (2019b). Reindeer turning maritime: Ice-locked tundra triggers changes in dietary niche utilization. *Ecosphere*, *10*(4), e02672. https://doi.org/10.1002/ecs2.2672

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep Residual Learning for Image Recognition. arXiv. Retrieved from http://arxiv.org/abs/1512.03385

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, *9*(5), 1199–1211. https://doi.org/10.1111/2041-210X.12955

Iannino, E., Linnell, J. D. C., Devineau, O., Odden, J., Mattisson, J., & Horntvedt Thorsen, N. (2024). Assessing the potential of camera traps for estimating activity pattern compared to collar-mounted activity sensors: a case study on Eurasian lynx *Lynx lynx* in south-eastern Norway. *Wildlife Biology*, e01263. https://doi.org/10.1002/wlb3.01263

James, K., & Bradshaw, K. (2020). Detecting plant species in the field with deep learning and

drone technology. *Methods in Ecology and Evolution*, *11*(11), 1509–1519. https://doi.org/10.1111/2041-210X.13473

Kellenberger, B., Marcos, D., & Tuia, D. (2018). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, *216*, 139–153. https://doi.org/10.1016/j.rse.2018.06.028

Khan, M. N., & Ahmed, M. M. (2019). Snow Detection using In-Vehicle Video Camera with Texture-Based Image Features Utilizing K-Nearest Neighbor, Support Vector Machine, and Random Forest. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(8), 221–232. https://doi.org/10.1177/0361198119842105

Langlois, A., Johnson, C.-A., Montpetit, B., Royer, A., Blukacz-Richards, E. A., Neave, E., et al. (2017). Detection of rain-on-snow (ROS) events and ice layer formation using passive microwave radiometry: A context for Peary caribou habitat in the Canadian Arctic. *Remote Sensing of Environment*, *189*, 84–95. https://doi.org/10.1016/j.rse.2016.11.006

Lundquist, J. D., Neiman, P. J., Martner, B., White, A. B., Gottas, D. J., & Ralph, F. M. (2008). Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level. *Journal of Hydrometeorology*, *9*(2), 194–211. https://doi.org/10.1175/2007JHM853.1

Mölle, J. P., Kleiven, E. F., Ims, R. A., & Soininen, E. M. (2022). Using subnivean camera traps to study Arctic small mammal community dynamics during winter. *Arctic Science*, *8*(1), 183–199. https://doi.org/10.1139/as-2021-0006

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., et al. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*, *8*(9), 808–812. https://doi.org/10.1038/s41558-018-0236-4

Pall, P., Tallaksen, L. M., & Stordal, F. (2019). A Climatology of Rain-on-Snow Events for Norway. *JOURNAL OF CLIMATE*, *32*, 22.

Pan, C. G., Kirchner, P. B., Kimball, J. S., Kim, Y., & Du, J. (2018). Rain-on-snow events in Alaska, their frequency and distribution from satellite observations. *Environmental Research Letters*, *13*(7), 075004. https://doi.org/10.1088/1748-9326/aac9d3

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2019). Automatic differentiation in PyTorch.

Prugh, L.R., Lundquuist, J.D., Sullender, B.K., Cunningham, C.X., Dechow, J., Borg, B., Sousanes, P.J., Stehn, S., Durand, M.T. (2024) Landscape heterogeneity buffers the impact of an extreme weather event on wildlife. *Commun Biol 7, 1515 (2024).* *https://doi.org/10.1038/s42003-024-07195-1*

Putkonen, J., Grenfell, T. C., Rennert, K., Bitz, C., Jacobson, P., & Russell, D. (2009). Rain on Snow: Little Understood Killer in the North. *Eos, Transactions American Geophysical Union*, *90*(26), 221–222. https://doi.org/10.1029/2009EO260002

Rizzi, J., Nilsen, I. B., Stagge, J. H., Gisnås, K., & Tallaksen, L. M. (2018). Five decades of warming: impacts on snow cover in Norway. *Hydrology Research*, *49*(3), 670–688. https://doi.org/10.2166/nh.2017.051

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

Sidumo, B., Sonono, E., & Takaidza, I. (2022). An approach to multi-class imbalanced problem in ecology using machine learning. *Ecological Informatics*, *71*, 101822. https://doi.org/10.1016/j.ecoinf.2022.101822

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

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., et al. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Van de Kerk, M., Arthur, S., Bertram, M., Borg, B., Herriges, J., Lawler, J., et al. (2020). Environmental Influences on Dall's Sheep Survival. *The Journal of Wildlife Management*, *84*(6), 1127–1138. https://doi.org/10.1002/jwmg.21873

Wayand, N. E., Stimberis, J., Zagrodnik, J. P., Mass, C. F., & Lundquist, J. D. (2016). Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA. *Journal of Geophysical Research: Atmospheres*, *121*(17), 9929–9942. https://doi.org/10.1002/2016JD025387

Wayand, N. E., Clark, M. P., & Lundquist, J. D. (2017). Diagnosing snow accumulation errors in a rain-snow transitional environment with snow board observations. *Hydrological Processes*, *31*(2), 349–363. https://doi.org/10.1002/hyp.11002

Chapter 6. FINE-GRAIN WINTER PRECIPITATION CLASSIFICATION USING VISUAL AND AUDIO INFORMATION

Publication history: This study was co-authored with Laura Prugh. It was accepted as a paper on the non-archival track to the 11th Workshop on Fine-grain Visual Categorization as part of the 2024 Conference on Computer Vision and Pattern Recognition (available at: https://sites.google.com/view/fgvc11/papers?authuser=0) . At the time this dissertation was submitted, no version of this manuscript was in review, accepted or published elsewhere.

Abstract: Differentiating rain from snow is important for snow modeling and wildlife monitoring, but limited on-the-ground sources are available. We present the first dataset for fine-grained winter precipitation classification from camera trap images using combined visual and audio information. Using images from camera traps and audio signatures from passive acoustic recorders (AudioMoths), we present a library of 872 annotated images for rain, snow, sleet, and no precipitation. When validated using nearby weather stations, we show that this method can distinguish between rain and snow, especially for moderate to high-intensity events (> 0.8 mm/hr). Furthermore, we demonstrate how this library can be used to develop three baseline models trained on image-only information, audio-only information, and combined image and audio information. We show that models can better detect winter weather precipitation events using combined audio and image inputs than with either signal alone, obtaining a final balanced accuracy of 72%.

Keywords: camera traps, rain, snow, acoustics, AudioMoth, machine learning

6.1 Introduction

Differentiating between active winter rainfall and snowfall is essential for accurate snow modeling and predicting ecological impacts on animal movement, foraging, and survival. Incorrect precipitation type in snow models during the early winter season decreases model accuracy for the remaining winter (Wayand et al., 2017). Notably, winter rain followed by refreezing can create an icy layer and restrict vegetation access for grazing animals, leading to catastrophic die-offs in case studies (Dolant et al., 2016; Hansen et al., 2019; Stien et al., 2010). Weather stations are the primary mode for winter weather monitoring, but stations often do not differentiate precipitation type and locations are not always optimized for regional studies or in rugged terrain (Lundquist et al., 2008; Boelman et al., 2019). Specific precipitation detection devices, called "disdrometers," can differentiate between rain and snow (Wayand et al., 2016), but these devices are expensive and large, and therefore difficult to scale up. Additionally, winter precipitation is spatially variable, switching between rain, snow, hail, or sleet depending on micro-climate factors such as forest canopy, elevation, and humidity. Thus, at present, no low-cost, easy-to-install solution exists for fine-scale observation of winter weather events. Especially as warming temperatures increase winter weather variability at fine spatial- and temporal- scales (Cohen et al., 2015), new tools are urgently needed to monitor for and differentiate among rain and snow events.

Camera traps provide a solution for fine-scale weather observation, because they can be set up in networks in hard-to-reach places and scheduled to take images at regular intervals (Burton et al., 2015). Previous work has demonstrated the ability of camera traps to record meteorological variables, such as overcast, sunshine, snow, and hail, based on ground conditions (Alison et al., 2024). Other work has documented the presence of precipitation in camera traps

without differentiating among precipitation types (Dertien et al., 2017). However, to the best of our knowledge, no dataset or model exists for detecting active rainfall or snowfall in camera traps. Computer vision models, such as those using images from traffic cameras and in-vehicle cameras, have detected both rainfall and snowfall (Bossu et al., 2011; Khan & Ahmed, 2019), but success can depend on camera parameters (Garg & Nayar, 2007). Camera traps present a new domain for rain and snow detection, as camera traps take photos with infrared flash, resulting in low-light and grey-scale images, which can make information extraction more difficult (Fig. 6.1; Beery et al., 2018). Additionally, in forests and varied terrain, there can be additional obstacles, such as rocks or trees, that prevent visual cues from reaching the camera, leading to incorrect weather detection.

Given difficulties in detecting precipitation from camera traps, additional data may be useful to ensure detection and differentiation between rain and snow. Passive acoustic recorders are a suite of tools designed to measure ambient sound to monitor wildlife, but they have been increasingly used to monitor biodiversity and surrounding environmental variables (Buxton et al., 2018; Hill et al., 2018). One recording device, the AudioMoth (from Open Acoustic Devices), has supplemented camera trap data for detecting external variables (Hill et al., 2019), and pre-existing models demonstrate the ability to detect rain from AudioMoth signatures (Metcalf et al., 2020). However, no study has yet to analyze the improvement in winter precipitation detection when using AudioMoths and camera traps together. For precipitation detection, AudioMoths and camera traps may supplement each other where the other is limited. Specifically, camera traps may better detect snow due to the bright and white appearance of snow droplets, but fail to detect rain, which is often transparent and hard to see. Conversely, AudioMoths may better detect rain, which is loud when interacting with ground objects, but fail

to detect to snow, which has a dampening effect. The two tools combined may also better identify sleet events, which are tricky to classify using current methods from the field and have characteristics of both rain and snow (Lundquist et al., 2008). Hence, we provide the first annotated dataset containing rain, snow, and sleet information from camera traps and AudioMoths, validated with nearby weather station data. We demonstrate that this dataset can train and test convolutional neural network models, establishing a baseline accuracy for future model development.

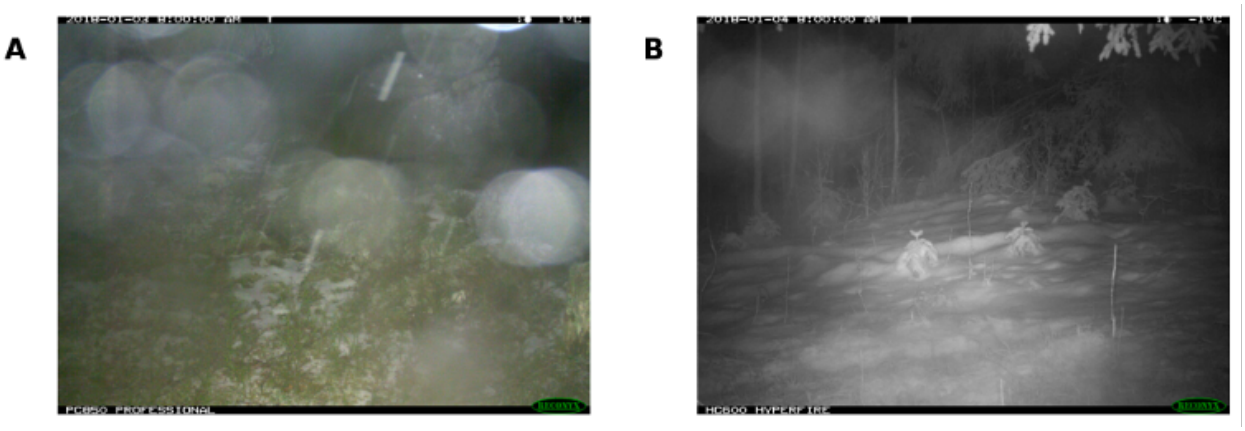


Figure 6.1 Camera trap images A) and B) both indicate precipitation, but additional information is necessary for fine-grain differentiation (i.e., identifying rain, snow, or sleet).

6.2 Methods

6.2.1 *Data Collection*

Between February and April 2023, we deployed an AudioMoth (Open Acoustic Devices model v1.1.0) alongside two camera traps (Reconyx model PC900) in two different sites in boreal Norway (Table 6.1). The southern site (latitude: 61.4º, longitude: 11.1 º, elevation: 318 m a.s.l.) and the northern site (latitude: 63.4º, longitude: 10.3º, elevation: 286 m a.s.l.) represent contrasting winter conditions. The southern site, in Inland county, Norway, experiences dry,

continental winters, with temperatures often well below freezing. Snowfall is common, but intense rainfall events are becoming more frequent. The northern site, in Trondelag county, Norway, experiences maritime winter conditions, with temperatures hovering around freezing, and precipitation often switching between rain and snow in the same storm (Pall et al., 2019). Dates were non-continuous as we had access to only one AudioMoth, and we switched locations to broaden weather events collected. The cameras were installed near each other but facing opposite directions to reduce field deployment efforts while diversifying image background conditions. Both cameras and AudioMoths were set to an hourly timelapse at the top of each hour, so that each hour had an image and a corresponding audio file. AudioMoths recorded for 30 seconds every hour and were programmed to a sampling rate of 32 kHz, matching previous studies for the mean frequency of rain (Metcalf et al., 2020).

Table 6.1 Audio and camera deployment dates and sample sizes.

County	Dates	Audio Samples	Image Samples
Inland	February 2-13	270	537
Trondelag	February 22-24	35	70
Trondelag	March 26-27	13	26
Inland	April 19-24	120	239

6.2.2 *Labeling and Validation*

Images and audio signatures were manually labeled by one person (KB). The image and the corresponding audio file were displayed at the same time, and the image was plotted in RGB and the audio file was plotted using the raw intensity values from the audio waveform relative to

time using Python's *scipy* package (Virtanen et al., 2020). The categories for precipitation matched categories used by processed disdrometer observations (e.g., snow, rain, sleet, or no precipitation) (Wayand et al., 2017). Following labeling, we summarized average temperature, decibel value, and pitch. Temperature was extracted from camera metadata, and decibel value and pitch were extracted using Python's *torchaudio* package (Yang et al. 2021). We selected decibels to mimic the scale of the human ear, and because it is recommended best practice for AudioMoth processing (Hill et al., 2019). For digital audio signals, loud decibel values start at 0 dB, with quieter values in the negatives. We selected pitch to assess any differences in perceived frequencies, after converting from the sample rate.

As an external source of validation, we compared our labels to precipitation detected from nearby weather stations. The weather station near the southern site was 4.1 km away and at the same elevation (i.e., 262 m versus 260 m). The weather station near the northern site was 3.2 km away and at a near-similar elevation (i.e., 283 m versus 290 m). The distance of our camera and AudioMoth sites represented the closest possible distance to the weather station while also adhering to land permissions in both areas. The weather stations did not record precipitation type, but instead recorded hourly precipitation (in mm/hr) and temperature (in Celsius). Following approaches used in regional and national Norway models (Saloranta, 2012), we classified precipitation as rain when temperature was >1ºC, and snow when temperature was < 1ºC. We combined our sleet and snow classifications to simply "snow" for dataset validation. We calculated overall accuracy (number of correct predictions / total number of predictions) of our labeling compared to weather station observations. We calculated precision, recall, and f-1 score (the harmonic mean for precision and recall) for each class. We also used a normalized confusion matrix to inspect for instances when we labeled images for rain or snow but the

weather station observed no precipitation (i.e., false positives), instances when we labeled for no precipitation, but the weather station observed rain or snow (i.e., false negatives), and instances when we mixed rain or snow labels. A normalized confusion matrix shows overall distribution of class labels after dividing by number of true instances for each label and allows for easier interpretation when classes are imbalanced (Fawcett, 2006).

Given the difficulty and importance of detecting winter rain (Harpold et al., 2017), we evaluated our ability to label for rain events with the combined audio and visual signal. To do so, we took all the instances when the weather station observed rain and identified whether we also labeled for rain. This metric is referred to as the "sensitivity" of our rain labeling. We counted a value as a true positive (1) if we labeled the combined image and audio signature for rain and counted a value as false negative (0) if we did not label for rain. We then fit a logistic regression model with the true positive or false negative value as the response and the corresponding precipitation, as measured by the weather station, as the predictor using Python's *sklearn* package (Pedregosa et al., 2011). We inspected the model output to determine if there was increasing sensitivity in our labeling with increasing precipitation amount. We used the Youden's Index, or the maximum difference between the true positive and false negative rate (i.e., top left corner of logistic regression model; Youden, 1950), to identify the precipitation threshold after which we are likely to label a combined image and audio signature correctly.

6.2.3 *Model Training*

After correcting for the misidentified labels identified by the weather stations from 6.4 (n = 9 misidentifications), we prepared three convolutional neural network (CNN) models using our library to assess improvement in model predictions when combining image and audio

information to improve model prediction abilities. The models represented the three combinations of image and audio information, including a model trained on only images, only audio signals, or combined image and audio information. The *image-only model* used the images downsized to 224 x 224 pixels and converted to grey-scale as the input data. The *audio-only model* used all audio samples downsampled to 16 kHz and converted to normalized 224 x 224 Mel-frequency spectrograms. A Mel-frequency spectrogram composes both the frequency and amplitude information as a matrix and is a common way to pre-process audio information for machine learning inputs (Chalmers et al., 2021; Vargas-Masis et al., 2021). To fit the required 3-channel input for the CNN model architecture, we repeated the spectrogram three times, or the same spectrogram in each input channel. Finally, the *image + audio model* combined image and audio information as input. We conducted the same pre-processing as the *image-only model* and *audio-only model*, but we replaced the last channel (i.e., third channel) of the input with the Mel-frequency spectrogram of the audio signature (see example in Fig. 6.2). We conducted an 80/10/10 training, validation, and testing split across models. We ensured that the testing split was the same across all three models for consistent performance comparison. All models were trained using a ResNet50 model, pre-trained on ImageNet, using Python's *pytorch* package (Paszke et al., 2019). We trained all models on 300 epochs, with the option for early stopping if validation accuracy did not increase after five epochs. We used a batch size of 12, a stochastic gradient descent optimizer, learning rate of 0.001, and a weight decay of 0.001. We also weighted classes due to the class imbalance, reflecting the inverse proportions of the classes normalized between 0 and 1 (i.e., no precipitation: 0.04; rain: 0.94; and snow: 0.31). We compared precision, recall, and f-1 score across all classes for each model. To determine the best

model, we calculated balanced accuracy, or the average of the recalls, which is a better metric for accuracy with imbalanced datasets (Whytock et al., 2021).

Figure 6.2 Example input data for our *image+audio* model. The ResNet-50 model architecture requires input data to have 3 channels, so we used the first two channels for the red and green channels of the RGB camera image, and for the third channel, we used the spectrograph from the AudioMoth recording.

6.3 Results

6.3.1 *Agreement with Weather Stations*

Overall, snow, rain, sleet, and no precipitation events demonstrated different visual and audio signatures (Fig. 6.3, Table 6.2). As expected, snow was brighter, quieter (i.e., lower decibel value), and occurred at colder temperatures. Rain was louder (i.e., higher decibel value), occurred at higher temperatures, and droplets appeared transparent. “No precipitation” images had audio information similar to snow, and visual information similar to rain. Interestingly, the minimum temperature for rain (0.6°C) closely matched, although was slightly higher than, the 1ºC temperature threshold for rain used by regional snow models in Norway (Saloranta, 2012). Although our sample size was small, a slightly higher average temperature for rain matched

findings that temperature thresholds for rain can depend on micro-climates (Lundquist et al., 2008). We calculated pitch information but found no clear pattern and large standard deviations across classes (Appendix E1). This finding might indicate that pitch depends on other weather characteristics, such as wind, that would be separate from precipitation events. Sleet information contained characteristics in between rain and snow for both temperature and loudness (Appendix E1).

Table 6.2 Summary information of winter precipitation from combined AudioMoths and camera trap observations, including number of samples, average temperature in Celsius reported from the camera trap, and average loudness in decibels reported from the AudioMoth. Snow and sleet samples are combined as snow.

	No Precipitation	Rain	Snow
Samples	732	35	87
Temperature ºC	-2.2 ± 9.4	2.0 ± 1.4	-5.0 ± 5.7
Decibels (dB)	-82.2 ± 7.1	-73.0 ± 6.6	-82.3 ± 7.6

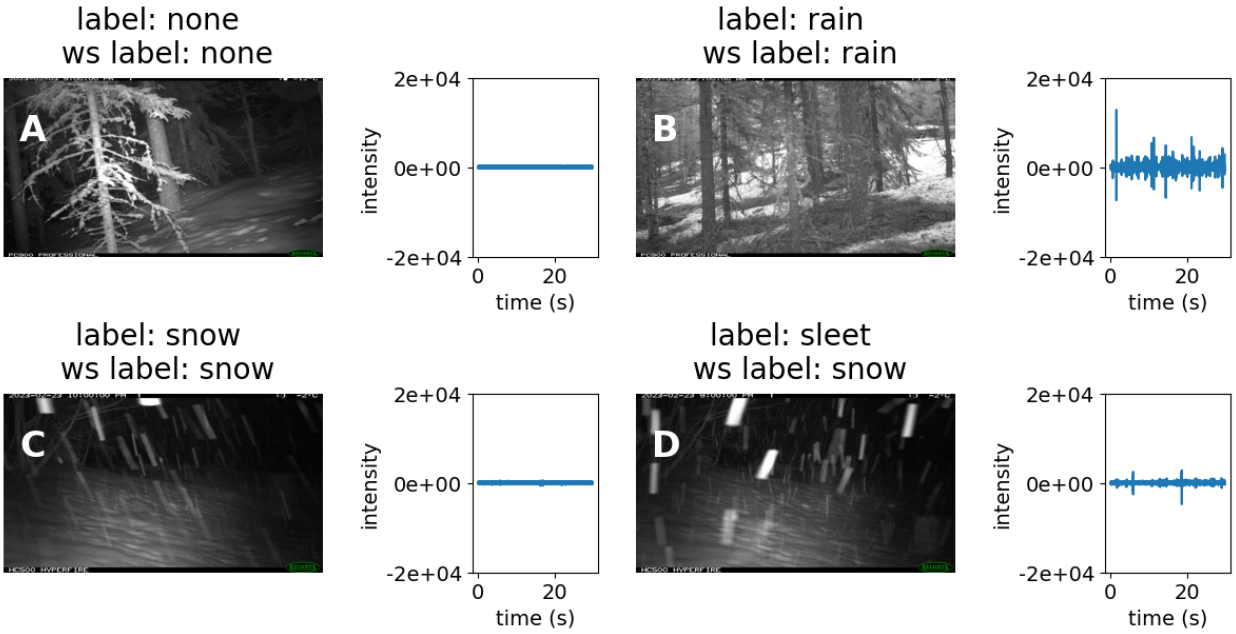


Figure 6.3 Example labels for A) no precipitation, B) rain, C) snow, and D) sleet with raw intensity sound values plotted against time (i.e., 30 seconds for each audio sample). We combined snow and sleet into one classification for dataset validation.

We found an overall accuracy of 93% when comparing manually labelled images to weather station information. For each class, we found f-1 scores of 95% for no precipitation, 72% for rain, and 74% for snow (Table 6.3). Inspection of the confusion matrix reveals that we labeled 5% of the images (n = 39; top row of Fig. 6.3A) as rain or snow, but the weather station detected no precipitation (i.e., false positives). Upon second inspection, in 3 cases, we confused no sound for falling snow although there was no visual information to suggest snow. In 6 cases, we confused drippings as rain. In 30 cases, however, second inspection confirmed the presence of rain and snow, either by observing bright, white snow droplets (i.e., snow) or hearing rain in the image. The discrepancy between our labels and weather station observations for these images suggest that, as stated in 6.1, precipitation is spatially variable, and a 3 to 4 km distance between

weather station and observation site was likely too far during some storm systems for the sites to always match.

In contrast to the false positives, we had 19 instances when we missed precipitation in labeling, but the weather station detected precipitation (i.e., false negatives). These false negatives corresponded to 12% of rain images and 17% of snow images (furthest left column of Fig. 6.4A). Precipitation was lower than 0.5 mm/hr in all cases ($\mu = 0.18 \pm 0.09$ mm/hr). Upon second inspection, we could not discern visually or with audio cues the presence of rain or snow. Rain < 0.7 mm/hr is categorized as light precipitation (National Meteorological Library and Archive 2023), suggesting that low intensity precipitation may not be detected by this method. Alternatively, as with the false positives, the weather may have been different between the weather station and the camera. We had 5 instances where we misclassified rain as snow (15% of rain images, see Fig. 6.4A), and 0 instances where we misclassified snow as rain. Inspecting these misclassifications revealed that the weather station reported temperatures ranging from 1.3 to 2.5 ºC, suggesting that the conditions at the camera and audio were possibly mixed, and separating between rain and snow is difficult.

Fitting a logistic regression for all true rain events using sensitivity as the response and precipitation amount as the predictor variable indicated that weather stations and our assigned labels were more likely to agree with more intense rain events (Fig. 6.4B). The Youden's Index suggested that we detected rain when intensity was > 0.8 mm/hr. This finding is advantageous for wildlife ecologists and snow hydrologists, as it suggests that this method can detect more intense rain events, which is the priority for snow and wildlife monitoring (Hansen et al., 2019; Yang et al., 2023).

Table 6.3 Precision, recall, f-1 scores, and sample sizes for winter weather events when validated using nearby weather stations. Overall accuracy between weather stations and manual labels was 93%.

Class	Precision	Recall	f-1 score	Samples
No Weather	0.95	0.97	0.96	732
Rain	0.74	0.71	0.72	35
Snow	0.83	0.68	0.74	105

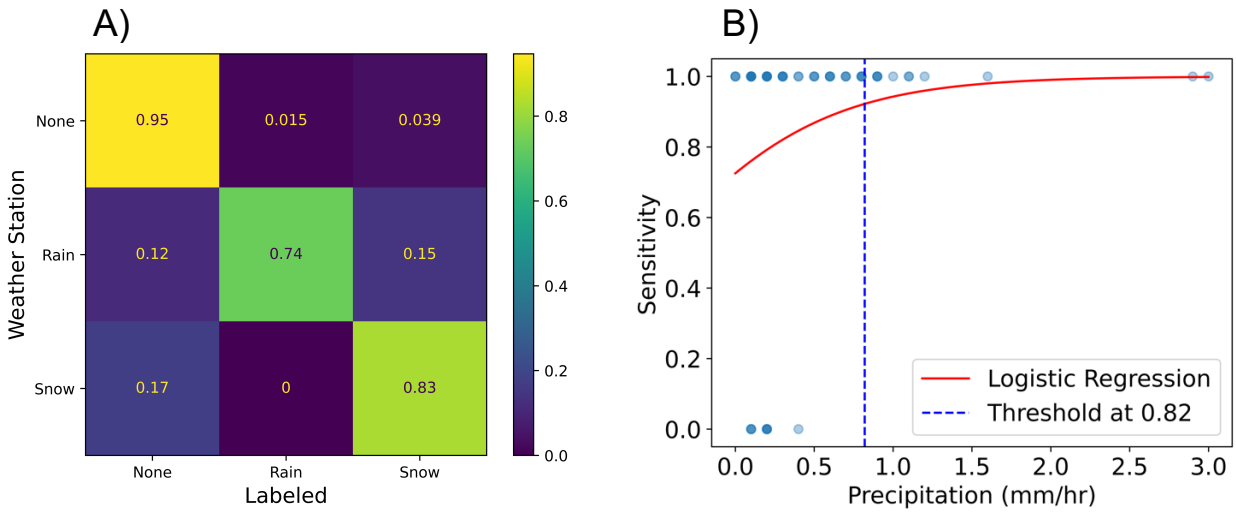


Figure 6.4 A) Normalized confusion matrix for validation using nearby weather station data. We divided row by the total number of samples (None = 732 samples, Rain = 35 samples, Snow = 105 samples) to ease interpretation of percent of correct predictions. B) There was some discrepancy between weather station and manually labeled data for detecting rain but increasing agreement for with precipitation intensity. The Youden index of a logistic regression suggests that we can detect rain above 0.8 mm/hr (dashed line).

6.3.2 *Baseline Models*

Only the *image-only* and the *image + audio models* were able to distinguish among precipitation classes (Table 6.4), and the *image + audio model* performed with a better overall balanced accuracy (72%). The model performed reasonably well across all three classes with f1-scores for no precipitation, rain, and snow equal to 0.93, 0.75, and 0.5, respectively. Surprisingly, the *audio-only model* failed to detect rain in all test cases, despite weighted classes and the distinct signature of rain. It is possible the failed predictions was due to few examples in the training set (n = 17 instances).

The *image + audio model* predicted five images as snow or rain when in fact there was no precipitation (see examples in Fig. 6.5A). In three of the five cases, the audio indicates the presence of wind, which the model may have learned is associated with storms. Additionally, in one image the snow "sparkles," which can occur under cold temperatures and after fresh snowfall (Horton & Jamieson, 2017). The model may have learned that the light refraction on the snow surface indicates snowfall, rather than the precipitation droplets themselves. Of the four samples of rain in the test set, the *image + audio model* predicted two incorrectly as snow (Fig. 6.5B). The temperature at the camera at both incorrect images was 1 ºC, a boundary condition, which is difficult to correctly differentiate even with current methods (Lundquist et al., 2008). There were no instances when the model misclassified rain for snow. There were eight cases when the model predicted no precipitation when in fact there was snow (Fig. 6.5C). In five of the eight cases, the weather station reported precipitation < 0.3 mm/hr during this event, so it is possible there were not enough droplets in the image for the model to detect. We hope that with more training examples, the *image + audio model* will continue to improve and reduce these errors.

Table 6.4 Baseline *image-only*, *audio-only*, *image+audio models*. We present the balanced accuracy (BA), and precision (P), recall (R), and f-1 score for all three classes. Bold indicates best model.

Model	Class	Precision	Recall	f-1 score	Samples
Image-only	None	0.97	0.96	0.95	73
	Rain	1.0	0.6	0.75	5
	Snow	0.57	0.44	0.5	9
	B.A,	--	--	0.67	87
Audio-only	None	0.85	0.99	0.91	36
	Rain	0	0	0	3
	Snow	0	0	0	5
	B.A.	--	--	0.67	44
Image+Audio	None	0.96	0.90	0.93	73
	Rain	1.0	0.6	0.75	5
	Snow	0.4	0.66	0.5	9
	B.A.	--	--	**0.72**	87

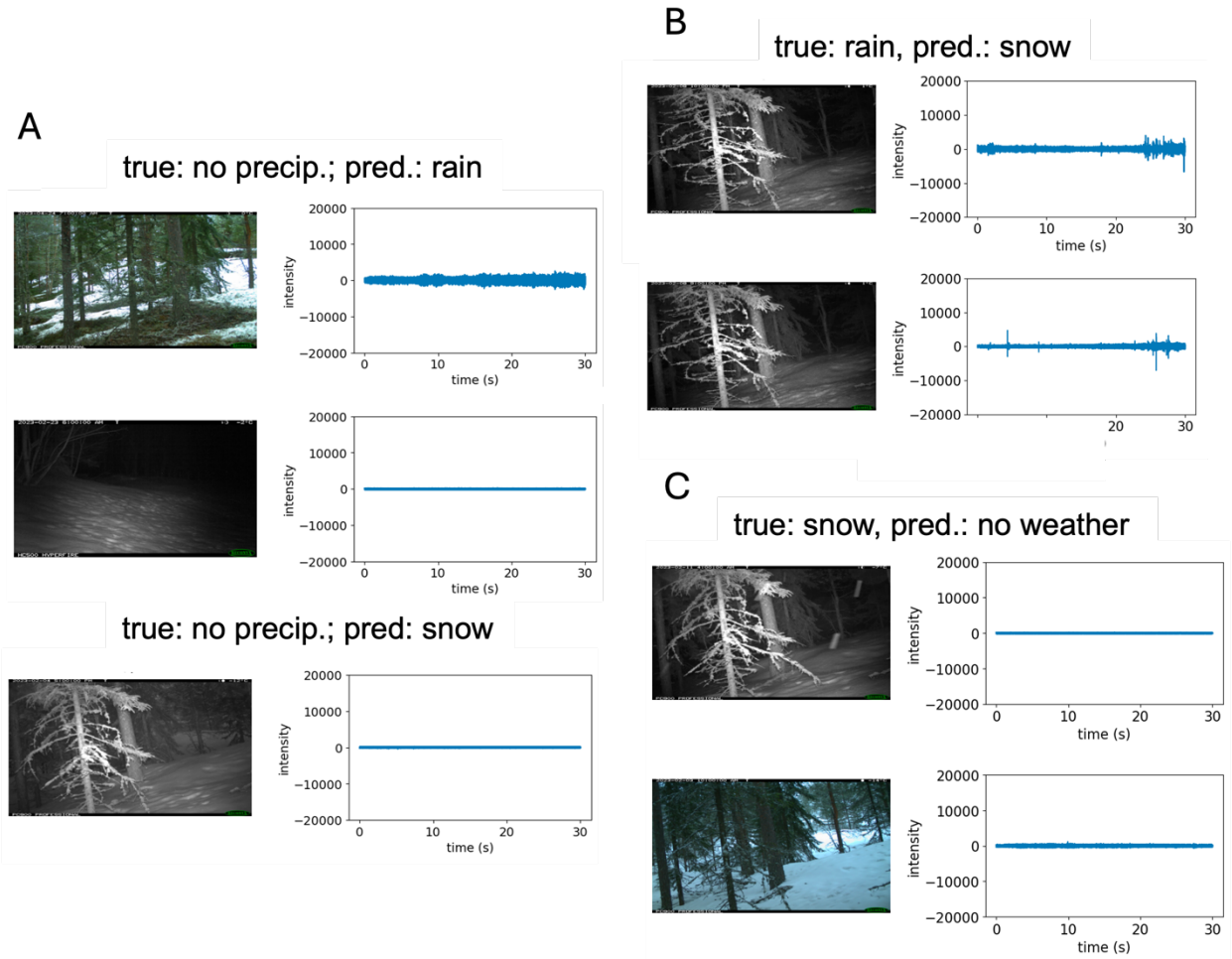


Figure 6.5 Error detections from *image+audio model* for A) no precipitation, B) rain, and C) snow.

6.4 Conclusion and Future Work

We provide a validated dataset and initial models for fine-grained winter precipitation using a novel bi-modal approach. We found that snow, rain, and "no precipitation" events have different profiles when audio and visual information are combined. Specifically, snow is bright and white, but quiet. Rain is transparent and creates loud decibel values when hitting surrounding surfaces. "No precipitation" events are both quiet and precipitation droplets are absent. While we

may miss smaller precipitation events, we can detect precipitation events above 0.8 mm/hr, and with a more direct approach than relying on temperature thresholds. We occasionally confused drippings for rain, and no snow for snow, but our baseline models suggest that combined image and audio information improves prediction, with an overall balanced accuracy of 72%. Future work should continue to grow both the training and testing set for further experimentation with larger sample sizes. The models and dataset presented here demonstrate that combining AudioMoth and camera trap data offers a new solution for differentiating between rain and snow using tools that wildlife ecologists and snow hydrologists already employ.

6.5 References

Alison, J., Payne, S., Alexander, J. M., Bjorkman, A. D., Clark, V. R., Gwate, O., et al. (2024). Deep learning to extract the meteorological by-catch of wildlife cameras. *Global Change Biology*, *30*(1), e17078. https://doi.org/10.1111/gcb.17078

Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in Terra Incognita. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Vol. 11220, pp. 472–489). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-01270-0_28

Boelman, N. T., Liston, G. E., Gurarie, E., Meddens, A. J. H., Mahoney, P. J., Kirchner, P. B., et al. (2019). Integrating snow science and wildlife ecology in Arctic-boreal North America. *Environmental Research Letters*, *14*(1), 010401. https://doi.org/10.1088/1748-9326/aaeec1

Bossu, J., Hautière, N., & Tarel, J.-P. (2011). Rain or Snow Detection in Image Sequences Through Use of a Histogram of Orientation of Streaks. *International Journal of*

Computer Vision, *93*(3), 348–367. https://doi.org/10.1007/s11263-011-0421-7

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., et al. (2015). REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Buxton, R. T., Lendrum, P. E., Crooks, K. R., & Wittemyer, G. (2018). Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance. *Global Ecology and Conservation*, *16*, e00493. https://doi.org/10.1016/j.gecco.2018.e00493

Cohen, J., Ye, H., & Jones, J. (2015). Trends and variability in rain-on-snow events. *Geophysical Research Letters*, *42*(17), 7115–7122. https://doi.org/10.1002/2015GL065320

Dertien, J. S., Doherty, P. F., Bagley, C. F., Haddix, J. A., Brinkman, A. R., & Neipert, E. S. (2017). Evaluating dall's sheep habitat use via camera traps: Dall's Sheep Habitat Use. *The Journal of Wildlife Management*, *81*(8), 1457–1467. https://doi.org/10.1002/jwmg.21308

Dolant, C., Langlois, A., Montpetit, B., Brucker, L., Roy, A., & Royer, A. (2016). Development of a rain-on-snow detection algorithm using passive microwave radiometry. *Hydrol. Process.*, 14.

Garg, K., & Nayar, S. K. (2007). Vision and Rain. *International Journal of Computer Vision*, *75*(1), 3–27. https://doi.org/10.1007/s11263-006-0028-6

Hansen, B. B., Pedersen, Å. Ø., Peeters, B., Le Moullec, M., Albon, S. D., Herfindal, I., et al. (2019). Spatial heterogeneity in climate change effects decouples the long-term dynamics of wild reindeer populations in the high Arctic. *Global Change Biology*, *25*(11), 3656–3668. https://doi.org/10.1111/gcb.14761

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. *Methods in Ecology and Evolution*, *9*(5), 1199–1211. https://doi.org/10.1111/2041-210X.12955

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Khan, M. N., & Ahmed, M. M. (2019). Snow Detection using In-Vehicle Video Camera with Texture-Based Image Features Utilizing K-Nearest Neighbor, Support Vector Machine, and Random Forest. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(8), 221–232. https://doi.org/10.1177/0361198119842105

Lundquist, J. D., Neiman, P. J., Martner, B., White, A. B., Gottas, D. J., & Ralph, F. M. (2008). Rain versus Snow in the Sierra Nevada, California: Comparing Doppler Profiling Radar and Surface Observations of Melting Level. *Journal of Hydrometeorology*, *9*(2), 194–211. https://doi.org/10.1175/2007JHM853.1

Metcalf, O. C., Lees, A. C., Barlow, J., Marsden, S. J., & Devenish, C. (2020). hardRain: An R package for quick, automated rainfall detection in ecoacoustic datasets using a threshold-based approach. *Ecological Indicators*, *109*, 105793. https://doi.org/10.1016/j.ecolind.2019.105793

National Meteorological Library and Archive. *Open to Everyone: An Overview of the National Meteorological Library and Archive Services.* Met Office, https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/library/publications/factsheets/factsheet_3-water-in-the-

atmosphere_2023.pdf

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., …Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pall, P., Tallaksen, L. M., & Stordal, F. (2019). A Climatology of Rain-on-Snow Events for Norway. *JOURNAL OF CLIMATE*, *32*, 22.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.

Saloranta, T. M. (2012). Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model. *The Cryosphere*, *6*(6), 1323–1337. https://doi.org/10.5194/tc-6-1323-2012

Stien, A., Loe, L. E., Mysterud, A., Severinsen, T., Kohler, J., & Langvatn, R. (2010). Icing events trigger range displacement in a high-arctic ungulate, *91*(3), 7.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wayand, N. E., Stimberis, J., Zagrodnik, J. P., Mass, C. F., & Lundquist, J. D. (2016). Improving simulations of precipitation phase and snowpack at a site subject to cold air intrusions: Snoqualmie Pass, WA. *Journal of Geophysical Research: Atmospheres*, *121*(17), 9929–9942. https://doi.org/10.1002/2016JD025387

Wayand, N. E., Clark, M. P., & Lundquist, J. D. (2017). Diagnosing snow accumulation errors in

a rain-snow transitional environment with snow board observations. *Hydrological Processes*, *31*(2), 349–363. https://doi.org/10.1002/hyp.11002

Yang, Z., Chen, R., Liu, Y., Zhao, Y., Liu, Z., & Liu, J. (2023). The impact of rain-on-snow events on the snowmelt process: A field study. *Hydrological Processes*, *37*(11), e15019. https://doi.org/10.1002/hyp.15019

Yang, Y.-Y., Hira, M., Ni, Z., Chourdia, A., Astafurov, A., Chen, C., Yeh, C.-F., Puhrsch, C., Pollack, D., Genzel, D., Greenberg, D., Yang, E. Z., Lian, J., Mahadeokar, J., Hwang, J., Chen, J., Goldsborough, P., Roy, P., Narenthiran, S., Watanabe, S., Chintala, S., Quenneville-Bélair, V, & Shi, Y. (2021). TorchAudio: Building Blocks for Audio and Speech Processing. arXiv preprint arXiv:2110.15018.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Chapter 7. CONCLUSIONS

This dissertation analyzed the ability of remote cameras to operate as snow and rain monitoring devices, connecting the snow information from cameras to satellites and wildlife activity models. Chapter 2 established the utility of cameras to supplement satellites on cloudy days and for customizing satellite products for regional applications. Chapter 3 identified the use of machine learning to extract snow information, specifically snow depth, providing an efficient approach for scaling up remote cameras for snow depth monitoring in the future. Chapters 4 and 5 presented applications of how snow information and environmental metadata extracted from the camera image can improve wildlife activity models (Chapter 4) or remote sensing (Chapter 5), highlighting new insights into changes in wildlife diel activity due to snow conditions and passive microwave accuracy for rain-on-snow events. Chapter 6 set a new direction for cameras by integrating camera traps with another novel wildlife tool, the AudioMoth, finding improved rain and snow differentiation when using both tools together compared to with cameras alone. Recurring themes emerged from these chapters regarding the ability of camera traps to observe snow. First, as described in Chapters 2 and 5, cameras can supplement satellite observations, especially in dense forest regions, despite differences in scale. Cameras have a field-of-view of ~20-30 m, and the camera's small field-of-view can be an asset if the user is interested in information at fine-scale resolutions. Second, as demonstrated in Chapters 3, 5, and 6, machine learning will be essential to extract environmental information from images. Creative train/test splits, such as the split by state (Washington or Colorado) I performed in Chapter 3, allowed for insight into the model performance for new datasets, and set an example of adjusting machine learning development pipelines for ecological applications. Third, and lastly, snow presents a complex phenomenon both to measure with on-the-ground and remote sensing tools (Chapters 2, 3, 5, and 6) and to interpret as a driver of animal activity (Chapter

4), and monitoring snow will continue to be important to uncover novel wildlife behavior patterns and the trade-offs animals must make during winter.

This study set out to extract snow properties from cameras in an extensive Norwegian wildlife network, and I was successful when leveraging manual inspection, computer vision, and machine learning techniques. As demonstrated in Chapter 2, manual inspection for snow cover provided success, however, given the over 2,000 camera trap studies worldwide (Steenweg et al., 2017), successful snow information extraction will hinge on consistent, robust approaches. I achieved adequate accuracy in models for snow depth (Chapter 3), rain detection (Chapter 5), and rain vs. snow detection (Chapter 6), demonstrating that machine learning approaches offer advantages for labeling camera imagery. For Chapter 3, I also processed and published two datasets for snow depth, totally over 20,000 labeled images. The first dataset is for the snow poles in Washington state, hosted on the University of Washington data archive, and the second dataset is for the NASA snow poles in Colorado state, hosted on the National Snow and Ice Data Center archive. For Chapter 5, I developed a dataset of over 20,000 images for rain or not rain. For Chapter 6, I developed a dataset of almost 1,000 images labeled for rain, snow, or no weather. The datasets can contribute to future models, and the models and datasets presented here will lower the training and labeling time for future researchers.

I developed models that were sufficiently accurate, but I found that developing machine learning models was time-intensive. In Chapters 3 and 5, I invested years of effort to assemble training sets, and in some cases, as I demonstrated in Chapter 6, confidently differentiating between rain and snow required the addition of an AudioMoth to improve performance in both labeling by humans and predictions from the model. Labeling and training time will likely decrease as more ecologists and hydrologists adopt machine learning pipelines. However, new models from the computer science community may benefit ecology and hydrology as well. As of dissertation submission, recent developments in large language models (LLMs) present the ability to identify

features of an image using a prompt without the need for additional labeling (Farrell et al., 2024). These “foundational” models trained to conduct various tasks may replace smaller models trained to perform one task, such as those presented here. However, current LLM models cost ~$0.17 per image, which may be unfeasible for processing thousands of images for academic, government, and non-profit research groups. Smaller models, such as the ones in this study that can be fine-tuned on one’s desktop or laptop, will likely play an important role in achieving scalability in ecology. In addition, some information from wildlife camera imagery may be difficult to perceive from visual data alone, as we demonstrated in Chapter 6 with the combined AudioMoth plus camera approach, and custom models that combine signals will ensure trustworthiness moving forward.

Besides the contributions to machine learning, utilizing a wildlife camera network for remote sensing supplementation was the first of its kind. It provided new insights into remote sensing accuracy in Norway, a densely forested and cloudy region that presents one of the most challenging landscapes for remote sensing products. Using cameras as a source of validation, I showed the decrease in MODIS Cloud Gap Filled products after 3 days and the limited abilities of AMSR2 for detecting rain. I then showed how cameras can be a source of supplementation. For example, cameras can be used to customize MODIS snow cover products to create binary snow/no snow maps and combined with passive microwave satellite observations to improve accuracy for winter rain detection. The study demonstrates the untapped potential of other wildlife camera networks (i.e., wildlifeinsights.org) for environmental monitoring, and future work should continue to utilize cameras in hard-to-reach places with a critical need for observation.

Utilizing a wildlife network for environmental monitoring also achieved new insights into animal activity. I connected information derived from the camera, including time of day, time of year, and location to estimate hourly snow hardness, an important and understudied snow property in wildlife studies. I was able to connect patterns in roe deer diel activity with freeze-thaw conditions and demonstrate increased morning activity on harder snow in deep snow areas. Recent work has

highlighted the need to understand animal diel activity in response to environmental conditions, and this study provides evidence for the role of cameras in measuring environmental variables as well.

The work presented in this dissertation confirmed that cameras can operate as weather stations for snow detection, and opportunities for future work demonstrate a potential for even more applications. I recommend future work extract snow information from camera images using global, publicly available datasets (e.g., wildlifeinsights.org) to validate satellites and monitor snow in places where we have not been able to before. I also recommend future work identify the feasibility of extracting other snow properties from cameras, beyond the properties of snow cover, depth, hardness, and precipitation presented here. For example, camera images may indicate the presence or absence of snow surface impurities (e.g., small organic matter) and age of snow (e.g., fresh fallen snow or aged snow), both of which are driving factors in snow surface albedo. Accurate account of snow interception by trees improves hydrology modeling, and previous work has documented the ability to identify snow-canopy interception in timelapse images (Lumbrazo et al. 2022). Given the abundance of animals that inhabit in forested areas, wildlife cameras may provide a rich dataset to identify snow-canopy interception information.

With the ability to report visual signatures of snow and weather, cameras provide important environmental information for ecologists and hydrologists to measure winter conditions and assess the impacts of snow on wildlife. Researchers can use the framework presented here for connecting cameras with satellite validation and wildlife studies in other regions, and they can use the models presented here towards future, larger models to extract snow information from camera imagery. Evaluating camera traps as ground-based remote sensing devices showed that a tool, traditionally from ecology, yielded new insights into multiple disciplines, including wildlife science, machine learning, remote sensing, and snow science.

APPENDIX A

Supplemental Information for Chapter 2

Evaluating MODIS snow products using an extensive wildlife camera network

Appendix A1: 12 tiles from MOD10A1F were processed for cloud-gap filled analysis: h18v00, h18v01, h18v02, h18v03, h18v04, h19v00, h19v01, h19v02, h19v03, and h19v04. All tiles were downloaded in the default projection, SR:ORG 6974.

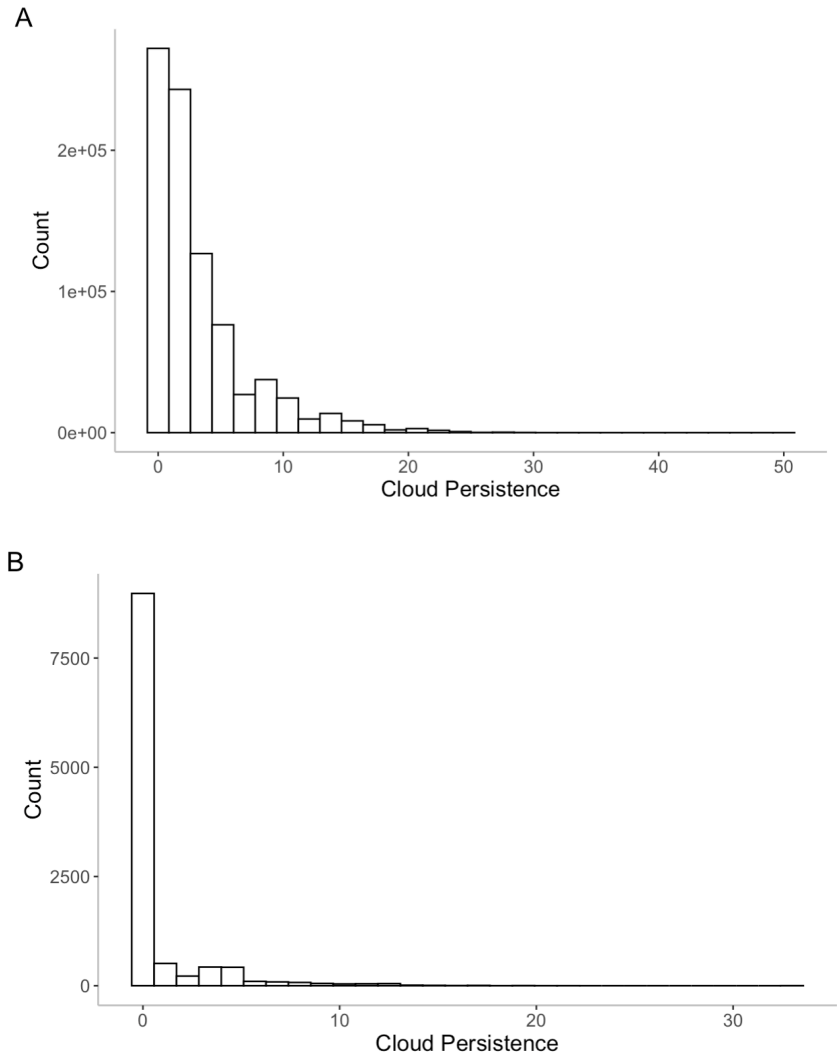


Figure A2: Histogram of cloud persistence values from the (A) MOD10A1F product and (B) our dataset. The MOD10A1F represents cloud persistence values from 1,181 cameras for three winter seasons: 1) January 1 – March 2018, 2) October 1 2018 – March 2019, and 3) October 2019 – March 2020. The histogram below represents the datapoints used in the analysis for the effect of cloud persistence on agreement. While there are many more values on clear sky days (Cloud Persistence = 0), we included these values to assess agreement on clear and cloudy days.

Table A3: Correlation among covariates used in the general linear mixed model to assess factors that affect agreement between snow cover values from camera images and MODIS. All continuous variables (i.e., latitude, tree canopy cover, daily NDVI) were scaled by subtracting the mean and dividing by the standard deviation. Image color mode refers to color images (1) or grey-scale images (0).

	Latitude	Tree canopy cover	Daily NDVI
Tree canopy cover	-0.404	--	--
Daily NDVI	-0.281	0.0951	--
Image brightness	0.0780	-0.0267	-0.0524

Table A4: Variance inflation factors among covariates used in the general linear mixed model to assess factors that affect agreement between snow cover values from camera images and MODIS. All continuous variables (i.e., latitude, tree canopy cover, daily NDVI) were scaled by subtracting the mean and dividing by the standard deviation. Image color mode refers to color images (1) or grey-scale images (0).

	Latitude	Tree canopy cover	Daily NDVI	Image color mode
Variance inflation factor	1.19	1.02	1.16	1.01

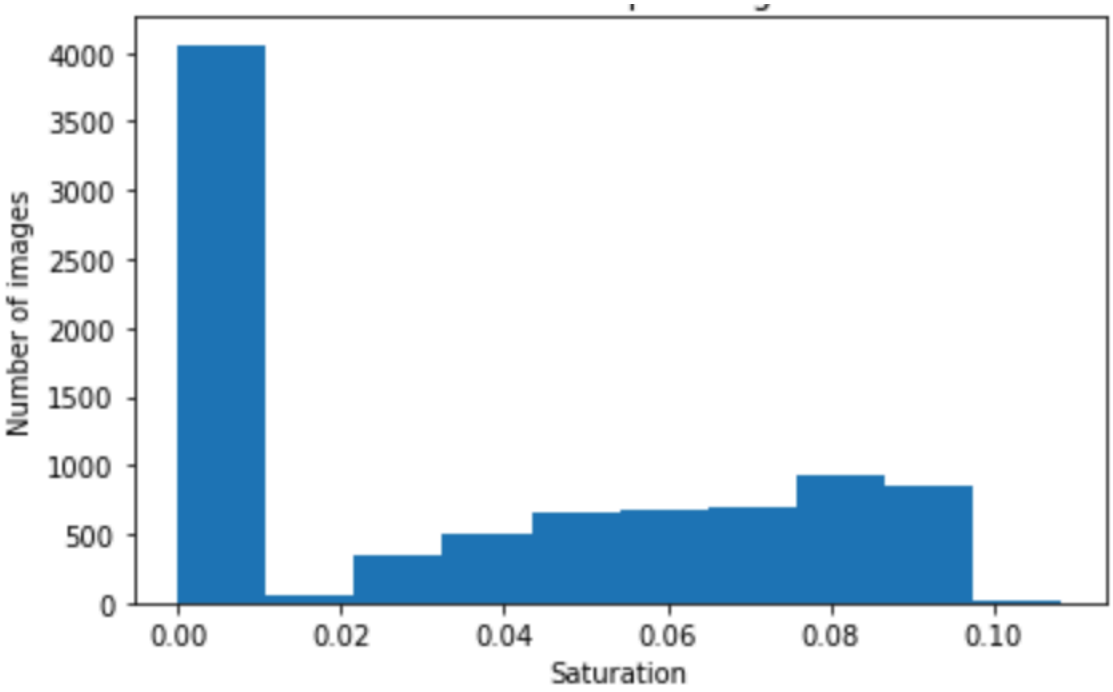


Table A5: Histogram depicting average saturation values of camera trap images. Images with an average saturation >0.02 were images taken with white flash or in daytime (i.e. color image). Images with an average saturation below <0.02 were images taken with infrared flash (i.e., grey-scale image).

Table A6: Table: Average monthly MODIS NDVI for all cameras (n = 665) for our study period (2018 – 2020). NDVI ranges between 0 and 1.

Month	Average NDVI
October	0.64
November	0.34
December	0.24
January	0.24
February	0.10
March	0.16
April	0.15

Figure A3: Average MODIS NDVI maps for each month of the study period: A) October, B), November, C) December, D) January, E) February, and F) March. For each month, NDVI was averaged for the years included in our study period. Thus, January, February and March included three years (2018-2020), and October, November and December included two years (2018 and 2019).

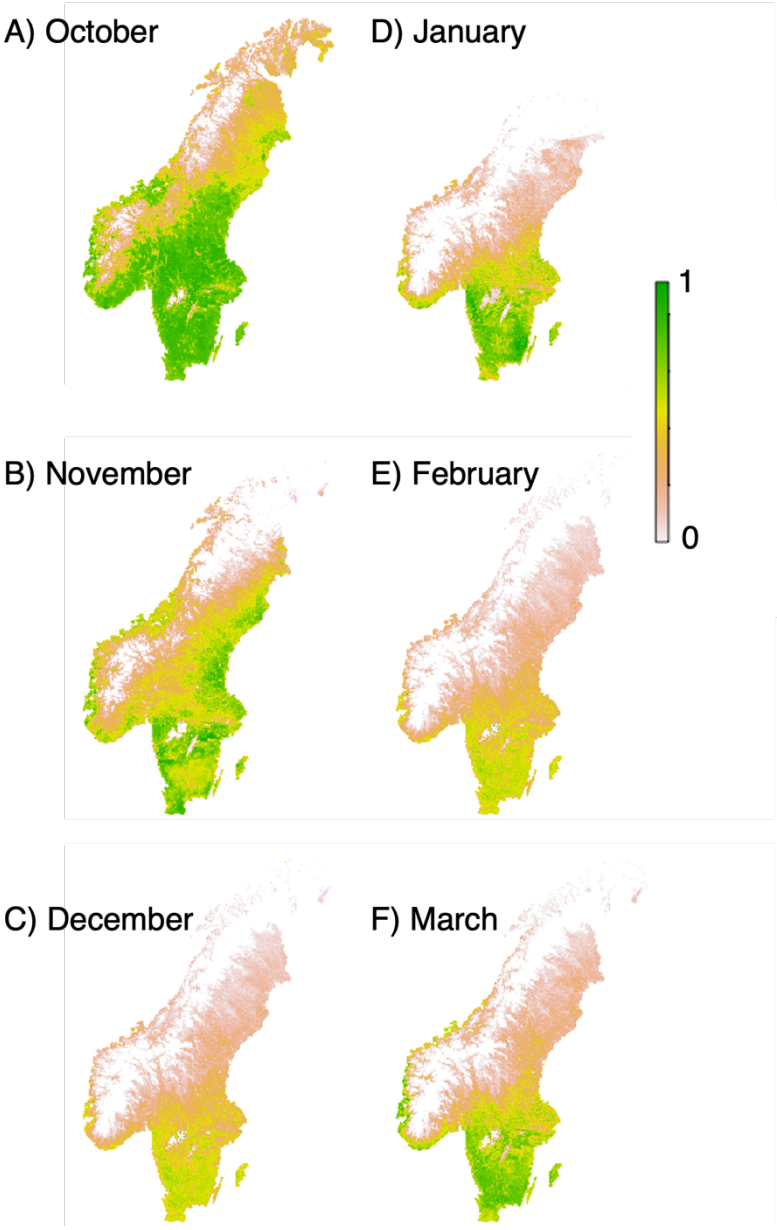


Table A7. Coefficient estimates, standard error (SE), t-values, and *p*-values from a general linear mixed model assessing factors that affect MODIS and camera agreement (n = 4,355) for months between November to March (excluding October). Continuous variables were normalized by subtracting the mean and dividing by the standard deviation prior to analysis. Image color mode is a categorical variable (1: color image; 0: grey-scale image). Camera identification was included as a random effect (n = 658).

	Estimate	SE	t-value	*p*-value
Intercept	73.12	0.46	159.43	<0.05
Latitude	-1.08	0.42	-2.54	<0.05
Daily NDVI	-0.075	0.25	-0.30	0.76
Tree canopy cover	-1.09	0.40	-2.75	<0.05
Image color mode (color image)	0.63	0.51	1.23	0.21

Table A8. Coefficient estimates, standard error (SE), t-values, and *p*-values from a general linear mixed model assessing factors that affect MODIS and camera agreement (n = 8,808). Max annual NDVI, calculated by taking the max NDVI for each of the three winter seasons (i.e., January 1 – March 2018, October 1, 2018 – March 2019, and October 2019 – March 2020) was included instead of daily NDVI. Continuous variables were normalized by subtracting the mean and dividing by the standard deviation prior to analysis. Image color mode is a categorical variable (1: color image; 0: grey-scale image). Camera identification was included as a random effect (n = 658).

	Estimate	SE	t-value	*p*-value
Intercept	76.27	0.44	174.46	<0.05
Latitude	-1.57	0.38	-4.081	<0.05
Max annual NDVI	6.18	0.23	27.00	<0.05
Tree canopy cover	-0.73	0.38	-1.90	0.058
Image color mode (color image)	7.66	0.45	17.02	<0.05

Table A9: Alternative options for thresholds to create a binary map from the MOD10A1 product and the corresponding True Positive Rate (TPR) and False Positive Rate (FPR). Users can specify different thresholds with subsequent changes in TPR and FPR.

Threshold	**True Positive Rate (Sensitivity)**	**False Positive Rate (Specificity)**
5	0.97	0.69
10	0.97	0.69
15	0.96	0.74
20	0.95	0.77
25	0.94	0.80
30	0.93	0.83
35	0.91	0.86
40	0.89	0.88
45	0.86	0.91
50	0.80	0.94
55	0.72	0.96
60	0.61	0.97
65	0.48	0.98
70	0.32	0.99
75	0.19	1.00
80	0.09	1.00
85	0.04	1.00
90	0.01	1.00
95	0.00	1.00

A)

B)

C)

Figure A4: Simulated model residuals using the *DHARMa* package in R indicated non-significant dispersion, but a total of 106 outliers out of 8,808 images (< 2%) were identified (Hartig, 2022). Outliers were defined as observations that either had extremely high or low agreement compared to the simulated range. Upon inspection, examples of outliers included images with A) narrow field of view, B) active weather, and C) low snow depths.

APPENDIX B

Supplemental Information for Chapter 3

Snow depth extraction from time-lapse imagery using a keypoint deep learning model

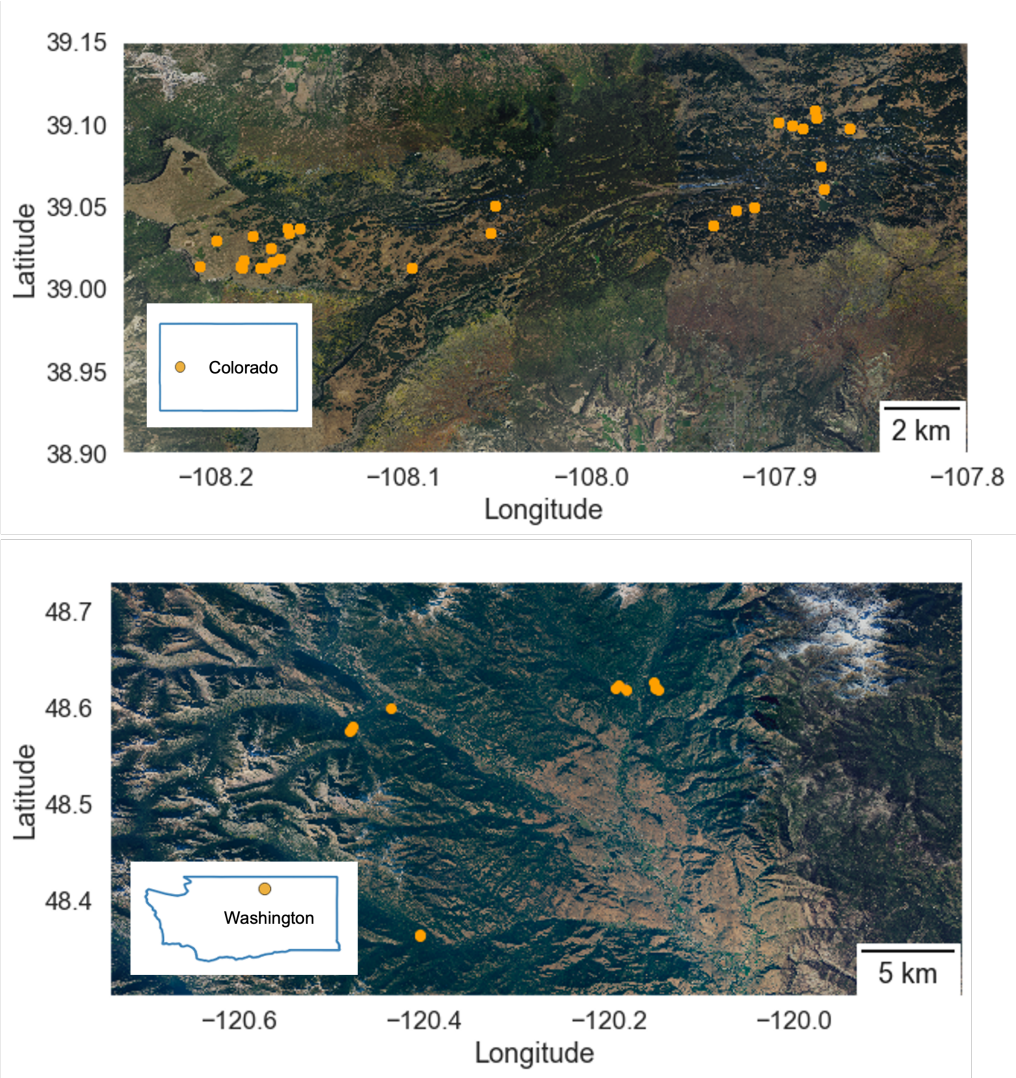
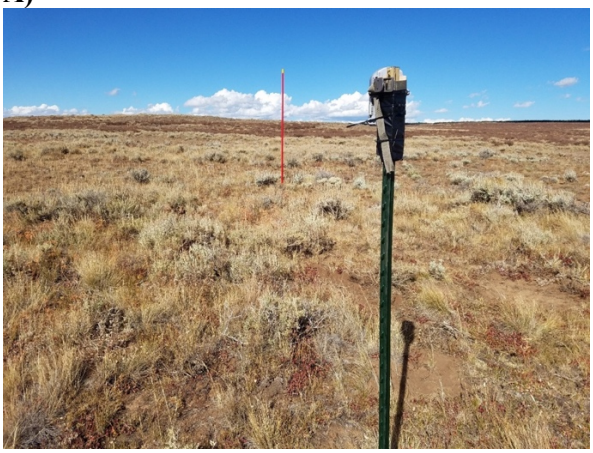


Figure B1. Locations of 20 sites in Grand Mesa, CO, (top) and 12 sites in Okanogan, WA, (bottom) used to train and test a keypoint detection model. The Grand Mesa, CO study area covers 470 km^2 and the Okanagan County, WA, study area comprises approximately 11,000 km^2. Aerial imagery is from the National Agriculture Imagery Program (NAIP).

A)



B)

Figure B2. A) Example set-up on Grand Mesa, Colorado. Wingscape cameras (WCT-00126 TimeLapseCam Pro were installed on t-post driven into the ground. B) Example set-up in Okanogan County, Washington. Reconyx cameras were set up on trees using a cable lock to secure the camera in place.

Table B1. Pole heights (cm) and pixel to centimeter conversion using either the yellow band on the top 10 cm of pole or the full pole height. Pole height varied depending on installation design and local snow conditions. Camera sites in Colorado (i.e., cameras with naming convention E** or W**) were all 304.8 cm tall. Camera sites in Washington (i.e., cameras with naming conventions CHE*, TWISP-*-**, or CUB-*-**) varied between 1-2 m depending on local snow conditions. Average and standard deviation (SD) of the length in pixels of the yellow band on the top 10 cm of the pole and full length of the pole show that the standard deviation for the full pole was consistently smaller.

Camera	**Pole Height (cm)**	**Top 10-cm (cm/pixels)**		**Full Pole (cm/pixels)**	
		Mean	**SD**	**Mean**	**SD**
E6A	304.8	0.108	0.00570	0.110	0.000225
E6B	304.8	0.132	0.0117	0.126	0.000303
E9A	304.8	0.120	0.0101	0.122	0.000472
E9E	304.8	0.135	0.0126	0.137	0.000459
E9F	304.8	0.137	0.0130	0.137	0.00608
W1A	304.8	0.151	0.00813	0.149	0.000660
W2A	304.8	0.147	0.00878	0.136	0.000552
W2B	304.8	0.187	0.0234	0.177	0.001902
W5A	304.8	0.130	0.00998	0.128	0.000737
W6A	304.8	0.148	0.01185	0.144	0.000814
W6B	304.8	0.165	0.0158	0.158	0.00112
W6C	304.8	0.203	0.0237	0.187	0.000662
W8A	304.8	0.185	0.0162	0.164	0.000611
W8C	304.8	0.433	0.0620	0.228	0.00196
W9A	304.8	0.115	0.00897	0.114	0.000267
W9B	304.8	0.125	0.00738	0.130	0.000619
W9C	304.8	0.145	0.0143	0.138	0.00133
W9D	304.8	0.139	0.00847	0.126	0.000462
W9E	304.8	0.135	0.0128	0.130	0.00037
W9G	304.8	0.1401	0.0146	0.131	0.000634
BUNKHOUSE-01	151.5	0.294	0.0405	0.269	0.00161
CEDAR-H-01	125	0.2009	0.0208	0.195	0.000646
CEDAR-L-01	125	0.17139	0.0140	0.194	0.00188
CEDAR-M-01	125	0.1666	0.0234	0.168	0.000938
CUB-H-01	126	0.2565	0.0110	0.253	0.00125
CUB-H-02	154	0.225	0.0230	0.212	0.000861
CUB-L-02	122.1	0.221	0.0153	0.224	0.00103
CUB-M-01	121.7	0.2736	0.0159	0.281	0.00115
CUB-M-02	151.5	0.186	0.0180	0.224	0.00560
CUB-U-01	151	0.234	0.0670	0.203	0.000946
TWISP-R-01	120.7	0.182	0.0107	0.202	0.00120
TWISP-U-01	121.5	0.215	0.0345	0.222	0.000783

Table B2. Error and difference in snow depth in centimeters for automated estimation of depth from snow poles for each camera. Error was defined as the Euclidean distance between the target point and the predicted point. The difference in snow depth was defined as the snow depth calculated from the manual labels minus the snow depth calculated from the automated labels.

Camera	Top Error (pixels)	Bottom Error (pixels)	Residual Error (cm)
E6A	1.17 ± 0.82	2.01 ± 1.34	-2.46 ± 2.49
E6B	1.44 ± 0.57	2.39 ± 0.43	-1.01 ± 2.74
E9A	1.18 ± 0.47	1.73 ± 0.88	-2.05 ± 1.95
E9E	1.14 ± 0.41	1.96 ± 0.69	-3.57 ± 2.12
E9F	1.58 ± 0.46	1.85 ± 0.72	-1.62 ± 1.69
W1A	0.81 ± 0.75	0.89 ± 0.57	-0.95 ± 2.38
W2A	0.79 ± 0.37	1.60 ± 0.69	-2.91 ± 2.41
W2B	0.86 ± 0.33	1.77 ± 0.85	-4.54 ± 2.84
W5A	1.00 ± 1.18	1.22 ± 1.21	-0.76 ± 2.48
W6A	0.89 ± 0.35	1.18 ± 0.66	-1.10 ± 2.87
W6B	0.62 ± 0.44	0.90 ± 0.57	-0.99 ± 1.85
W6C	1.41 ± 0.37	1.92 ± 0.55	-3.40 ± 2.47
W8A	0.65 ± 0.28	1.10 ± 0.74	-0.55 ± 2.71
W8C	1.18 ± 1.26	1.83 ± 0.73	-4.66 ± 8.18
W9A	1.03 ± 0.43	1.73 ± 0.89	-2.22 ± 2.37
W9B	0.79 ± 0.45	1.70 ±1.15	-1.31 ± 3.31
W9C	1.10 ± 0.43	1.75 ±0.83	-2.23 ± 2.45
W9D	0.90 ± 0.70	1.28 ± 0.99	-1.15 ± 2.45
W9E	0.88 ± 0.65	1.42 ± 0.82	-1.57 ± 2.50
W9G	1.62 ± 0.59	2.88 ± 1.06	-3.74 ± 2.78

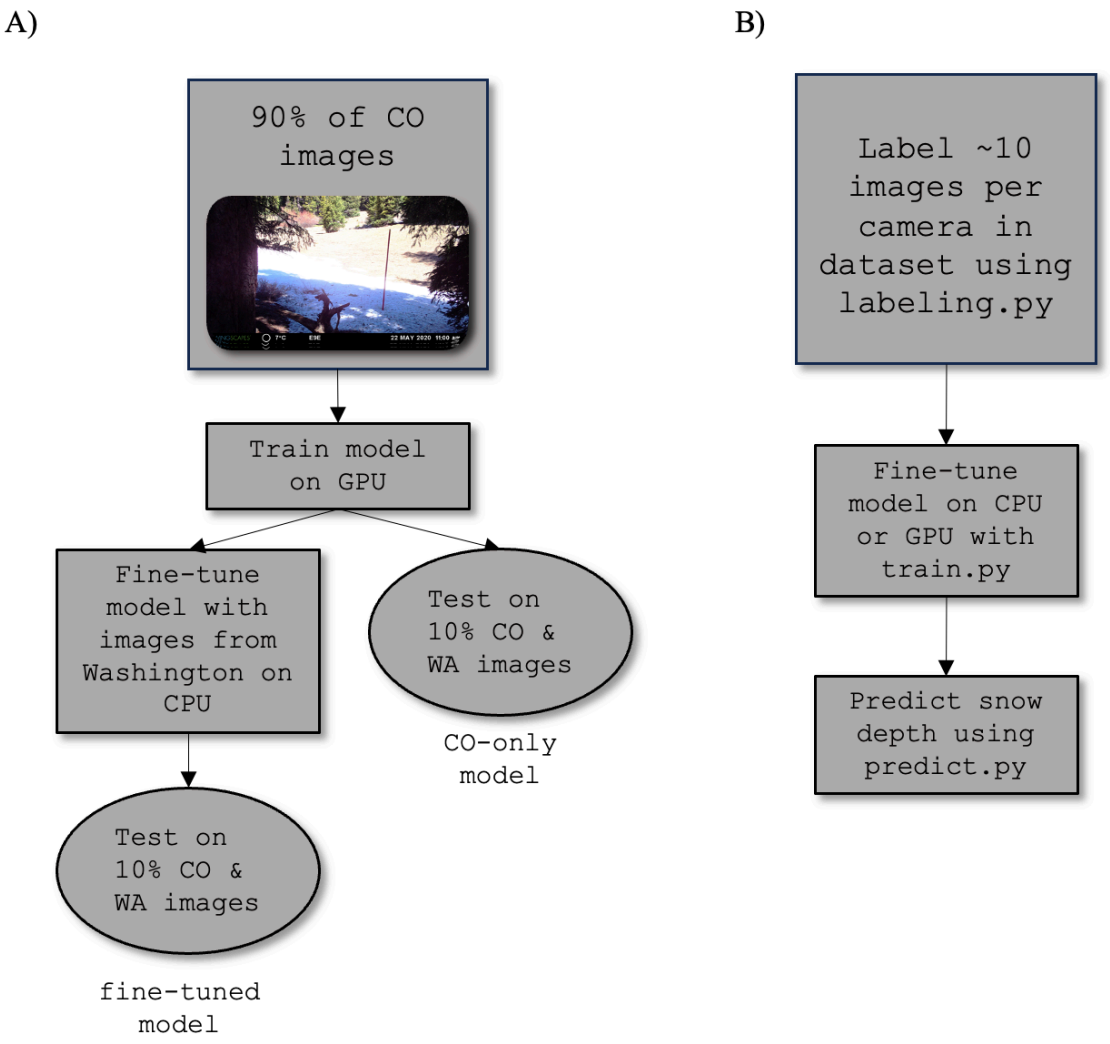


Figure B3. Workflow for experiment for the CO-only model and the fine-tuned model, as well as proposed workflow for user. The WA-only model would be derived from training the Washington images without using the CO-only model weights. The CO+WA model would be to combine both sets of images during training.

Table B3. To identify the reprojection error we found the downscaling factor for the width and height of each camera to convert from the native resolution to the resolution at 224x224. We list the width and height of all the images for each camera (i.e., width and height columns). We then multiplied that factor by the cm/pixels for each pole (i.e., cm/pixels conversion column) to get the error in both the x and y direction (i.e., X and Y columns). We then took the hypotenuse of those numbers by calculating h = sqrt(x^2 + y^2), so that the final reprojection error for each camera is listed in the hypotenuse column.

Camera	**Size**		**cm/pixels conversion**		**Error (cm)**		**Hypotenuse (cm)**
	width	**height**	**Mean**	**SD**	**X**	**Y**	
E6A	6080	3420	0.110	0.000225	1.49	0.84	1.71
E6B	6080	3420	0.126	0.000303	1.71	0.96	1.96
E9A	6080	3420	0.122	0.000472	1.66	0.93	1.9
E9E	6080	3420	0.137	0.000459	1.86	1.05	2.13
E9F	6080	3420	0.137	0.00608	1.86	1.05	2.13
W1A	6080	3420	0.149	0.000660	2.02	1.14	2.32
W2A	6080	3420	0.136	0.000552	1.85	1.04	2.12
W2B	6080	3420	0.177	0.001902	2.40	1.35	2.76
W5A	6080	3420	0.128	0.000737	1.74	0.98	1.99
W6A	6080	3420	0.144	0.000814	1.95	1.10	2.24
W6B	6080	3420	0.158	0.00112	2.14	1.21	2.46
W6C	6080	3420	0.187	0.000662	2.53	1.43	2.91
W8A	6080	3420	0.164	0.000611	2.22	1.25	2.55
W8C	6080	3420	0.228	0.00196	3.09	1.74	3.55
W9A	6080	3420	0.114	0.000267	1.55	0.87	1.78
W9B	6080	3420	0.130	0.000619	1.76	0.99	2.02
W9C	6080	3420	0.138	0.00133	1.87	1.05	2.15
W9D	6080	3420	0.126	0.000462	1.71	0.96	1.96
W9E	6080	3420	0.130	0.00037	1.76	0.99	2.02
W9G	6080	3420	0.131	0.000634	1.78	1.00	2.04
BUNKHOUSE-01	2048	1440	0.269	0.00161	1.22	0.86	1.5
CEDAR-H-01	2048	1536	0.195	0.000646	0.89	0.67	1.11
CEDAR-L-01	2048	1536	0.194	0.00188	0.89	0.67	1.11
CEDAR-M-01	2048	1536	0.168	0.000938	0.77	0.58	0.96
CUB-H-01	1920	1080	0.253	0.00125	1.08	0.61	1.24
CUB-H-02	2048	1536	0.212	0.000861	0.97	0.73	1.21
CUB-L-02	2048	1536	0.224	0.00103	1.024	0.77	1.28
CUB-M-01	1920	1080	0.281	0.00115	1.20	0.68	1.38
CUB-M-02	1920	1080	0.224	0.00560	0.96	0.54	1.1
CUB-U-01	2048	1536	0.203	0.000946	0.93	0.70	1.16
TWISP-R-01	1920	1080	0.202	0.00120	0.87	0.49	1.99
TWISP-U-01	2048	1536	0.222	0.000783	1.01	0.76	1.27

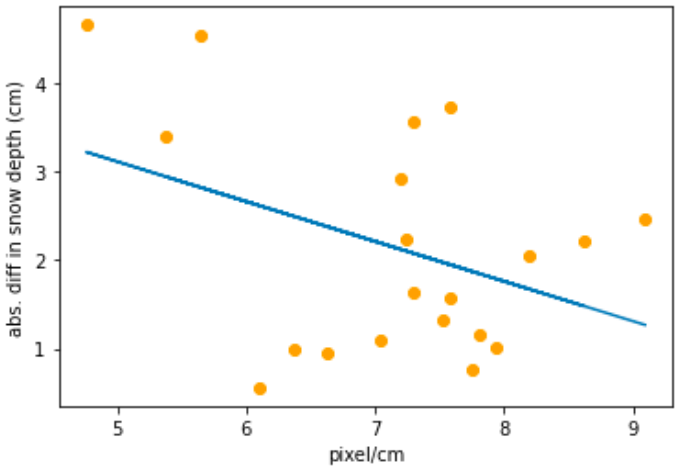


Figure B4. Measurement error (y-axis) was not significantly related to the pixel/cm conversion factor (x-axis), suggesting that the model performed well on poles both close and far from the camera (slope = -0.45, $F_{1, 20} = 3.09$, $p = 0.09$). The measurement error was calculated by subtracting the actual snow depth by the predicted snow depth. The pixel/cm conversion was found by identifying the number of pixels of a snow-free pole that represented the full length of the pole in centimeters.

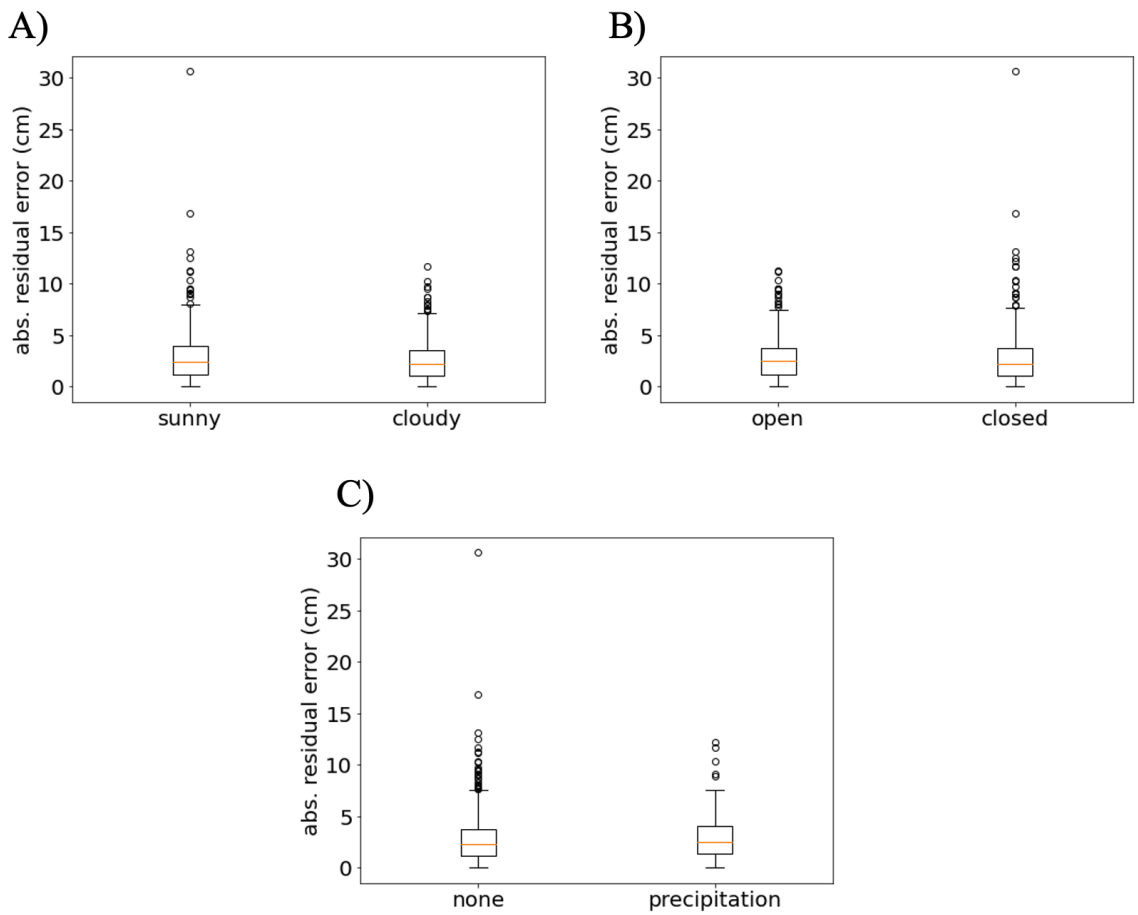


Figure B5. ANOVA and t-test results showed no differences among error means for multiple comparisons: A) sunny vs. cloudy (*t(890) = 0.89, p = 0.37*), B) closed vs. open canopy (*t(890) = -0.37, p = 0.71*), and C) no precipitation vs. precipitation (*t(890) = 1.60, p = 0.11*).

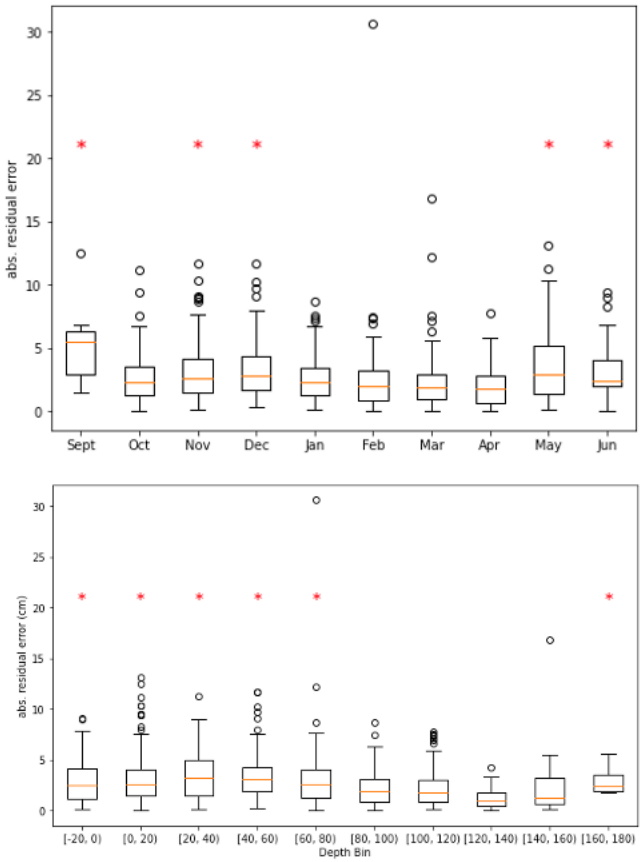
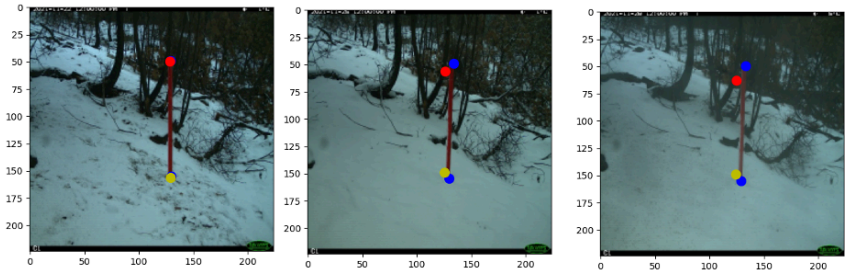
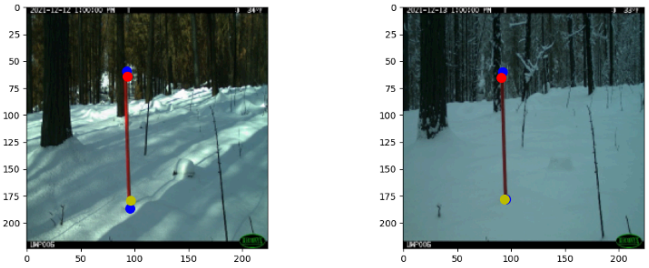


Figure B6. ANOVA results returned several significant groups, marked with an "*" to indicate significant differences from the error mean for A) month of year *($F = 12.8$, $df = 890$, $SS = 73.8$, $p < 0.05$)* and B) binned snow depth values *($F = 5.5$, $df = 890$, $SS = 277.5$, $p < 0.05$)*.

A)



B)



C)

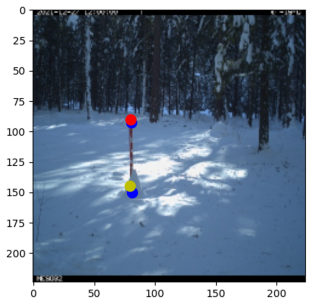


Figure B7. A) The “fine-tuned” model over-estimated snow depth in multiple cases during the early season at site CEDAR-M-01. Upon inspection, the pole tilted slightly on a slope. The slope and the pole tilt together may affect model performance before the model learned the new placement of the pole. B) High-contrast shadows at the CUB-M-02 site may have affected model performance during the early and melt season. The left image shows the model predicting the snow line where the shadow meets the pole. The right image shows the model’s ability to predict the snow depth correctly in the absence of the high-contrast shadows. C) Snow occasionally loads at the base of the pole, forming a small mound. These mounds can be up to a couple centimeters along the pole. This image suggests the fine-tuned model was able to predict snow depth in the presence of snow mounds (yellow points), but there is some discrepancy between the manually-labelled and predicted points.

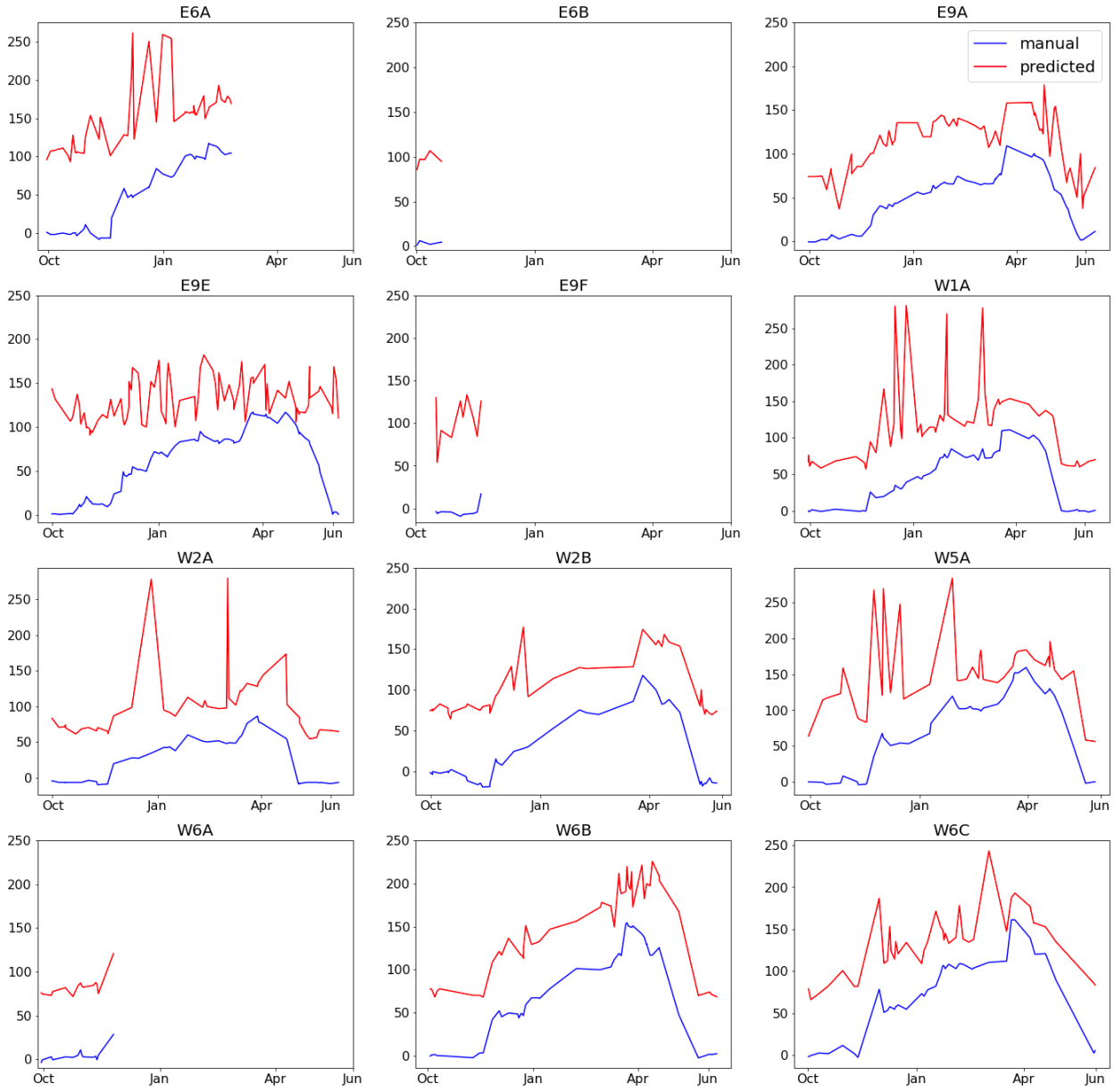


Figure B8. The fine-tuned model trained on Washington data lowers the accuracy of the model's performance on Colorado data, shown here. The predicted line shows a consistent offset with the manually labelled data and overestimation at high snow depth values.

Table B4**.** Performance of the fine-tuned model improved with more training data. The data subset indicates how many images from each of the 12 Washington cameras we used. We tested the fine-tuned model on the test dataset for each state (Colorado = 802 images, Washington = 892 images).

Data Subset	Dataset	Training time (min)	Epochs	Res. Error (cm)	MAE (cm)	R^2
1	Washington	5.7	63	27.7 ± 23.7	29.0	0.33
	Colorado			-38.1 ± 37.9	47.9	0.48
2	Washington	6.3	60	-0.8 ± 12.6	9.6	0.80
	Colorado			-70.8 ± 37.6	72.5	0.52
5	Washington	7.3	43	-2.9 ± 9.2	7.1	0.89
	Colorado			-72.3 ± 36.0	72.9	0.59
10	Washington	10.2	42	-1.4 ± 5.1	4.0	0.96
	Colorado			-67.2 ± 36.6	67.7	0.61
15	Washington	17.6	40	-4.7 ± 4.9	3.9	0.97
	Colorado			-78.0 ± 39.8	78.1	0.53
20	Washington	14.3	31	-2.6 ± 4.2	3.5	0.98
	Colorado			-67.8 ± 35.9	68.0	0.61
25	Washington	19.8	62	-2.0 ± 3.3	2.8	0.99
	Colorado			-55.5 ± 36.0	56.3	0.62
30	Washington	22.9	36	-1.9 ± 2.8	2.5	0.99
	Colorado			-60.1 ± 34.9	60.4	0.62

Table B5. Residual error (Res. Error), root mean square error (RMSE), mean absolute percent error (MAPE), and mean absolute error (MAE) for model when WA and CO data are combined in training. All poles were included during training except for the two poles listed for testing.

Poles for training	Poles for testing	Res. Error (cm)	RMSE (cm)	MAPE (%)	MAE (cm)
18 poles from CO dataset and 11 poles from WA dataset	E9E, TWISP-U-01	15.37	27.49	2281.7	20.35

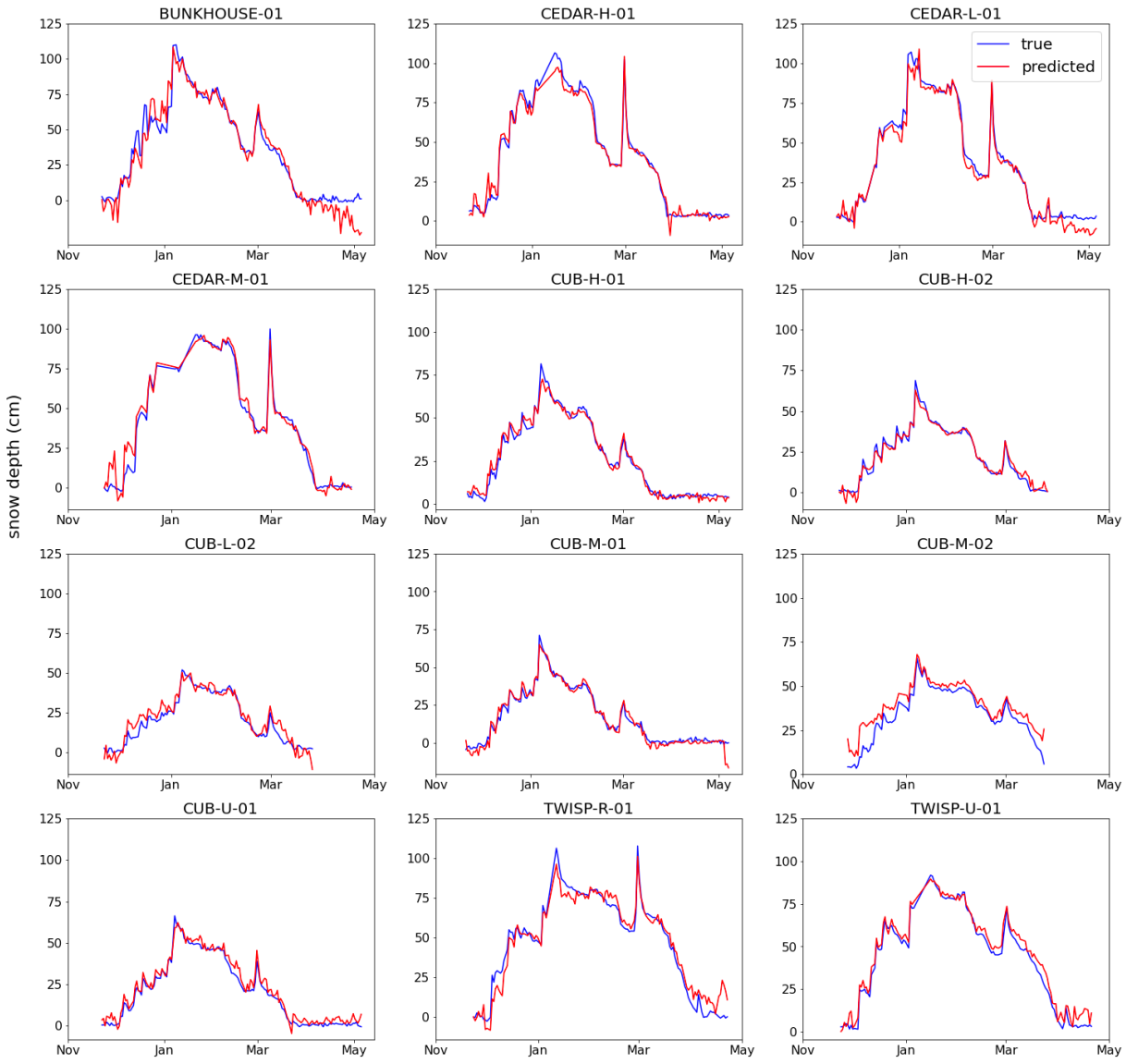


Figure B9. Predicted snow depth from model when including a maximum snow depth value from each camera into the training set in addition 10 images for every camera.

APPENDIX C

Supplemental Information for Chapter 4

Diel and seasonal cycles of snow hardness predict changes in activity patterns for roe deer but not mountain hares

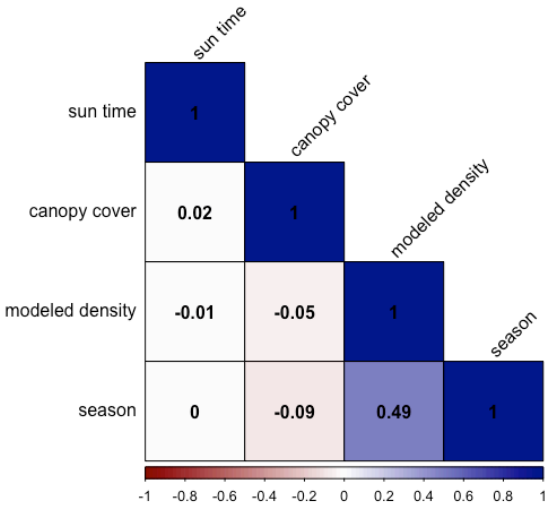


Figure C1. Correlation plot for all predictor and response variables used in the models. Sun time, canopy cover, and season were derived from field metadata. Modeled snow density was derived from Norway's daily snow model run at 1-km resolution (Saloranta et al. 2012). All variables were below the threshold of 0.5.

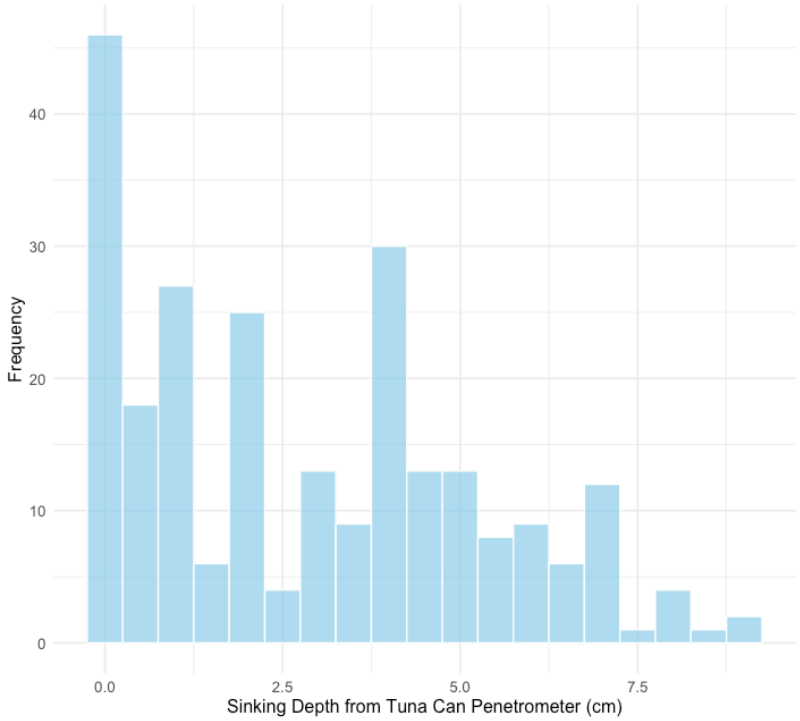


Figure C2. Distribution of snow hardness data collected from the penetrometer for snow hardness modeling. We used a Tweedie distribution to reflect the continuous, non-zero distribution of our snow hardness data.

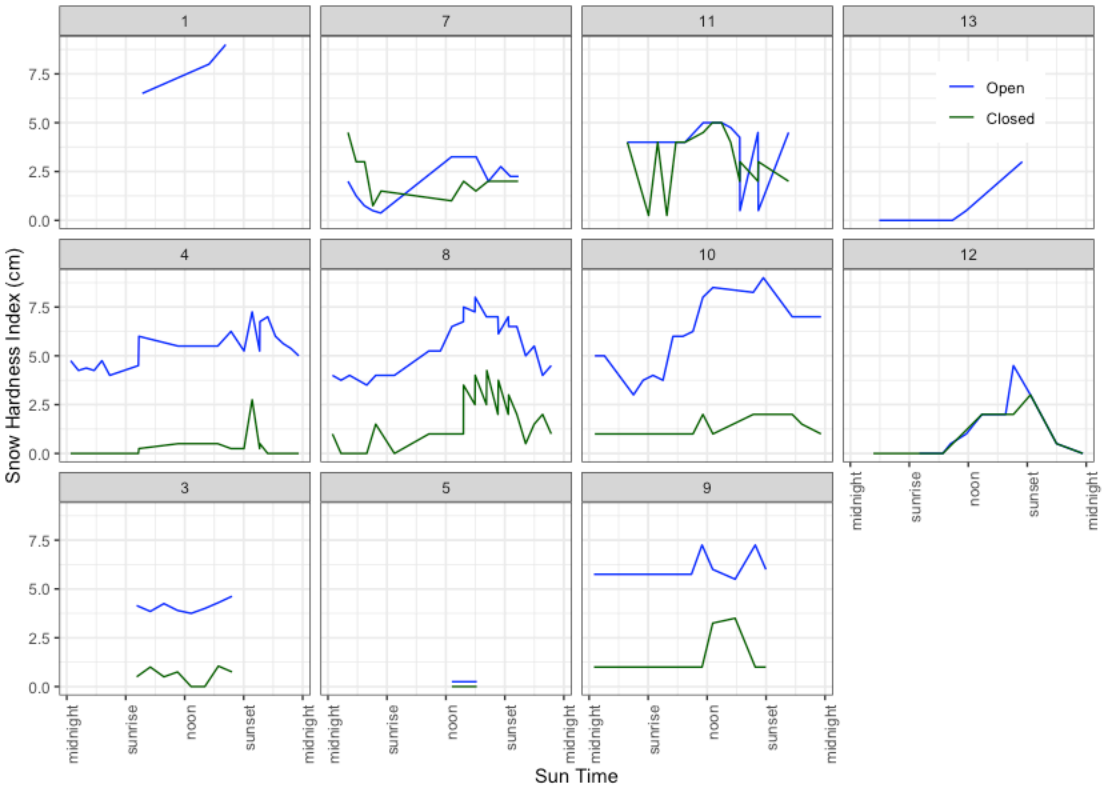


Figure C3. Hourly snow hardness index values obtained from 13 visits across three sites (*n* = 247 observations total). Snow surface hardness was indexed by dropping a 200 g penetrometer (tuna can) from a height of 50 cm and measuring the depth it sank into the snow. Higher values indicate softer snow and lower values indicate harder snow. Colors indicate whether the observation occurred in open (blue) or closed (green) canopy. Closed canopy had fewer observations due to no snow at later measurements in the spring. Visits 1, 2, 5, 9 were only daytime observations due to lack of lodging to conduct overnight observations. All times were converted to times relative to sunrise/sunset, i.e., 'sun time,' using the *suntime* package in R.

Table C1. AIC comparison table for generalized additive models predicting snow hardness with ΔAIC values < 4 from the top model. Models with interactions also include main effects.

Model	k	AIC	ΔAIC
canopy * cos(sun time) + canopy * density + cos(sun time) * season + sin(sun time) * season	11	835.7	0.00
canopy * cos(sun time) + canopy * density + canopy * sin(sun time) + cos(sun time) * season + season * sin(sun time)	12	835.8	0.03
canopy * cos(sun time) + canopy * density + cos(sun time) * season + density * season + season * sin(sun time)	12	836.3	0.61
canopy * cos(sun time) + canopy * density + canopy * sin(sun time) + density * season + season * sin(sun time) + canopy * sin(sun time)	12	836.7	1.02
canopy * cos(sun time) + canopy * density + canopy * season + cos(sun time) * season + season * sin(sun time)	12	836.8	1.03
canopy * cos(sun time) + canopy * density + canopy * season + canopy * sin(sun time) + cos(sun time) * season + season * sin(sun time)	12	837.7	1.95
canopy * cos(sun time) + canopy * density + canopy * season + cos(sun time) * season + density * season + season * sin(sun time)	13	838.2	2.49
canopy * density + canopy * sin(sun time) + cos(sun time)*season + season * sin(sun time)	11	838.5	2.77
canopy * cos(sun time) + canopy * density + canopy * season + cos(sun time) * season + season * sin(sun time) + canopy * cos(sun time) * season	12	838.8	3.07
canopy * density + cos(sun time) * season + season * sin(sun time)	10	838.9	3.13
canopy * cos(sun time) + canopy * density + canopy * season + canopy * sin(sun time) + cos(sun time) * season + density * season + season * sin(sun time)	14	839.2	3.46
canopy * density + canopy * sin(sun time) + cos(sun time) * season + density * season + season * sin(sun time)	11	839.6	3.87
canopy * density + cos(sun time) * season) + density * season + season * sin(sun time)	11	839.6	3.90
full model	17	842.6	7.30
null model	1	1259.3	423.6

Table C2. Model output summary from a generalized additive model predicting snow hardness using sun time (transformed using cosine and sine), snow density, and interactions with canopy class and season. Canopy was defined as either "open" or "closed," and season was defined as either "winter" (February and March observations) or "spring" (April and May observations). We transformed sun time using sine and cosine transformations to account for circularity. Combining sine and cosine allowed for flexible horizontal shifts along the x-axis (i.e., phase shifts) and changes in amplitude.

Coefficient	Estimate	Std. Error	T value	*P value*
Intercept	0.22	0.088	2.55	0.01
canopy (open)	1.23	0.10	13.12	< 0.05
cos(sun time)	-0.45	0.13	-3.37	< 0.05
density	0.16	0.08	1.9	< 0.05
season (late)	-2.90	0.52	-5.51	< 0.05
sin(sun time)	-0.21	0.05	-4.10	< 0.05
canopy (open) * cos(sun time)	0.31	0.15	2.11	< 0.05
canopy (open) * density	-0.31	0.09	-3.26	< 0.05
cos(sun time) * season (late)	-1.14	0.50	-2.27	< 0.05
season (late) * sin(sun time)	-2.58	0.50	-5.17	< 0.05

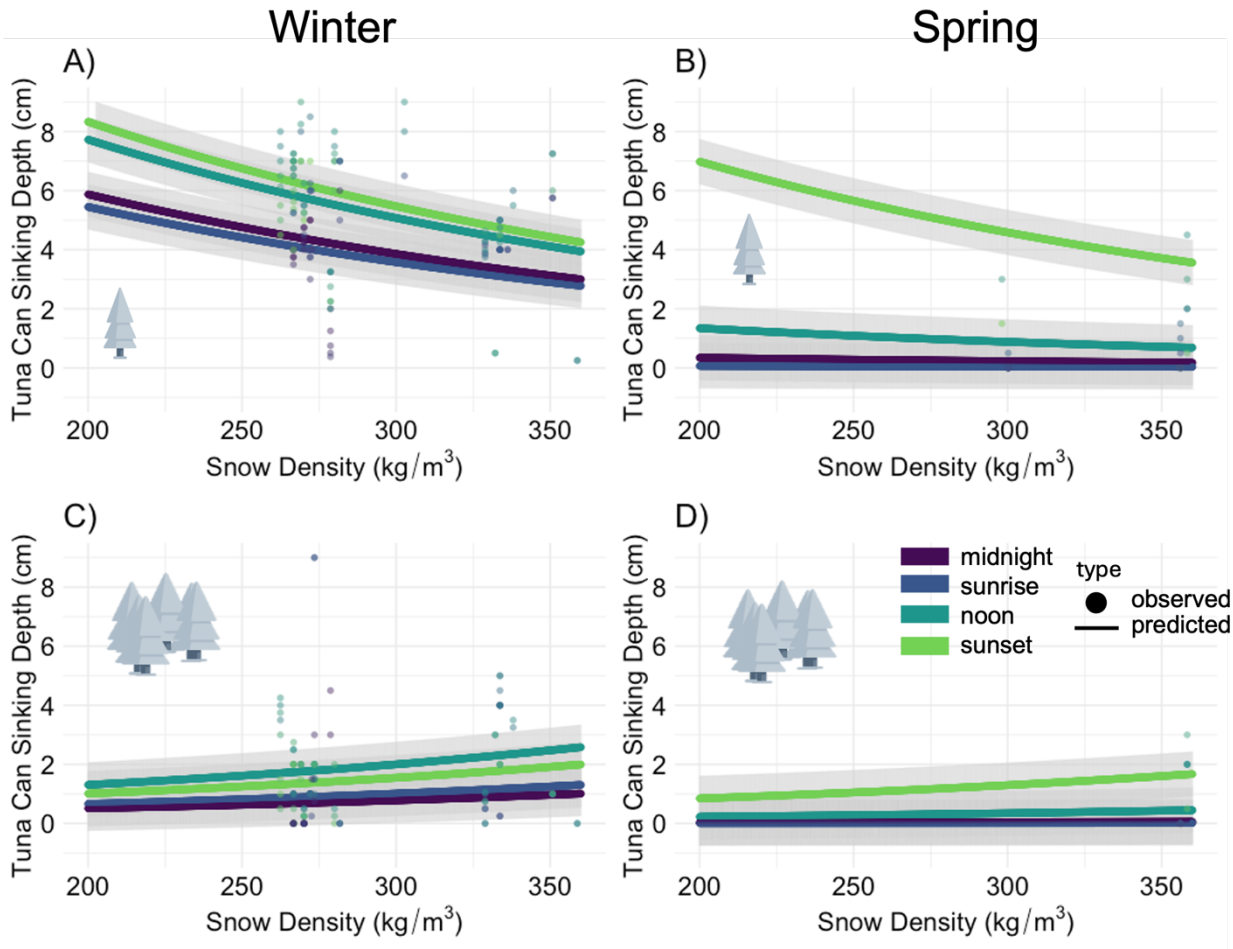


Figure C4. Observations of our tuna can sinking depth with snow density outputs from Norway's daily snow model in the open canopy for A) winter and B) spring and closed canopy for C) winter and D) spring. In all panels, lower sinking depth values correspond to harder snow, and higher values correspond to softer snow. Points represent observations, lines represent predicted values with corresponding 95% confidence intervals from the top model predicting snow hardness when sun time is held constant at either midnight, sunrise, noon, or sunset.

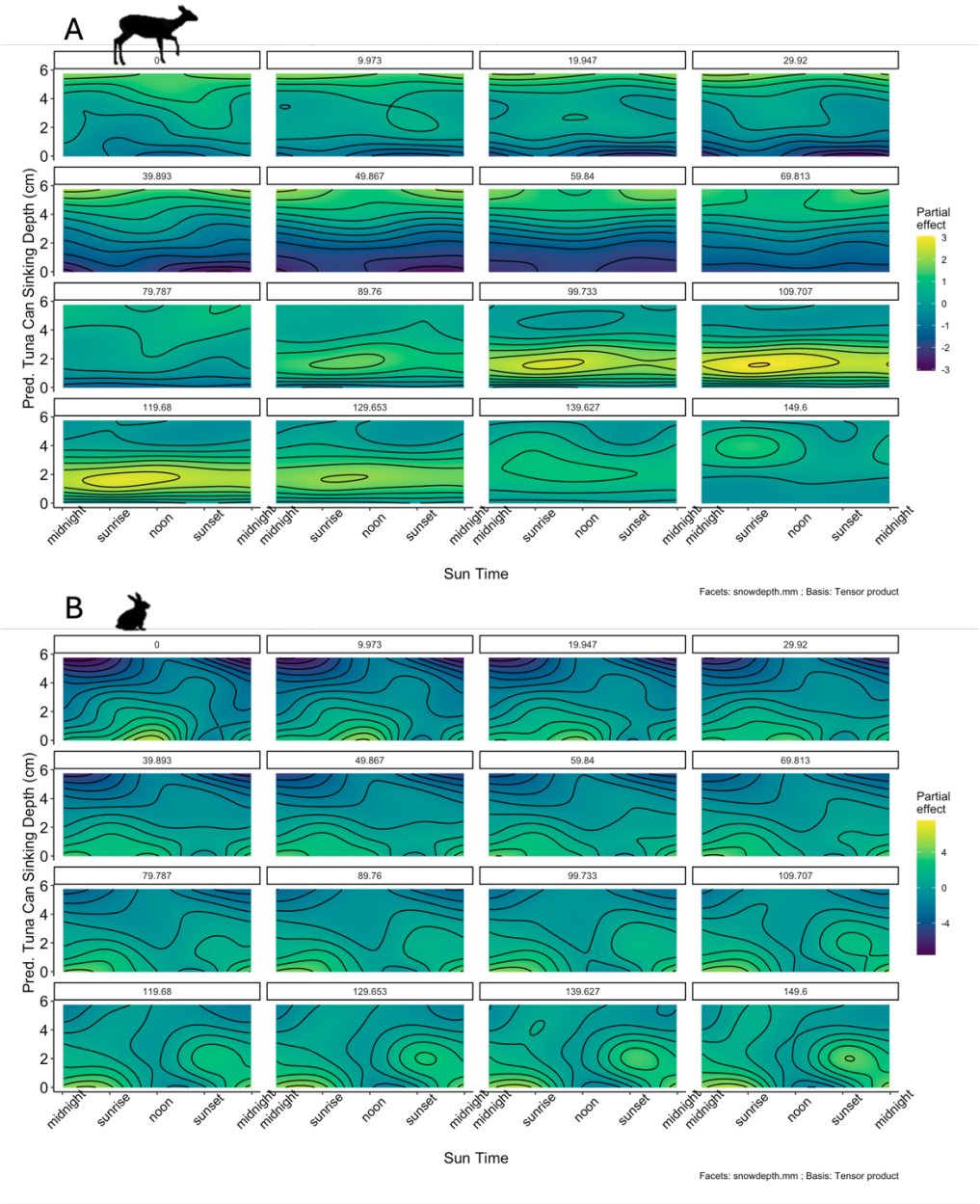


Figure C5. Partial effects from the interaction terms for both the roe deer (top row) and hare (bottom) GAM models, which regressed animal occurrence with sun time, snow depth, snow hardness, an interaction for snow hardness, sun time, and snow depth. Yellow values are associated with higher occurrence and purple values are associated with lower occurrence. Predicted tuna can sinking depth was derived from our snow hardness model, with higher values associated with softer snow.

Table C6. Statistics from the generalized additive models predicting weekly hourly occurrence of roe deer and mountain hares. All variables for both models were found significant ($p < 0.05$). EDF represents effective degrees of freedom.

Species	Variable	EDF	p-value	Deviance Explained
Roe Deer	Sun time	6.30	< 2e-16	12.2%
	Snow Hardness	1.00	0.07	
	Snow Depth	5.72	< 2e-16	
	Sun time * Snow Hardness * Snow Depth	32.78	< 2e-16	
Hare	Sun time	7.37	< 2e-16	13.7%
	Snow Hardness	6.80	< 0.001	
	Snow Depth	6.94	< 0.001	
	Sun time * Snow Hardness * Snow Depth	33.85	< 2e-16	

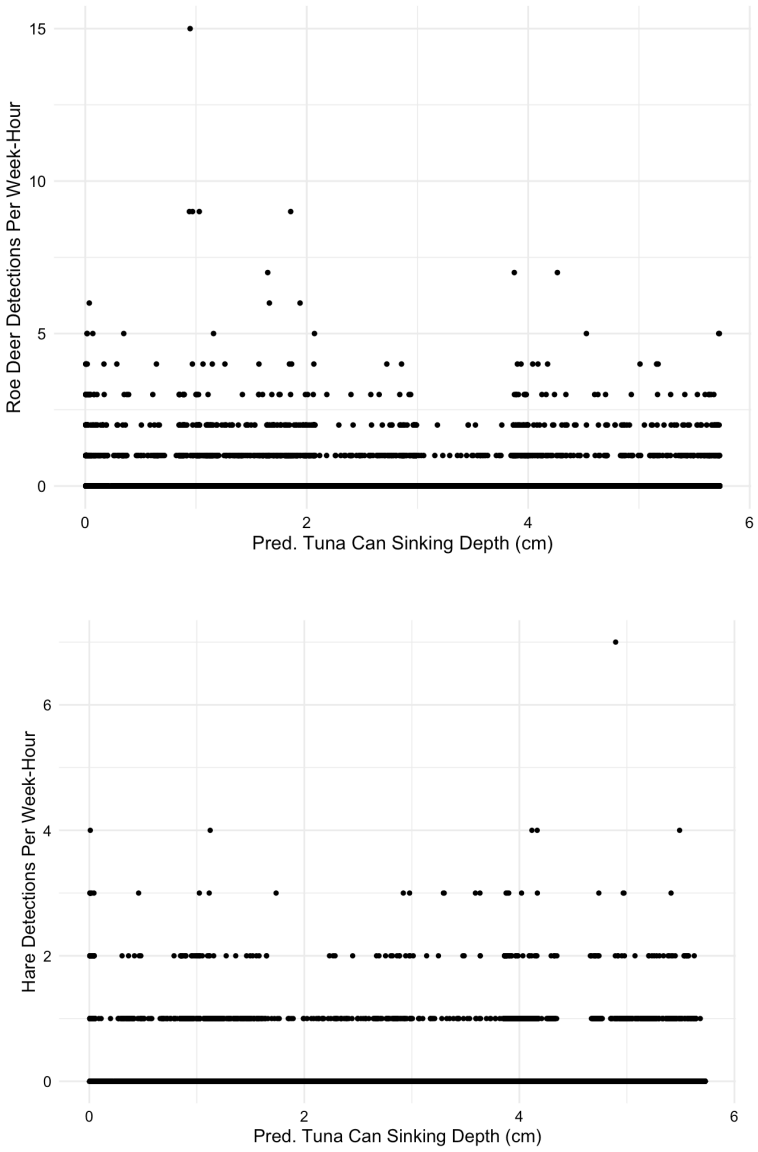


Figure C6. Distribution of A) roe deer detections (n= 417) and non-detections (n = 113,583), and B) hare detections (n = 1012) and non-detections (n = 112,988) for our diel activity GAM model, where each point represents the sum of detections for one of hour of each week for a camera. The y-axis represents the number of roe deer that was detected in each hour for a camera week, and the x-axis represents the corresponding estimated snow hardness value from our snow hardness model. Higher sinking depth values are associated with softer snow. A jitter was applied (height = 0.1) to show data more easily.

APPENDIX D

Supplemental Information for Chapter 5

Rain-on-snow detection from passive microwave remote sensing, weather stations, and wildlife cameras

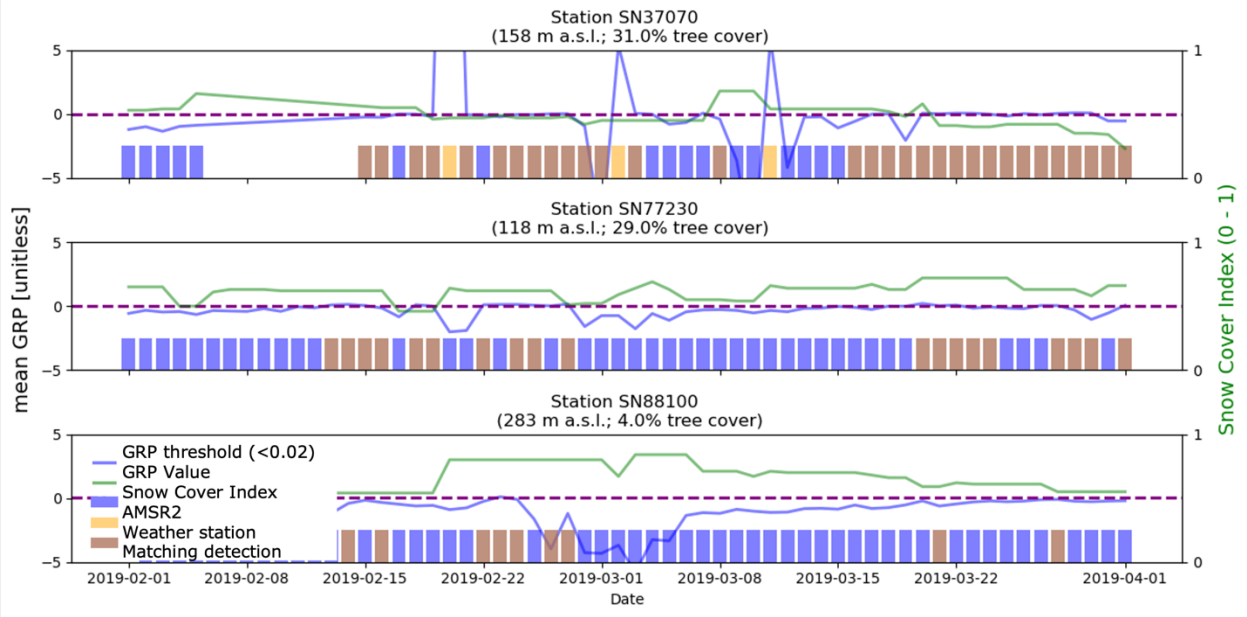


Figure D1. A) ROS time-series for three weather stations using the gradient ratio polarization < 1 from Pan et al. (2018) for 2 months from our study period, February 1, 2019, to April 1, 2019. Stations from top to bottom include, SN38140 (58.3°N 8.5°E; n = 50 events), SN37070 (58.9°N 8.6°E; n = 47 events), and SN77230 (65.8°N 13.2°E; n = 41 events). We focus on 2 months from our study period, February 1, 2019, to April 1, 2019, to facilitate visualization. Blue bars indicate AMSR2 detected using a GRP < 0.02 with a check for dry snow, cream bars indicate a weather station detected a ROS event, and brown indicates both detected an ROS event. The blue line shows the daily GRP value, and the green line shows the snow cover index

derived from the daily MODIS cloud-gap-filled product. The purple dashed line shows the threshold value used, GRP < 0.02, identified using an ROC analysis.

Table D1. Accuracy for various GRP thresholds for identifying ROS from AMSR2 when supplemented using camera image information. Cameras were labeled using a machine learning model trained to detect rain or no rain.

	GRP < 1		GRP < 1 plus dry snow check		GRP < 0.02		GRP < 0.02 plus dry snow check	
	Before	After	Before	After	Before	After	Before	After
Accuracy	31	27	86	87	58	55	88	89
TPR	0.99	0.99	0.08	0.10	0.30	0.30	0.10	0.11
TNR	0.27	0.21	0.92	0.93	0.60	0.56	0.93	0.93

Appendix D2. Spatial Analysis: **Rain on snow event on December 5, 2018**

On December 5, 2018, we identified 19 cameras with ML-labeled detections. Of the cameras, 2 detected a rain-on-snow event, and 17 detected no rain. For the weather stations, no weather stations recorded positive ROS detections, but AMSR2 detected events at XX stations. Upon inspection of the AMSR2 record using a GRP < 0.02 threshold with a dry snow check shows how cameras can further validate and supplement the AMSR2 record, but more work is needed to broaden sample size for days with ROS events (Figure D2).

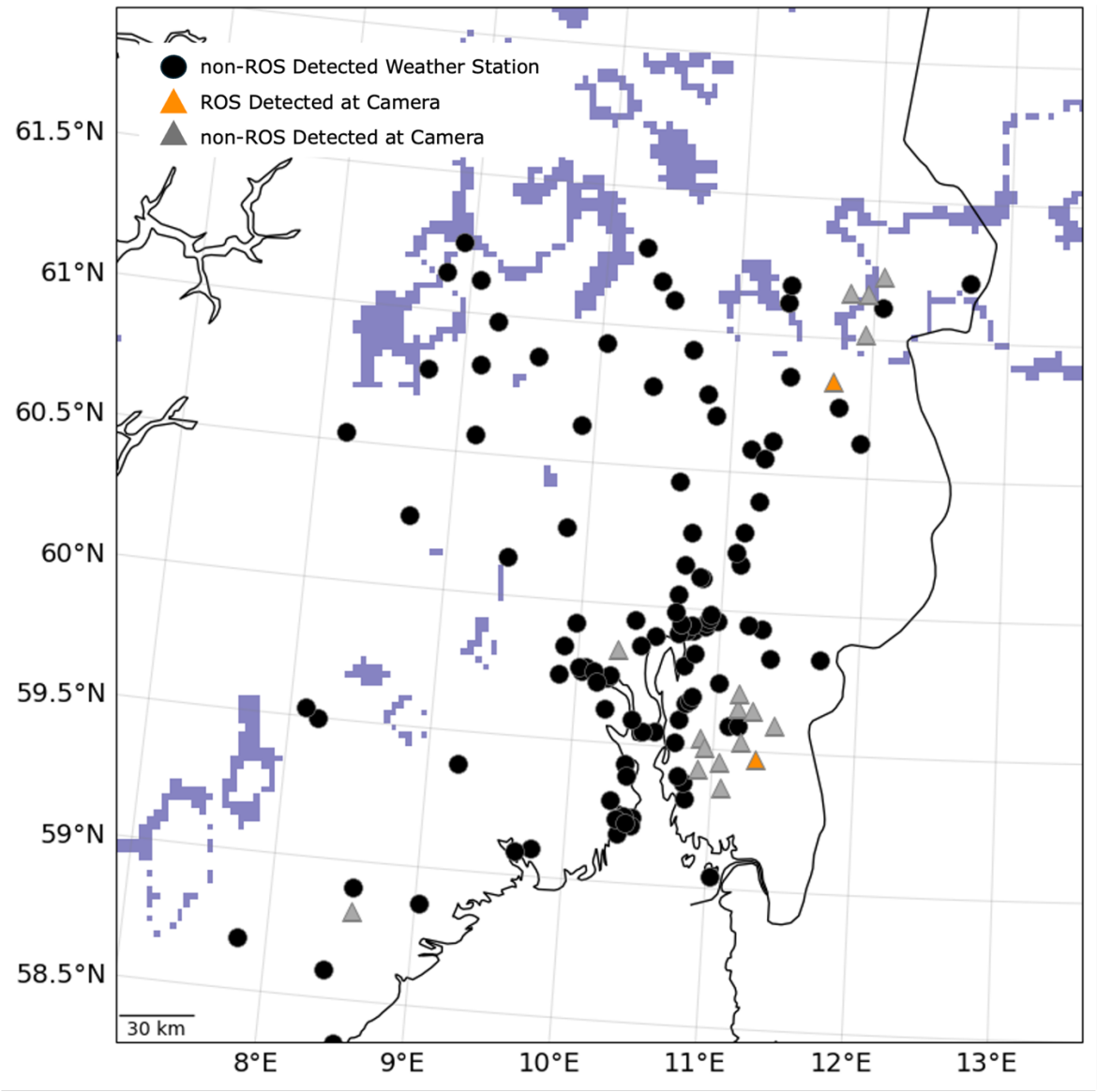


Figure D2. Map of example ROS event on December 5, 2018. Purple indicates the pixel was detected as ROS by AMSR2, white indicates non-detection. No weather stations reported a ROS event. Our ML model detected a ROS event at 2 cameras, and no rain at 17 cameras. When mapped over the region, AMSR likely overestimated the spatial extent, and cameras can provide additional information on AMSR extent.

APPENDIX E

Supplemental Information for Chapter 6

Appendix E1. Additional Library Information

Additional library information for pitch across all classes and specific information for the sleet class:

- Pitch information for each classification (Hz). No weather: 1244.5 ± 847.1; rain: 562.0 ± 577.0; snow: 1523.7 ± 736.8; sleet 1004.2 ± 738.7.
- Sleet information. *n* = 18 samples; avg temp: -0.39 ± 2.83 ºC. Avg. decibels -75.9 ± 9.1 dB.

VITA

Catherine Marina Breen is a graduate researcher in the Prugh Lab at the University of Washington School of Environmental and Forest Sciences. Her research focuses on extracting snow information from wildlife camera trap images to improve satellite observations and wildlife monitoring. Ms. Breen holds a Bachelor of Arts in Ecology and Evolutionary Biology from Princeton University.