UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Statistical Characterization and Development of Snowpack Predictions

Permalink

https://escholarship.org/uc/item/4mv5q07v

Author

Yu, Michelle

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Statistical Characterization and Development of Snowpack Predictions

by

Michelle Yu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Adjunct Professor Christopher Paciorek, Co-chair
Associate Professor Fernando Pérez, Co-chair
Associate Teaching Professor Andrew Bray

Summer 2024

Statistical Characterization and Development of Snowpack Predictions

Copyright 2024
by
Michelle Yu

Abstract

Statistical Characterization and Development of Snowpack Predictions

by

Michelle Yu

Doctor of Philosophy in Statistics

University of California, Berkeley

Adjunct Professor Christopher Paciorek, Co-chair

Associate Professor Fernando Pérez, Co-chair

Accurate high-resolution spatiotemporal data on environmental variables is critical for informing natural resource management and guiding climate change mitigation and adaptation practices. From a hydrological perspective, mountain snowpack is a primary source of freshwater for meeting societal drinking and agricultural needs in various regions. As a result of climate change, these water resources are becoming increasingly vulnerable and declining at a steady pace. To support informed decision-making around water resource planning, accurate data on regular fine-scale spatial and temporal intervals is crucial. Several such gridded data products on key snowpack properties such as snow water equivalent (SWE) and snow depth (SD) have been introduced and provide great utility to the scientific community and practitioners in the environmental space. However, errors and uncertainties in these gridded snow products are not well understood and assumptions underlying the generation of these products are inadequately investigated. In this dissertation, questions around uncertainty and representativeness of gridded snow data are examined and quantified, and a novel statistical approach to estimating SWE is introduced.

The first chapter introduces key terms, concepts, and datasets used throughout the dissertation. The second chapter quantifies error and uncertainty underlying a widely-used gridded SWE product. The third chapter discusses limitations in this gridded product and proposes an alternative framework for empirical SWE prediction that is intuitive, scalable, and statistically sound. Finally, the fourth chapter explores questions around representativeness of point measurements and gridded estimates of true SD and SWE, addressing concerns associated with standard point-to-grid comparisons commonly used in evaluation methodologies.

Contents

Contents i

List of Figures iii

List of Tables x

1 Introduction 1
- 1.1 Snow Water Equivalent (SWE) . . . 1
- 1.2 Annual Cycle of SWE . . . 3
- 1.3 Datasets . . . 4
- 1.4 Study Area & Period . . . 9
- 1.5 University of Arizona Gridded SWE Product (UA SWE) . . . 11
- 1.6 Reproducibility . . . 12
- 1.7 Overview . . . 25

2 Quantifying Variability and Error of a Gridded SWE Product 26
- 2.1 Introduction . . . 26
- 2.2 Datasets . . . 26
- 2.3 Methods . . . 27
- 2.4 Results . . . 37
- 2.5 Conclusion . . . 59

3 Advancing SWE Prediction through Statistical Modeling 61
- 3.1 Introduction . . . 61
- 3.2 Datasets . . . 62
- 3.3 Methods . . . 64
- 3.4 Results . . . 75
- 3.5 Conclusion . . . 99

4 Consideration of Spatial Scales and Representativeness in Assessing Gridded Snowpack Estimates 101
- 4.1 Introduction . . . 101
- 4.2 Datasets . . . 102

4.3 Assessing Gridded SWE against Point Measurements . . . 107
4.4 Assessing Gridded SWE against Gridded Observations . . . 119
4.5 Discrepancy Between Gridded SWE and True SWE . . . 128
4.6 Conclusion . . . 139

Bibliography **141**

A Supplementary Material for Chapter 2 **147**
A.1 Sensitivity Analysis . . . 147

B Supplementary Material for Chapter 3 **150**
B.1 Conditions for Daily Ablation and Daily SWE Models . . . 150
B.2 Gridded SWE . . . 151

List of Figures

1.1 Typical SWE cycle over a water year [74, 60, 64], exemplified by measurement data from a single SNowpack TELemetry (SNOTEL) station for a particular water year (Colorado Site 737 for WY 2016). 4
1.2 Geographical and topographic distribution of the UCRB hydrologic unit with overlaid Colorado ASO basins (dark purple) and SNOTEL sites (red dots) for reference. 10
1.3 SNOTEL SWE compared to UA SWE and our reconstruction of UA SWE, averaged across sites and years for each water year day (October 1st through September 30th). 14
1.4 SNOTEL SWE compared to SNOTEL net snowfall and PRISM net snowfall at SNOTEL sites, averaged across sites and years on each day. 15
1.5 SNOTEL and PRISM accumulated snow and cumulative snow ablation at SNOTEL sites, averaged across sites and years on each day. 16
1.6 SNOTEL and PRISM precipitation at SNOTEL sites, averaged across sites and years on each day. 16
1.7 SNOTEL and PRISM temperature at SNOTEL sites, averaged across sites and years on each day. 17
1.8 Climatological mean of SNOTEL SWE compared to that of UA SWE and UA SWE reconstruction, including decomposition of UA SWE components. Note that SNOTEL and PRISM net snowfall decrease after peak SWE due to the zero-resetting approach used to handle snow ablation exceeding snow accumulation (i.e., Chapter 2 approach). Values on dates prior to June 30 are of interest. . . . 18
1.9 Climatological standard deviation of SNOTEL SWE compared to that of UA SWE and UA SWE reconstruction, including decomposition of UA SWE components. Values on dates prior to June 30 are of interest. 19
1.10 Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences positive change (accumulation). Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month. 20

1.11 Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences no change. Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month. 21
1.12 Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences negative change (ablation). Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month. 22
1.13 Total magnitude of differences in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) for each category of UA SWE behavior (accumulation, no change, ablation), across all sites and years for each month. 23

2.1 Jennings [29] dynamic rain-snow partitioning temperature threshold estimates, clipped to the UCRB region. 30
2.2 Aggregated site-specific ANOVAs of fractional and total variance explained by each factor on each day at SNOTEL sites. 38
2.3 Aggregated site-specific ANOVAs of fractional and total variance explained by each factor on each day at SNOTEL sites. Temporal and residual effects are excluded to highlight main effects of interest. 40
2.4 Percentage of sites and years that precipitation is partitioned differently by the default 2°C threshold and the dynamic Jennings threshold. 42
2.5 Distribution of grid cells in the UCRB corresponding to SNOTEL sites (dots or squares) for an arbitrary day (April 1, 2000). 43
2.6 Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at grid cells containing SNOTEL sites. 44
2.7 Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at grid cells corresponding to SNOTEL sites. Temporal and residual effects are excluded for clarity. 46
2.8 Distribution of snow covered grid cells in the UCRB, as defined by the SSM, for an arbitrary day (April 1, 2000). Snow covered grid cells are indicated in yellow. Red dots indicate SNOTEL sites. 48
2.9 Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at snow covered grid cells. 49
2.10 Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at snow covered grid cells. Temporal and residual effects excluded for clarity. 50

2.11 Error decomposition of UA SWE components, when interpolation is ignored in the algorithm. Solid colored bars indicate the average of the δ_C errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference (SNOTEL SWE in this case). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the error of the reference (SNOTEL SWE), which is 0% by definition. 51
2.12 Error decomposition of alternative UA SWE modeling decisions, when interpolation is ignored in the algorithm. Solid colored bars indicate the average of the δ_2 and δ_3 errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference ("SNOTEL" and "PRISM" for δ_2 and δ_3, respectively). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the mean SNOTEL SWE error (0% by definition) and the grey dashed vertical line marks the mean error of "PRISM" (i.e., UA SWE ignoring interpolation). . . . 53
2.13 Average precipitation across sites and years on each day from the SNOTEL, PRISM, GridMET, and Livneh data sources. 55
2.14 Average temperature across sites and years on each day from the SNOTEL, PRISM, GridMET, and Livneh data sources. 55
2.15 Error decomposition of UA SWE components, when interpolation is included in the algorithm. Solid colored bars indicate the average of the δ_C errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference (SNOTEL SWE and subsequently "krigSNOTEL"). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the error of the SNOTEL SWE reference, which is 0% by definition and the grey dashed vertical line marks the error of the "krigSNOTEL" reference. 56
2.16 Error decomposition of alternative UA SWE modeling decisions, when interpolation is included in the algorithm. Solid colored bars indicate the average of the δ_2, δ_3, and δ_4 errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference ("SNOTEL", "krigSNOTEL", and "PRISM" for δ_2, δ_3, and δ_4, respectively). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the mean SNOTEL SWE error (0% by definition) and the grey dashed vertical lines mark the mean error of "krigSNOTEL" and "PRISM" (i.e., UA SWE). 58

3.1 The UA SWE ablation regression as depicted in Figure 2b of [10]. Each blue dot represents a day at a SNOTEL site located in a 10° × 10° longitude-latitude box in the WUS during the 2004-2007 water years. The green line is the best fit regression line that is used to estimate cumulative ablation on each day for both SNOTEL and PRISM in the net snowfall components in the UA SWE algorithm. 62
3.2 Climatological snow accumulation derived from five different fixed rain-snow partitioning thresholds, from -4°C to +4°C. The snow accumulation quantity on each day is the average snow accumulation across all sites and years in the study area and period. 64
3.3 Overview of study workflow, including model training, evaluation, and application. 65
3.4 Distribution of non-zero differences between precipitation on cold days (snowfall) and change in SWE (dSWE). 67
3.5 Performance of the full daily ablation model trained only with static terrain covariates, compared to the baselines, under the daily ablation evaluation approach. 76
3.6 Performance of the full daily ablation model trained only with granular temperature covariates, compared to the baselines, under the daily ablation evaluation approach. 77
3.7 Performance of the full daily ablation model trained with historical weather condition covariates of differing temporal lags, compared to the baselines, under the daily ablation evaluation approach. 78
3.8 Performance of the full daily ablation model trained with static granular temperaturelags, compared to the baselines, under the daily ablation evaluation approach. 79
3.9 Performance of the full daily ablation model trained with all static terrain, granular temperature, and historical weather (seven day lag) covariates, compared to the baselines, under the daily ablation evaluation approach. 80
3.10 Performance of the full daily ablation model trained with covariates sets including previous day SWE, compared to the baselines, under the daily ablation evaluation approach. 81
3.11 Performance of the full daily ablation model trained with previous day SD in addition to previous day SWE, compared to the baselines, under the daily ablation evaluation approach. 82
3.12 Performance from the cumulative ablation evaluation scheme. The green bars correspond to the UA SWE-like daily ablation (dAbl) model, the orange bars correspond to the physical daily ablation model, and the blue bars correspond to the full daily ablation model. For each daily ablation model type, the solid bar corresponds to the model excluding $\widehat{\text{SWE}}_{t-1}$ from the covariate set, and the textured bar corresponds to the model including $\widehat{\text{SWE}}_{t-1}$ in the covariate set. . 84

3.13 Performance from the SWE evaluation scheme. The blue bars correspond to the full daily ablation model, whereas the red bars correspond to the dSWE model. For each model type, the solid bar corresponds to the model excluding $\widehat{\text{SWE}}_{t-1}$ from the covariate set, and the textured bar corresponds to the model including $\widehat{\text{SWE}}_{t-1}$ in the covariate set. 86
3.14 Climatological mean of observed and predicted SWE across the water year, comparing SNOTEL SWE, SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation (dAblation) and dSWE ML models. 87
3.15 Climatological standard deviation of observed and predicted SWE across the water year, comparing SNOTEL SWE, SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation and dSWE ML models. . . . 88
3.16 Climatological RMSE on each day across the water year for the best performing full daily ablation and dSWE ML models, compared to SNOTEL net snowfall and bias-corrected SNOTEL net snowfall. 89
3.17 Distribution of site RMSE (against SNOTEL SWE) of SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation and dSWE ML models. 90
3.18 Climatological mean of observed and predicted SWE across the water year at grid cells containing the SNOTEL sites, comparing SNOTEL SWE, UA SWE, and our full daily ablation and dSWE ML models when PRISM meteorological inputs are used. 91
3.19 Climatological standard deviation of observed and predicted SWE across the water year at grid cells containing the SNOTEL sites, comparing SNOTEL SWE, UA SWE, and our full daily ablation and dSWE ML models when PRISM meteorological inputs are used. 92
3.20 Taylor diagram. 93
3.21 Gridded SWE prediction from the full daily ablation RF model compared to UA SWE on April 1. 95
3.22 Differences in the mean and standard deviation of our full daily ablation RF model compared to UA SWE (i.e., UA SWE - full daily ablation RF SWE) on April 1, by 4 km $\times$ 4 km grid. 96
3.23 Difference in mean of our full daily ablation RF model versus UA SWE by 4 km $\times$ 4 km grid for each month. 97
3.24 Comparison of ECDFs of climatological April 1 means and standard deviations of SWE predicted by UA SWE compared to the full daily ablation RF model at the grid level over the UCRB. 98

4.1 Taylor River basin on April 21, 2022. (a) 3 m ASO SD; (b) 4 km Gridded ASO SD; (c) SD from 4 km UA SWE. 104
4.2 Taylor River basin on May 25, 2022. (a) 3 m ASO SD; (b) 4 km Gridded ASO SD; (c) SD from 4 km UA SWE. 105

4.3 Taylor River basin on April 21, 2022. (a) 50 m ASO SWE; (b) 4 km Gridded ASO SWE; (c) SWE from 4 km UA SWE; (d) SWE from 4 km MY SWE. . . . 106
4.4 Taylor River basin on May 25, 2022. (a) 50 m ASO SWE; (b) 4 km Gridded ASO SWE; (c) SWE from 4 km UA SWE; (d) SWE from 4 km MY SWE. 106
4.5 Distribution of elevation and slope quantities of SNOTEL sites in the UCRB and all 30 m grid cells in the UCRB. 107
4.6 Distribution of elevation and slope quantities of SNOTEL sites in the UCRB and all 4 km UCRB grid cells containing SNOTEL sites. 108
4.7 Distribution of elevation and slope quantities of SNOTEL sites within Colorado ASO basins and all 30 m pixels in the Colorado ASO basins. 109
4.8 Scatterplots of SD from SNOTEL compared to ASO (top row), Gridded ASO (middle row), and UA SWE (bottom row) during the peak SWE period (left column) and the ablation period (right column). 111
4.9 Example of regions of ASO basins overlaid by a selected SNOTEL site on dates in the peak SWE period, when ASO and SNOTEL at the point-level differ significantly. 112
4.10 Example of a region of an ASO basin (East River Basin) overlaid by a selected SNOTEL site on a date in the ablation period. (a) The 10 × 10 ASO pixels surrounding the SNOTEL site, and (b) the 4 km × 4 km grid cell containing the SNOTEL site. Generally, there is less heterogeneity in ASO pixels surrounding SNOTEL sites, especially during the ablation season. 114
4.11 Scatterplots of SWE from SNOTEL compared to ASO (top row), Gridded ASO (second row), UA SWE (third row), MY SWE with gridded PRISM meteorological inputs (fourth row), and MY SWE with point-level SNOTEL meteorological inputs (bottom row) during the peak SWE period (left column) and the ablation period (right column). 115
4.12 Comparison of SNOTEL and PRISM meteorology at SNOTEL sites within ASO basins. Each dot represents a monthly average over the study period for a SNOTEL site. 118
4.13 Scatterplots of Gridded ASO SWE compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, arranged in ascending order from top to bottom rows. Each dot in each panel corresponds to a grid cell in an ASO basin on a flight date. 121
4.14 Scatterplots of Gridded ASO SWE compared to UA SWE (left column) and MY SWE (right column) stratified by month. Each dot in each panel corresponds to a grid cell in an ASO basin on a flight date. 123
4.15 Total water volume of each basin-date according to Gridded ASO SWE versus UA SWE and MY SWE. Each dot in the scatterplots correspond to a basin-date at which an ASO flight took place. The black dots correspond to basin-dates in the peak SWE season and the red dots correspond to basin-dates falling in the ablation season. 126
4.16 Comparison of variograms of Gridded ASO SWE, UA SWE, and MY SWE, computed from grid cells aggregated across all basins. 127

4.17 Histogram and boxplot of differences between predictions from MY SWE with 4 km meteorological input data and MY SWE with 800 m meteorological input data. 129
4.18 Scatterplots comparing Gridded ASO SWE to UA SWE (top row), MY SWE with 4 km PRISM meteorological input data (middle row), and MY SWE with 800 m PRISM meteorological input data (bottom row). Each dot in the scatterplots represents a single grid cell in a basin-date between 2016 and 2019 when proprietary 800 m PRISM data are available. 131
4.19 Histogram and boxplot of the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside bands around the 4 km cell-wise average elevation (Average Elevation Filtered Gridded ASO SWE). 133
4.20 Scatterplots showing the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside 4 km cell-wise average elevation bands (Average Elevation Filtered Gridded ASO SWE) compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, stratified by month. 134
4.21 Histogram and boxplot of the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside the range of elevation and slope values represented by SNOTEL sites in the UCRB (Attribute Filtered Gridded ASO SWE). 136
4.22 Scatterplots showing the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside the range of elevation and slope values represented by SNOTEL sites in the UCRB (Attribute Filtered Gridded ASO SWE) compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, stratified by month. 137
4.23 Comparison of variograms of Gridded ASO SWE, UA SWE, MY SWE, Average Filtered Gridded ASO SWE, and Attribute Filtered Gridded ASO SWE computed from grid cells aggregated across all basins. 139

B.1 Gridded SWE prediction from the dSWE NN model compared to UA SWE on April 1. 151
B.2 Differences in mean and standard deviation of our dSWE NN model versus UA SWE by 4km x 4km grid. 152
B.3 Difference in mean of our dSWE NN model versus UA SWE by 4km x 4km grid for each month. 153
B.4 Comparison of ECDFs of climatological April 1 means and standard deviations of SWE predicted by UA SWE compared to the dSWE NN model at the grid level over the UCRB. 154

List of Tables

2.1 The four key drivers of UA SWE and their associated datasets or statistical approaches considered in our sensitivity analysis and error decomposition. *indicates the default dataset or approach employed by the UA SWE algorithm. 28
2.2 Estimates corresponding to each factor in our ANOVA setup. 32

3.1 Full list of covariates considered in daily ablation and daily SWE modeling. Variables for which site-level data exists, the corresponding data source at the grid scale, if available, is indicated in brackets. 70

4.1 List of basins in or adjacent to the UCRB surveyed by ASO. The associated flight dates and approximate basin area are provided. 103

B.1 Physical conditions of days used for training and prediction for each daily ablation (dAbl) and dSWE model. This is the tabular representation of the methodology discussed in Section 3.3. 150

Acknowledgments

I would like to express my heartfelt gratitude to my advisors, Chris Paciorek and Fernando Pérez, for their support and guidance over the years. Chris has generously devoted his time and energy to engaging in in-depth discussions with me while also creating a space for me to work and grow independently. I am equally appreciative of Fernando for fostering a welcoming environment and for helping me navigate various aspects of my research journey.

I have had the pleasure of working with Alan Rhoades and Mark Risser for several years. Their expertise and invaluable input have been instrumental in developing and advancing my research, and their enthusiasm has made our collaboration truly enjoyable. I would also like to thank Andrew Bray, Giles Hooker, Manuela Girotto, John Chiang, and the Pérez Group for their insightful suggestions and contributions.

I am extremely grateful to La Shana Porlaris, Ryan Lovett, and the entire UC Berkeley Statistics staff for everything they have done to make my graduate school experience as seamless as possible. Thank you for your dedication, care, and positive spirit.

I am fortunate to have encountered many teachers and mentors throughout my life who inspired me and believed in my potential. I would like to extend a special thanks to Cy Michael Ogle, David Damiano, Eric Ruggieri, Shannon Stock, Cristina Ballantine, Steven Levandosky, Bruce Bunke, Ken Prestwich, and Steve Chung, who helped shape my academic path and whose influence was crucial in my decisions to pursue mathematics and statistics. Without them, I would not be where I am today.

To my loved ones, thank you for your unwavering support, patience, and encouragement throughout this journey. From long conversations and thoughtful advice to celebrating my wins and sharing in my losses, your presence has been a constant source of strength. Your personal sacrifices and the countless ways you have stood by me have made all the difference.

This work was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, the National Science Foundation Earth Cube Program under awards 1928406, 1928374, and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1752814. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. I am also fortunate to have received support from the Environment and Society: Data Sciences for the 21st Century National Science Foundation Research Traineeship and Hearts to Humanity Eternal.

Computations for this research were performed on the National Energy Research Scientific Computing Center Cori and Perlmutter supercomputers as well as on UC Berkeley Statistical Computing Facility machines.

Chapter 1

Introduction

The dynamic intersection of statistics and the Earth sciences plays a crucial role in understanding and addressing complex challenges facing our planet today. As the global community confronts issues such as climate change, biodiversity loss, and environmental degradation, there is an increasing need for rigorous statistical methods to analyze intricate datasets, model physical processes, and inform evidence-based decision-making.

In recent decades, advancements in areas such as remote sensing technology, computational capabilities, and open data initiatives have fueled rapid growth in the volume and complexity of data in Earth science communities, e.g. [68, 43, 84]. These datasets, often characterized by spatial and temporal variability, present unique challenges for traditional statistical methods, e.g. [61, 22, 16, 32]. Innovative statistical approaches tailored to the characteristics of environmental data enable researchers to uncover meaningful patterns, quantify uncertainties, and develop predictive models to inform policies and conservation strategies.

This dissertation explores the application of statistical techniques within the field of hydrology, specifically addressing uncertainty, modeling, and fidelity of spatiotemporal data related to water resources derived from snow. It critically examines how errors propagate through models and improves the accuracy of representing and assessing complex hydrological processes through a statistical lens. Overall, it aims to enhance understanding and management of this critical natural resource.

1.1 Snow Water Equivalent (SWE)

Snow is integral to the Earth's climate system, influencing everything from regional water resources to global climate patterns. Snowpack, in particular, is a vital component of the hydrological cycle, serving as a natural reservoir that stores water during water and releases it gradually during the melt season in the spring and summer [69, 64]. The seasonal accumulation and melting of mountain snowpack significantly influences surface hydrology in many regions, making it an important water resource. Its impacts extend from local

to global scales, affecting water supply, hydropower, ecological processes, weather patterns, recreational activities, and global climate dynamics [75]. As a result, accurate measurement of snowpack properties is essential for understanding and effectively managing this critical resource.

Seasonal snowpack is characterized by snow depth (SD) and bulk density [48]. The product of these two quantities yields snow water equivalent (SWE), which is the amount of liquid water in a snowpack that would be formed if the snowpack were completely melted. SWE is a fundamental metric for quantifying snowpack and assessing its impact on water resources, carrying significant scientific, environmental, and socioeconomic implications [75].

Snowmelt from mountain snowpack replenishes rivers, streams, and groundwater, sustaining flow that supports aquatic ecosystems, preserves wildlife habitats, and fosters biodiversity [66]. SWE also aids in mitigating drought impacts in snow-dependent regions by ensuring consistent water supply and maintaining ecological health [67].

Economically, SWE is invaluable for sectors reliant on water availability [70]. Agriculture, for instance, heavily depends on predictable water supply for irrigation, directly influencing crop yields and food production. Understanding SWE helps farmers and agricultural planners anticipate water availability and manage planting and harvesting schedules effectively. Similarly, hydropower generation relies on accurate SWE assessments to optimize reservoir operations and predict energy production from snowmelt-driven hydroelectric plants. The economic significance of snowpack is substantial, contributing several hundred billion dollars in annual benefit in the Western United States (WUS) [12, 13].

Furthermore, SWE informs natural disaster planning, particularly in flood forecasting and reservoir management. Monitoring changes in snowpack depth and density allows for early assessment of flood risks associated with rapid snowmelt and heavy rainfall events [13]. This enhances public safety and minimizes economic losses by facilitating effective flood response and infrastructure planning.

In scientific research, SWE serves as a key parameter in snow model and climate model evaluation, which is essential for monitoring and detecting climate change impacts [9, 50].

Observational and model data have both shown that anthropogenic climate change has reshaped global and WUS water resources, including a declining snowpack, with model projections suggesting further mountain snowpack decline [64]. Rising temperatures are causing shifts in precipitation patterns, leading to more rain than snow in some regions and altering the timing, extent, and intensity of snowpack accumulation and melt [77]. This results in reduced SWE volumes and durations, affecting freshwater availability crucial for agriculture, hydropower generation, and urban water systems [35]. Accelerated snowmelt rates contribute to earlier and more rapid spring runoff in mountainous regions, disrupting ecosystems, increasing water scarcity, and posing challenges for infrastructure resilience and disaster preparedness [2].

Given the societal significance of snowpack as a natural resource and the far-reaching implications of its evolution due to climate change, accurate data on SWE, along with a comprehensive understanding of underlying uncertainties, is critically needed [19].

1.2 Annual Cycle of SWE

The water year (WY) is a standardized 12-month period between October 1 and September, used in hydrology to analyze water-related data. It encompasses the full annual cycle of SWE, which is characterized by a snow accumulation period and a snow ablation period. Snow accumulation is the amount of snow that falls and accumulates on the ground, while snow ablation is the reduction of the snowpack through processes such as snow melt and sublimation [81]. The snow accumulation period typically occurs from fall through spring, while the snow ablation period generally takes place from late spring through summer. These periods can be characterized by key dates such as the snowpack accumulation start date (SAD), peak SWE date (SPD), and the complete snowmelt date (CMD) [74, 60], which are defined as follows:

- **Snowpack Accumulation Start Date (SAD)**: First day when SWE $\geq$ 10% of maximum SWE, considered the beginning of the accumulation season.
- **Peak SWE (SPD)**: Day of maximum SWE.
- **March 1 SWE** and/or **April 1 SWE**: At times used as proxies for peak SWE as these are standardized dates facilitating easier comparison of peak SWE across years and locations. Although somewhat arbitrary, April 1 originates from monthly manual snow surveys initiated in the early twentieth century and has been formalized in water management as an indicator of streamflow during the warm [64].
- **Complete Melt Date (CMD)**: First day when SWE $\leq$ 10% of maximum SWE, considered the end of the melt season.

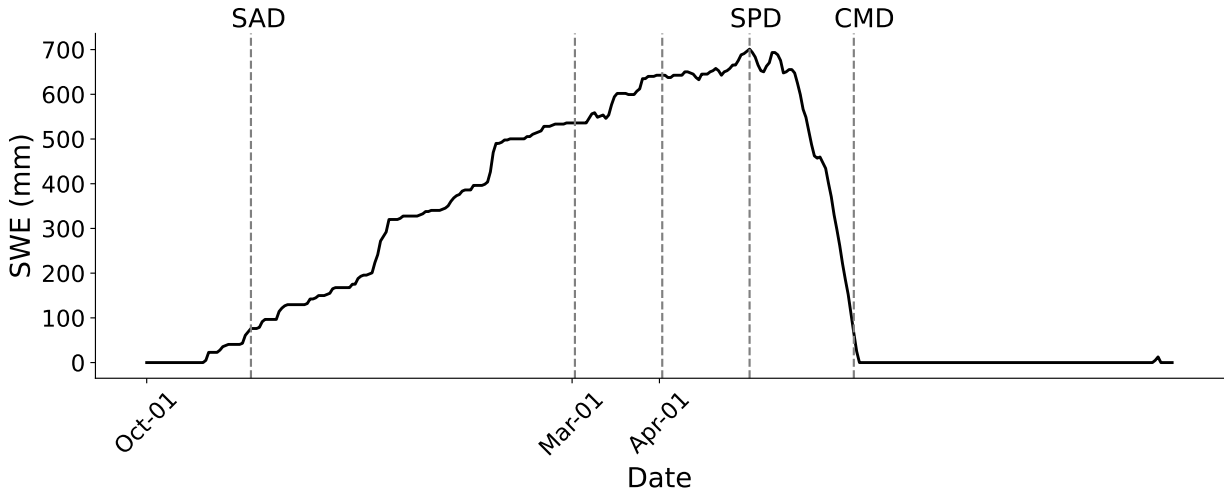


Figure 1.1: Typical SWE cycle over a water year [74, 60, 64], exemplified by measurement data from a single SNowpack TELemetry (SNOTEL) station for a particular water year (Colorado Site 737 for WY 2016).

1.3 Datasets

To provide clarity and context before delving into specific datasets, we define and distinguish several key terms. "Data" refers to measurements or in-situ observations collected directly from sources such as meteorological stations, satellites, sensors, and other instruments. "Gridded products" or "gridded data" are structured collections of values organized in grid formats typically at regular spatial and temporal intervals. These gridded values are derived from raw data through methods such as interpolation, statistical modeling, or physical models. Gridded products provide spatially continuous representations of climate variables or other environmental parameters over specific geographical areas and time periods.

In this section, we provide in-depth descriptions of the key datasets used in this dissertation. First, we discuss observational data derived from in-situ measurements and remote sensing. These datasets serve as groundtruth for point-level SWE and SD validation in subsequent chapters. Following this, we describe gridded SWE data products, with a specific focus on the University of Arizona SWE (UA SWE) product that is foundational to our studies.

Observational Data

SNOwpack TELemetry (SNOTEL) Network

The SNOwpack TELemetry (SNOTEL) network is a system of automated snowpack and weather monitoring stations primarily operated by the United States Department of Agriculture's Natural Resources Conservation Service (NRCS) [52]. These stations are located in remote mountainous regions across the WUS, where snowmelt serves as a major source of water supply for downstream communities and ecosystems. They primarily aim to monitor and assess snowpack conditions in real-time.

SNOTEL generates daily (and in some instances, hourly or subdaily) data on SD, SWE, temperature, precipitation, and other meteorological variables. SWE is measured using snow pillows, which are large bladders filled with a mixture of antifreeze and water, installed beneath the snowpack. As snow accumulates on top of the pillow, it compresses the bladder, and the pressure change is recorded by sensors. This pressure change correlates with the weight of the overlying snowpack. Snow depth data is collected by snow depth sensors that determine the distance from the sensor to the snow surface. Temperature and precipitation are respectively measured using temperature sensors in weatherproof enclosures and precipitation gauges that detect weight changes from accumulated liquid.

Airborne Snow Observatory (ASO)

The Airborne Snow Observatory (ASO) dataset, originally developed by NASA's Jet Propulsion Laboratory (JPL) and the California Department of Water Resources and now available through ASO, Inc., is a remotely sensed dataset encompassing SD, SWE, and other snow-related parameters over large mountainous regions. Since 2013, ASO has been collecting data across basins in parts of California, Colorado, Oregon, Utah, Washington, and Wyoming during the peak SWE period (February-April) into the early summer period (May-July). For each basin, ASO provides snow depth maps at 3 m and 50 m resolution and SWE maps at 50 m resolution on certain flight dates.

The data collection and processing procedure underlying ASO are described in detail in [18] and [55]. Here, we provide an overview of key steps. The ASO data collection methodology involves using an airborne LiDAR (Light Detection and Ranging) system combined with an imaging spectrometer to accurately measure snow depth and characterize snow properties. Specifically, LiDAR sensors mounted on an aircraft emit laser pulses towards the ground. These pulses bounce back off the snow surface, providing precise measurements of the distance from the aircraft to the snow surface. Snow depth is determined by comparing LiDAR-derived surface elevations on snow-off and snow-on days. On the other hand, an imaging spectrometer captures reflected light from the snow surface across various wavelengths, characterizing snow properties, such as albedo and grain size. The SD maps derived from LiDAR undergo various data processing and calibration steps that utilize information derived from the image spectrometer. These steps include handling pixels with

missing data, distinguishing between vegetated and non-vegetated (bare) areas, and merging separate gridded surfaces. The result is composite high-resolution (3 m) maps of snow depth.

After calibration, validation, and processing, the 3 m resolution snow depth maps are aggregated to obtain 50 m resolution snow depth maps. The snow depth measurements are then combined with albedo estimates to model SWE (via the iSNOBAL model), producing 50 m resolution SWE maps.

Examples of Other Observational Datasets

Besides SNOTEL, another standard in-situ SWE dataset is snow course data, a manual snow monitoring initiative across the WUS managed by the National Water and Climate Center (NWCC). Snow courses collect monthly measurements at fixed locations along transects, where surveyors measure snow depth and water content during winter. Typically 1,000 feet long, these sites consist of 5 to 10 sample points evenly spaced apart. Surveyors use snow samplers, often referred to as "snow tubes", to remove a core of snow at each sample point. The core is then weighed to determine its water content. The reported measurement is the average depth and water content from the 5-10 sample points.

Another widely used observational SWE dataset is the one operated by the National Oceanic and Atmospheric Administration (NOAA) National Operational Hydrologic Remote Sensing Center (NOHRSC). NOHRSC integrates a combination of ground-based measurements, satellite remote sensing data, and modeling outputs to estimate SWE at a high spatial resolution. Specifically, it utilizes the Snow Data Assimilation System (SNODAS) to assimilate observational data such as snow course measurements, SNOTEL data, and satellite-derived snow cover extent into a physically-based land surface model.

A valuable citizen science initiative in snow research is the Community Snow Observations (CSO) dataset, engaging individuals, citizen scientists, and community groups in collecting and sharing snowpack data. Participants measure SD during personal excursions using tools like snow probes. These SD measurements are integrated with other physical data to simulate SWE through snow modeling.

Gridded Products

University of Arizona SWE (UA SWE)

The University of Arizona SWE product (UA SWE) is a 4 km $\times$ 4 km SWE and SD gridded dataset spanning the U.S., at present available for the 1982 to 2022 water years. The methodology is fully empirical, integrating ground measurement data (including SNOTEL) and model data (PRISM gridded climate data) to estimate SWE and applying further numerical modeling (snow density parameterization) to estimate SD [10, 11, 83].

Details on the mathematical formulation of UA SWE are available in Section 1.5. Below, we summarize the algorithm qualitatively, highlighting the main components of the algorithm.

$$\text{UA SWE} = \phi\left(\frac{\text{SNOTEL SWE}}{\text{SNOTEL net snowfall}}\right) \cdot \text{PRISM net snowfall} \tag{1.1}$$

where ϕ is interpolation (specifically, an ordinary kriging approach with no nugget using a spherical variogram) and net snowfall is defined for both SNOTEL and PRISM as:

$$\text{net snowfall} = \text{accumulated snow} - \text{cumulative snow ablation}$$

Accumulated snow is estimated by summing observed snowfall determined based on partitioning precipitation using a 2°C rain-snow partitioning threshold. In other words, on each day, precipitation is considered snowfall if temperature falls below 2°C. Otherwise, precipitation is considered rainfall.

Cumulative snow ablation is estimated by a quadratic regression on cumulative degree days exceeding 0°C. Degree days above 0°C on each day is the total number of degrees by which mean daily temperature exceeds 0°C for that day.

While several other established gridded SWE products exist, this dissertation primarily employs the UA SWE product. The decision to focus on UA SWE is three-fold. First, UA SWE produces deterministic estimates of SWE that involves assumptions leading to uncertainties that could be challenged. Second, UA SWE is one of the most commonly used gridded SWE datasets in snow hydrology research. Third, despite difficulties with fully reconstructing UA SWE from scratch (as described in Section 1.6), it is arguably the most easily reproducible gridded SWE product available publicly.

Examples of Other Gridded Products

Other widely used gridded SWE products in the scientific community include the Livneh and the UCLA WUS SR products. These are not used in analyses in this dissertation, but the generation methodologies are briefly outlined below.

Livneh SWE The Livneh 2015 (Livneh) product is a gridded dataset on precipitation, maximum and minimum temperature, SWE, and other hydrometeorological variables. The dataset is available at a 1/6° (∼6 km) spatial resolution across Mexico, the Conterminous United States (CONUS), and southern Canada and at a daily temporal resolution, spanning 1950 through 2018 [40, 41].

To generate gridded precipitation and temperature data, an interpolation technique (the SYMAP algorithm) is applied to quality-controlled precipitation, minimum temperature, and maximum temperature observational data from the Global Historical Climatology Network (GHCN) to a 1/6° spatial resolution. The gridded temperature data is lapsed with elevation at a constant rate of 6.5°C/km and the gridded precipitation data is scaled on a monthly basis to match the PRISM climatological mean (in the case of CONUS). SWE and other hydrometeorological variables are then simulated using a land surface model forced with the processed gridded precipitation and temperature data in conjunction with reanalysis wind data.

UCLA Western United States Snow Reanalysis (UCLA WUS SR) SWE The UCLA Western United States Daily Snow Reanalysis (UCLA WUS SR), developed by researchers at the University of California, Los Angeles, is a daily 90 m spatial resolution SWE dataset, spanning the WUS from 1985 to 2021 [44, 23].

The dataset merges Landsat fractional snow-covered area (fSCA) datasets, meteorological reanalysis products, and snow modeling using a Bayesian data assimilation framework. Unlike other products whose SWE estimates are fully deterministic, UCLA WUS SR provides probabilistic SWE estimates. The Bayesian reanalysis approach involves two main steps. First, an ensemble guess of SWE is obtained by applying a land surface model to static and dynamic physical input variables. Then, the SWE estimates are iteratively updated (i.e., improved) using fractional snow cover area data. The result is a posterior distribution of SWE estimates.

Limitations of SWE Datasets

Observation-based SWE data are susceptible to measurement errors and spatiotemporal gaps. Inevitably, sensors and other data collection equipment experience mishaps from time-to-time, resulting in missing or erroneous data [50]. By definition, measurement data are taken at single point locations that are intrinsically limited by a lack of confidence in how they capture the landscape mean across coarse grid cells [47]. This is especially problematic in areas of mixed forest vegetation, open areas prone to wind redistribution, and complex topography. Most snow-covered regions fall into at least one of these categories. For example, snow course, SNOTEL, and CSO are dispersed at various irregular spatial locations, with snow course collecting monthly data, CSO on non-consecutive dates, and ASO collecting high spatial resolution data in a select number of watersheds on at most a few days per water year.

To address limitations in observed SWE data, bias correction methods are at times employed. Some examples of bias correction techniques include mathematical formulae to correct measurement biases in certain records as well as homogenization of data from old and new sensors. Gridded products, namely statistical and model-based methods that utilize point-level observations or outputs from atmospheric reanalyses to produce regular spatial and temporal SWE data, are another approach for addressing shortcomings in observational data.

Gridded data products are also limited in a number of ways [50]. Methods that leverage point measurement data, such as UA SWE and Livneh, involve a smoothing or interpolation step in the gridding algorithm. Assumptions underlying smoothing and interpolation procedures add additional uncertainty to the final SWE predictions, at times up to 60% of differences observed in estimates of WUS-wide peak SWE [64]. Moreover, gridded products that depend, at least in part, on gridded input datasets are susceptible to errors in these coarse-scale input datasets. If high-resolution datasets for relevant physical variables, such as precipitation and temperature, were available, then uncertainties produced by possible gridding would be minimized substantially or even eliminated. Other methods, such as Livneh,

utilize a physical model in the gridding procedure. The use of such models often adds uncertainty and error to the final SWE estimates. Furthermore, spatial variability in SWE due to topographic and land cover heterogeneity is not adequately captured by data with relatively coarse spatial resolution, which also complicates the validation of SWE products. Additionally, the coarse resolution, along with the averaging and smoothing steps, makes it challenging for gridded products to capture extreme SWE values.

1.4 Study Area & Period

Our study area is the Upper Colorado River Basin (UCRB) hydrologic unit, which consists of a river network spanning northern Arizona, western Colorado, northern New Mexico, eastern Utah, and southern Wyoming (Figure 1.2). The UCRB is a region of interest because it serves as a critical basin in recharging the Colorado River, largely through snowmelt, and provides water to nearly 40 million people and 5.5 million acres of farmland in the WUS, and drives an over $1 trillion economy, through agricultural, industrial, and recreational use [38, 80].

Climate change has significantly affected the UCRB. Events such as drought, over-apportionment, and continuous urban development have depleted the Colorado River's supplies [78]. Because the effects of these events are dire for water supply in the WUS, the UCRB serves as an interesting and important region for further study and analysis.

The UCRB is also home to several data collection campaigns that have launched in recent years. Specifically, the U.S. Department of Energy (DOE) Surface Atmosphere Integrated Field Laboratory (SAIL) and the Study of Precipitation, the Lower Atmosphere and Surface for Hydrometeorology (SPLASH) campaigns provide high quality state-of-the-art observational temperature and precipitation data at select locations in the UCRB [24, 6]. Additionally, the ASO campaign, providing fine-grain SD and SWE data, operates across certain watersheds around this region [18, 55].

The study period in Chapters 2 and 3 of this dissertation is the 1986 to 2016 water years. In alignment with ASO data availability, Chapter 4 focuses on the 2016 to 2022 water years.

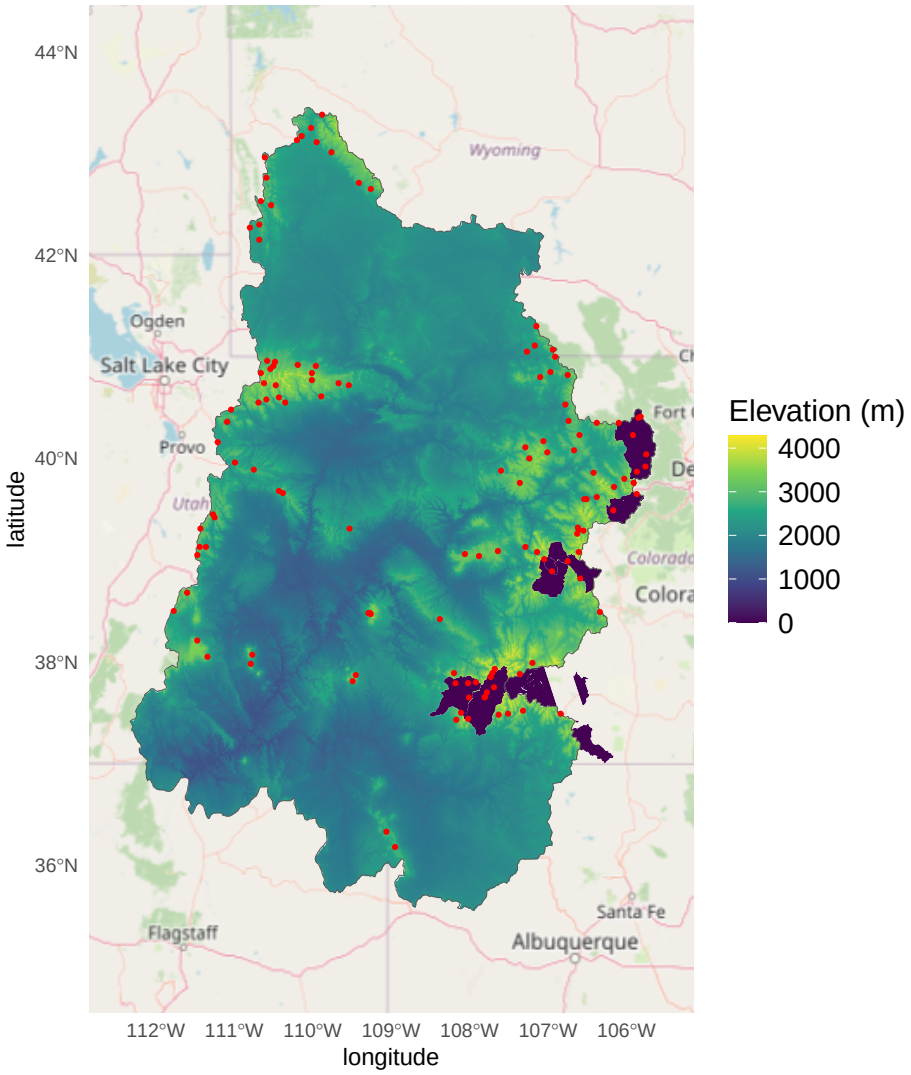


Figure 1.2: Geographical and topographic distribution of the UCRB hydrologic unit with overlaid Colorado ASO basins (dark purple) and SNOTEL sites (red dots) for reference.

The geographic location and spatial distribution of SNOTEL stations in the UCRB and ASO basins within or adjacent to the UCRB are shown in Figure 1.2.

In total, there are 135 SNOTEL stations of varying temporal lengths (depending on when they were set up and when data collection began) in the UCRB. After accounting for water years that contain missing or erroneous data, 712,845 site-days remain in our study period. The stations are generally located in higher elevation mountainous areas close to the eastern and western borders of the UCRB.

ASO flights in Colorado took place between 2016 and 2023. We consider basins with at least one flight within the 2016-2022 period, constituting ten basins within or adjacent to

the UCRB. In total, our study encompasses 26 basin-dates for which ASO data is available. All the basins are located on the east side of the UCRB, in more mountainous regions. A full list of ASO basins and associated flight dates is provided in Table 4.1.

1.5 University of Arizona Gridded SWE Product (UA SWE)

UA SWE combines high-quality in-situ SWE observations that are spatially and temporally limited with poorer-quality coarse datasets that have stronger spatiotemporal coverage. The method can be viewed as a bias correction of a physically-based estimate of SWE (based on PRISM climate data) using true SWE measured at observational stations (SNOTEL). It does so by interpolating (via kriging) scaled SNOTEL point measurements to match the resolution of PRISM and then applies a simple multiplication of the two quantities element-wise to produce a gridded SWE field matching the resolution of PRISM.

Mathematically, on each day t in a water year, the UA SWE algorithm combines SNOTEL point observations $i \in [1, \ldots, n]$ (where n is the number of SNOTEL sites) and PRISM grid values $j \in [1, \ldots, m]$ (where m is the number of PRISM grid cells) in the following formulation:

$$\text{UA SWE}_{j,t} = \phi(Y_{1,t}, Y_{2,t}, \ldots, Y_{n,t}) \cdot (N_{j,t}) \tag{1.2}$$

where ϕ is an interpolation from the point- to grid-level (specifically, ordinary kriging with spherical variogram, range = 20 km, sill = 0.2, and nugget = 0).

$Y_{i,t}$ is the SNOTEL ratio, defined as the ratio of observed SWE at SNOTEL sites, $O_{i,t}$, to SNOTEL net snowfall, i.e.,

$$Y_{i,t} = \frac{O_{i,t}}{S_{i,1:t}^{\text{SNOTEL}} - A_{i,1:t}^{\text{SNOTEL}}}. \tag{1.3}$$

Net snowfall is estimated by the difference between accumulated snow, $S_{i,1:t}$, and cumulative snow ablation, $A_{i,1:t}$. Accumulated snow is defined as the sum of precipitation values, $P_{i,t}$, on days where temperature, $T_{i,t}$ falls below a 2°C rain-snow partitioning threshold over time. Cumulative snow ablation is estimated by a quadratic regression (with coefficients $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$) on cumulative degree days, $D_{i,1:t}$ — sum of degree days, $d_{i,t}$, exceeding 0°C over time. Explicitly, accumulated snow is:

$$S_{i,1:t} = \sum_t s_{i,t}, \tag{1.4}$$

$$s_{i,t} = P_{i,t} \mid T_{i,t} < 2°\text{C}$$

and cumulative snow ablation is:

$$A_{i,1:t} = \hat{\beta}_0 + \hat{\beta}_1 D_{i,1:t} + \hat{\beta}_2 D_{i,1:t}^2, \tag{1.5}$$

$$D_{i,1:t} = \sum_{t} d_{i,t} \cdot 1_{d_{i,t}>0}$$

$N_{j,t}$ from Eq. (1.2) is PRISM net snowfall:

$$N_{j,t} = S_{j,1:t}^{\text{PRISM}} - A_{j,1:t}^{\text{PRISM}}$$

$S_{j,1:t}$ and $A_{j,1:t}$ are accumulated snow and cumulative snow ablation defined as in Eq. (1.4) and Eq. (1.5), respectively, but on the grid level using PRISM.

As is evident from the formulation of UA SWE outlined above, several assumptions were made in the algorithm's estimation of snow accumulation and snow ablation. UA SWE assumes that all precipitation falling under a certain fixed temperature threshold is snowfall and that all temperature exceeding a fixed threshold contributes to ablation. The algorithm also does not account for the effects of sublimation and rain on snow.

1.6 Reproducibility

Reconstruction of UA SWE

Unfortunately, the code to produce the UA SWE product was not available at the time of writing this dissertation. In order to be analyze errors and modeling decisions underlying the gridded dataset, we attempted to reconstruct the UA SWE algorithm from scratch to the best of our abilities based on methods described in [10], [11], and [83].

Our reconstruction of UA SWE differs from the UA SWE gridded product in the following ways:

- We reconstructed the UA SWE product only for our study area of the UCRB, rather than the WUS or CONUS. This has implications for kriging.
- We reconstructed the SWE variable of UA SWE and omitted the reconstruction of SD because it required a further snow density parameterization.
- We did not utilize snow depth data from the Cooperative Observer Program (COOP) network in our reconstruction.

Additionally, some specific details related to the UA SWE data assimilation method were missing or limited in the papers and documentation associated with the gridded product. As a result, we made a few decisions and assumptions in our reconstruction to ensure continuous time series across space. These are outlined below.

- We believe the rain-snow partitioning scheme used to estimate snow accumulation was a 2°C temperature threshold.

- We assumed that the model of cumulative ablation on cumulative degree days $> 0°C$ is a quadratic regression with zero intercept. We reconstructed this regression model based on visual comparison against Figure 3.1 from [10], since the exact regression coefficients were not provided. The regression model was also reconstructed based on a different set of SNOTEL sites — we specifically used data from SNOTEL sites in the UCRB in our reconstruction.

- On each day, for both SNOTEL and PRISM, ablation (via degree days) can only accumulate if the amount of accumulated snow will remain equal to or exceed the resulting cumulative ablation. In Chapter 2, we decided that snow accumulation and cumulative degree days would reset to zero on days where cumulative ablation derived from cumulative degree days would exceed the accumulated snow. In Chapter 3, we modified the approach and only allowed cumulative degree days to increase up to the point where the corresponding cumulative ablation is less than or equal to the total snow accumulation. This required working backward to find roots for the cumulative ablation quadratic regression.

- If the SNOTEL ratio $Y_{i,t} > 2$ (Eq. (1.3)), then we set $Y_{i,t} = 1$. This decision was made to avoid SNOTEL ratios approaching infinity.

- If SNOTEL net snowfall $= 0$ and SNOTEL observed SWE $O_{i,t} >= 0$, then we set $Y_{i,t} = 1$.

Diagnosis of UA SWE and UA SWE Reconstruction Underestimation

Underestimation of SNOTEL SWE by UA SWE and UA SWE Reconstruction

As seen in Figure 1.3, on average, at SNOTEL sites in the UCRB, throughout the water year, UA SWE underestimates SNOTEL SWE and our reconstruction of UA SWE underestimates UA SWE.

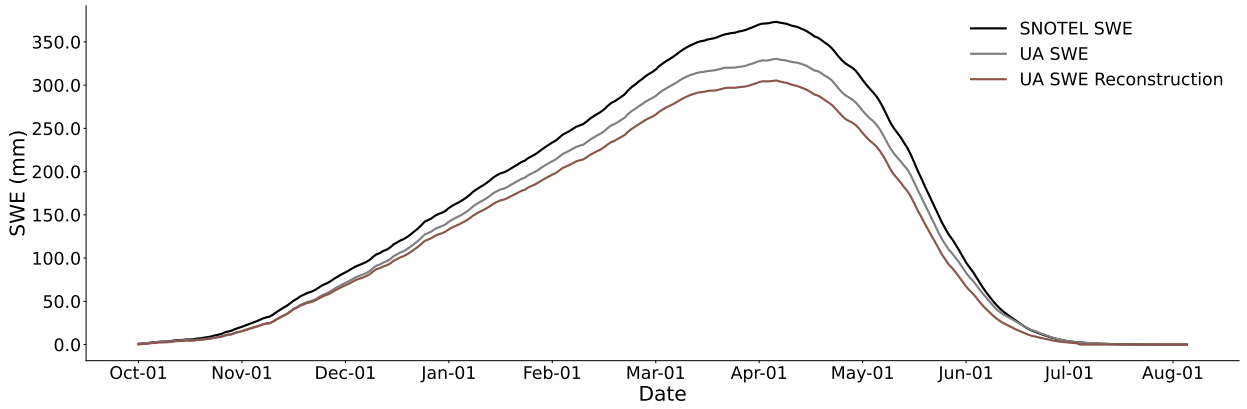


Figure 1.3: SNOTEL SWE compared to UA SWE and our reconstruction of UA SWE, averaged across sites and years for each water year day (October 1st through September 30th).

First, we aim to better understand the underestimation of UA SWE to SNOTEL. While acknowledging the expected lack of exact agreement between UA SWE and SNOTEL SWE due to their different spatial resolutions and resulting scale misalignment, we try to diagnose the underestimation. To do so, we examine the components and datasets underlying the algorithm. However, since the true decomposed components computed by UA SWE (i.e., the net snowfall, accumulated snowfall, cumulative ablation, precipitation, and temperature quantities) are not accessible, this exploration can only target the underestimation of our reconstruction of UA SWE to SNOTEL SWE.

According to Figure 1.4, the underestimation is attributed to the combination of PRISM net snowfall being insufficiently large and SNOTEL net snowfall being larger than SNOTEL SWE throughout the snow season. The latter results in SNOTEL ratios less than 1, which acts on PRISM net snowfall and further reduces the PRISM net snowfall estimates. The peak of SNOTEL net snowfall occurring later than SNOTEL SWE could suggest that the 2°C rain-snow partitioning threshold used to estimate accumulation causes systematic over-accumulation around the peak SWE period. Net snowfall showing persistent snow into August when SNOTEL SWE indicates all snow has melted reflects overestimation of snow accumulation and/or underestimation of cumulative ablation in the net snowfall calculation.

To diagnose the behavior of SNOTEL net snowfall contrasted against PRISM net snowfall, the net snowfall components can be decomposed into their snow accumulation and snow ablation counterparts (Figure 1.5). Evidently, throughout most of the water year, SNOTEL accumulated snow exceeds PRISM accumulated snow. SNOTEL cumulative ablation also exceeds PRISM cumulative ablation around and during the ablation season, but the magnitude of the difference is much smaller. The substantial difference in SNOTEL accumulated snow and PRISM accumulated snow drives the difference in SNOTEL net snowfall and PRISM net snowfall.

Further decomposing the accumulated snow and cumulative ablation components into their precipitation (P) and temperature (T) input datasets, we observe the striking underestimation of SNOTEL P by PRISM P (Figure 1.6), especially between November and May.

The underestimation of PRISM P to SNOTEL P can be attributed to several factors. The statistical interpolation procedure underlying PRISM smooths precipitation data across grid cells, which tends to diminish localized high precipitation events and overestimate low or zero precipitation areas [21, 17]. In complex mountainous regions, this smoothing effect leads to significant underestimation of large precipitation events because the interpolation struggles to capture the high spatial variability and intensity of these events [65]. Additionally, the reliance of PRISM on elevation for modeling may not fully account for the nuanced effects of wind direction, airflow dynamics, and complex terrain, further contributing to its dry bias [53, 27]. SNOTEL data themselves may experience measurement errors (e.g., wind-blowing effects and undercatch biases) [25, 62], exacerbating the discrepancy. Precipitation gauges are also often located in lower elevation areas (relative to surrounding landscape) that are easier to access, leading high elevation areas where precipitation is more variable and intense to be underrepresented in the PRISM interpolation [65]. Some differences between PRISM P and SNOTEL P may also arise from inherent challenges in comparing point data with grid cell averages [49, 65]. To assess whether this grid-to-point comparison significantly drives the discrepancy, analysis comparing PRISM P to other point-level precipitation datasets, such as from the GHCN, could be useful.

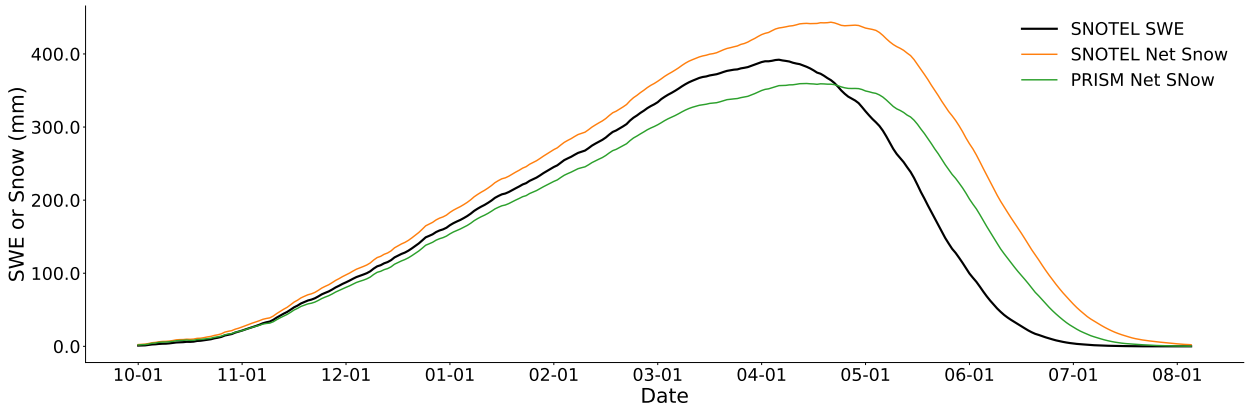


Figure 1.4: SNOTEL SWE compared to SNOTEL net snowfall and PRISM net snowfall at SNOTEL sites, averaged across sites and years on each day.

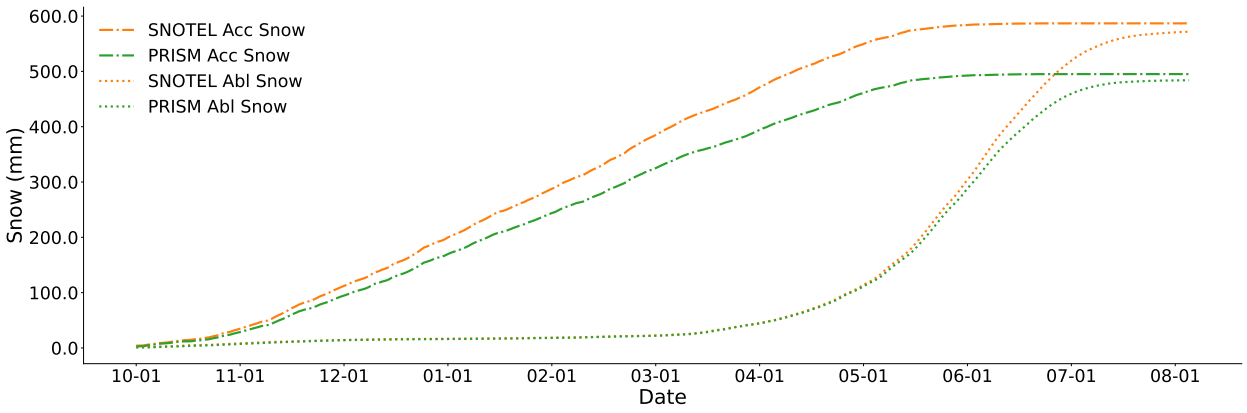


Figure 1.5: SNOTEL and PRISM accumulated snow and cumulative snow ablation at SNOTEL sites, averaged across sites and years on each day.

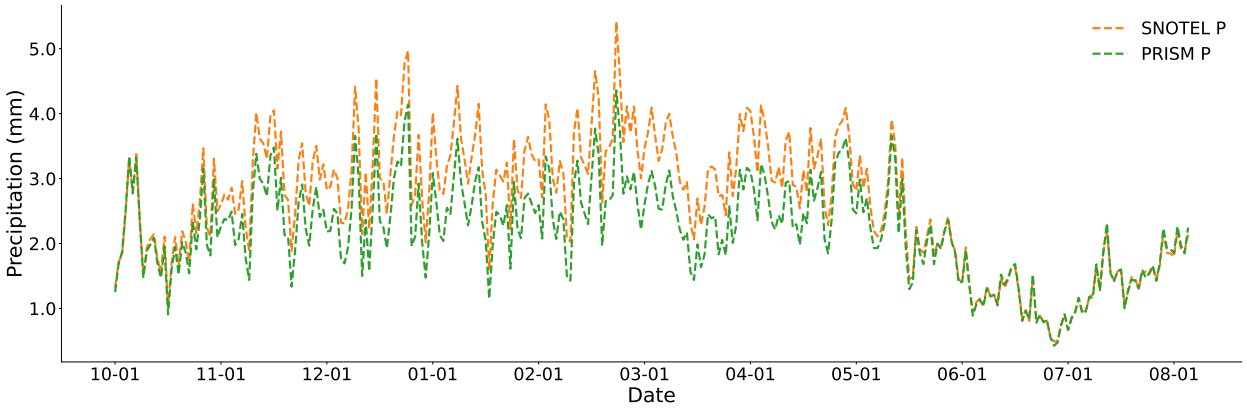


Figure 1.6: SNOTEL and PRISM precipitation at SNOTEL sites, averaged across sites and years on each day.

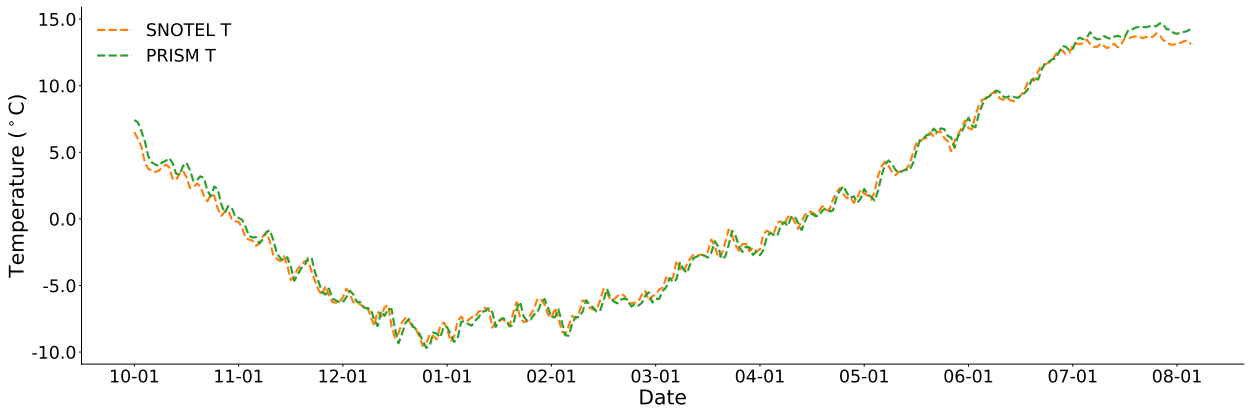


Figure 1.7: SNOTEL and PRISM temperature at SNOTEL sites, averaged across sites and years on each day.

Figures 1.8 and 1.9 compare the climatological mean and standard deviation (mean and standard deviation computed across all sites and years, for each day) of UA SWE, our reconstruction of UA SWE, and a systematic decomposition of UA SWE components against SNOTEL SWE at SNOTEL sites. In both figures, building upon SNOTEL net snowfall (orange line), "Avg Ratio SNOTEL Net Snow" (purple line) represents a simple bias-corrected SNOTEL net snowfall (where SNOTEL net snowfall on each day is scaled using a fixed SNOTEL ratio, which is the average SNOTEL ratio across all sites on that day) and "Krig Ratio SNOTEL Net Snow" (red line) represents an additional level of complexity in the scaling of SNOTEL net snowfall, through a more sophisticated interpolation approach that results in SNOTEL ratios that differ across sites on each day. "PRISM Net Snow" (green line), "Avg Ratio PRISM Net Snow" (blue line), and "Krig Ratio PRISM Net Snow" (purple line) are the analogous PRISM counterparts, where interpolation of the SNOTEL ratio increases in complexity, but is multiplied by PRISM net snowfall instead of SNOTEL net snowfall. The latter is equivalent to our reconstruction of UA SWE.

Collectively, the climatological time series show that net snowfall as a direct proxy for SWE is flawed, in line with what was observed in Figure 1.4. The seasonality is shifted and too much snow is predicted during the accumulation season. As a result, applying a correction to net snowfall is critical for offsetting these errors. While both the simple averaging and kriging approaches produce similar average SWE estimates, the kriging method exhibits greater variability, particularly during the ablation season, leading to a larger spread and closer alignment with SNOTEL SWE compared to the averaging approach (Figure 1.9). Moreover, importantly, the choice of input data (SNOTEL vs PRISM) strongly influences the resulting snow predictions. This is seen by the significant difference in net snowfall (and resulting SWE) derived from SNOTEL compared to PRISM, where PRISM underestimates. This strong impact of the choice of input data is also apparent from comparing the red and brown lines — when only SNOTEL data is used in the UA SWE algorithm,

the climatological mean and standard deviation of resulting SWE predictions is much closer to that of SNOTEL SWE. When PRISM data is incorporated into the algorithm, however, underestimation of the resulting SWE predictions becomes much more pronounced, visible throughout the entire water year.

In summary, Figures 1.4 to 1.9 suggest that the underestimation of SNOTEL SWE by UA SWE at SNOTEL sites is primarily due to discrepancies between PRISM and SNOTEL records. Specifically, the underestimation of PRISM P to SNOTEL P affects their respective accumulated snow and, consequently, the net snowfall estimations. The net snowfall formulation and interpolation approach do not fully account for the differences caused by discrepancies in the input datasets.

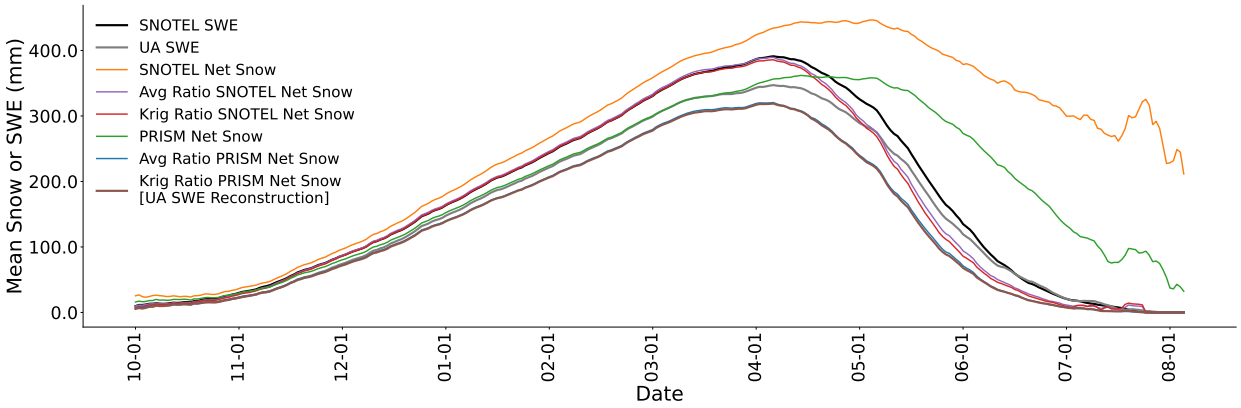


Figure 1.8: Climatological mean of SNOTEL SWE compared to that of UA SWE and UA SWE reconstruction, including decomposition of UA SWE components. Note that SNOTEL and PRISM net snowfall decrease after peak SWE due to the zero-resetting approach used to handle snow ablation exceeding snow accumulation (i.e., Chapter 2 approach). Values on dates prior to June 30 are of interest.

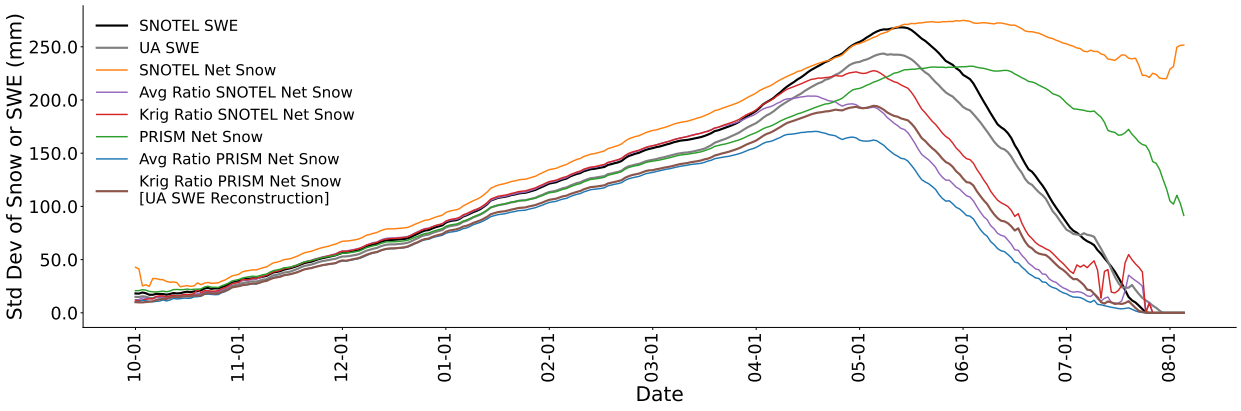


Figure 1.9: Climatological standard deviation of SNOTEL SWE compared to that of UA SWE and UA SWE reconstruction, including decomposition of UA SWE components. Values on dates prior to June 30 are of interest.

Underestimation of UA SWE by UA SWE Reconstruction

Next, we examine the underestimation of our reconstruction of UA SWE to the UA SWE product.

To investigate which behaviors are most influential in driving the underestimation of UA SWE Reconstruction, we categorize daily changes in UA SWE values into positive change (accumulation), no change, and negative change (ablation). No (zero) change in predicted SWE is possible because UA SWE publishes its SWE values as whole numbers. This is presumably a result of rounding their final predicted values. To align with UA SWE, we round our reconstructed SWE values to whole number for this analysis.

Within each of the three categories, we further stratify the daily changes in our UA SWE Reconstruction to identify systematic patterns. Figures 1.10, 1.11, and 1.12 illustrate this: the top panel displays the total difference in daily changes in UA SWE and UA SWE Reconstruction, aggregated across all days within each stratum of UA SWE Reconstruction behavior, for each of the three categories of daily UA SWE change, respectively. The bottom panel shows the proportion of days falling within these strata for each of the three categories of daily UA SWE change. Additionally, Figure 1.13 presents the total difference between daily changes in UA SWE and UA SWE Reconstruction (i.e., aggregated across all strata) for the three categories of daily UA SWE change, allowing us to assess the relative magnitude of differences among these categories.

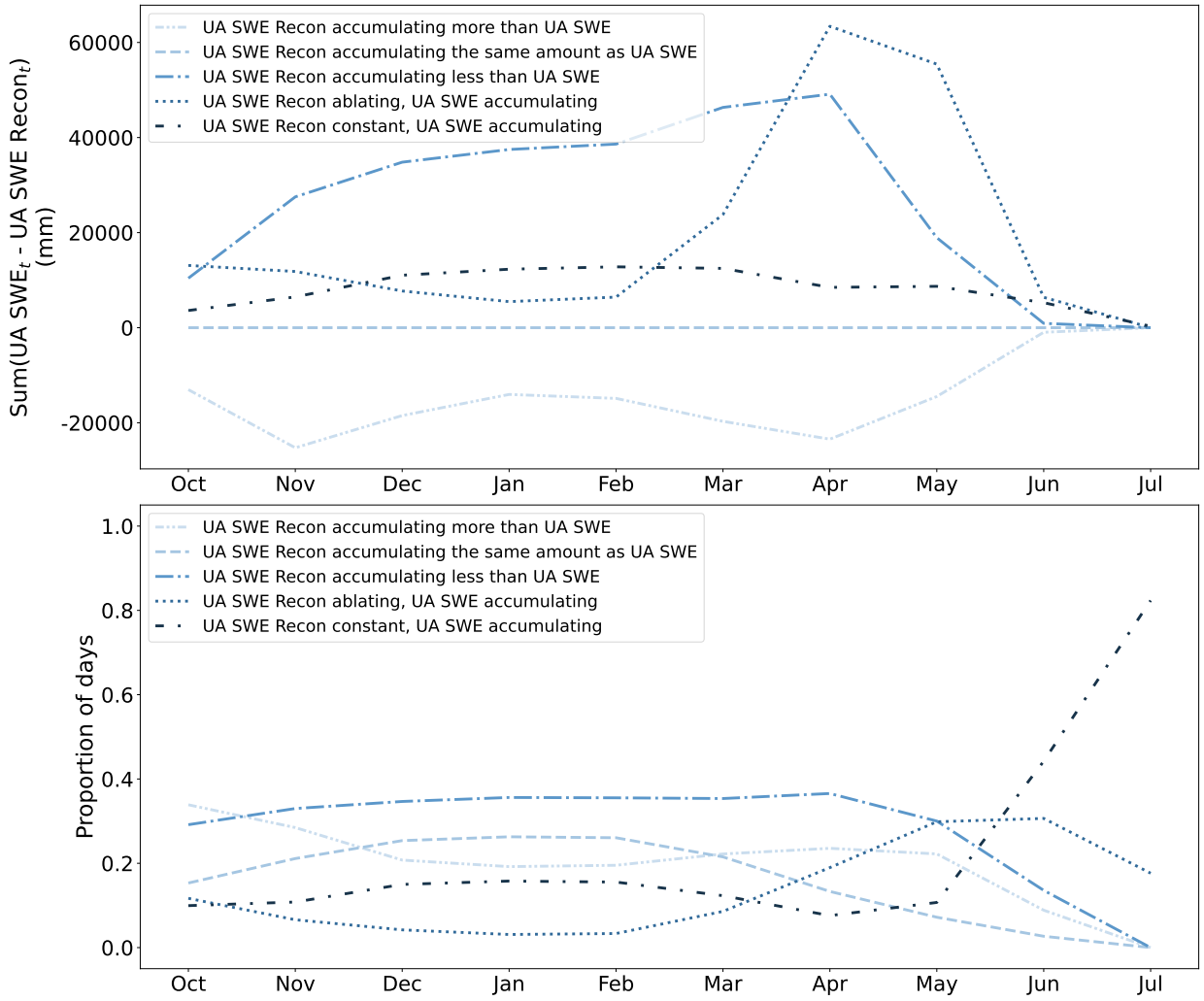


Figure 1.10: Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences positive change (accumulation). Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month.

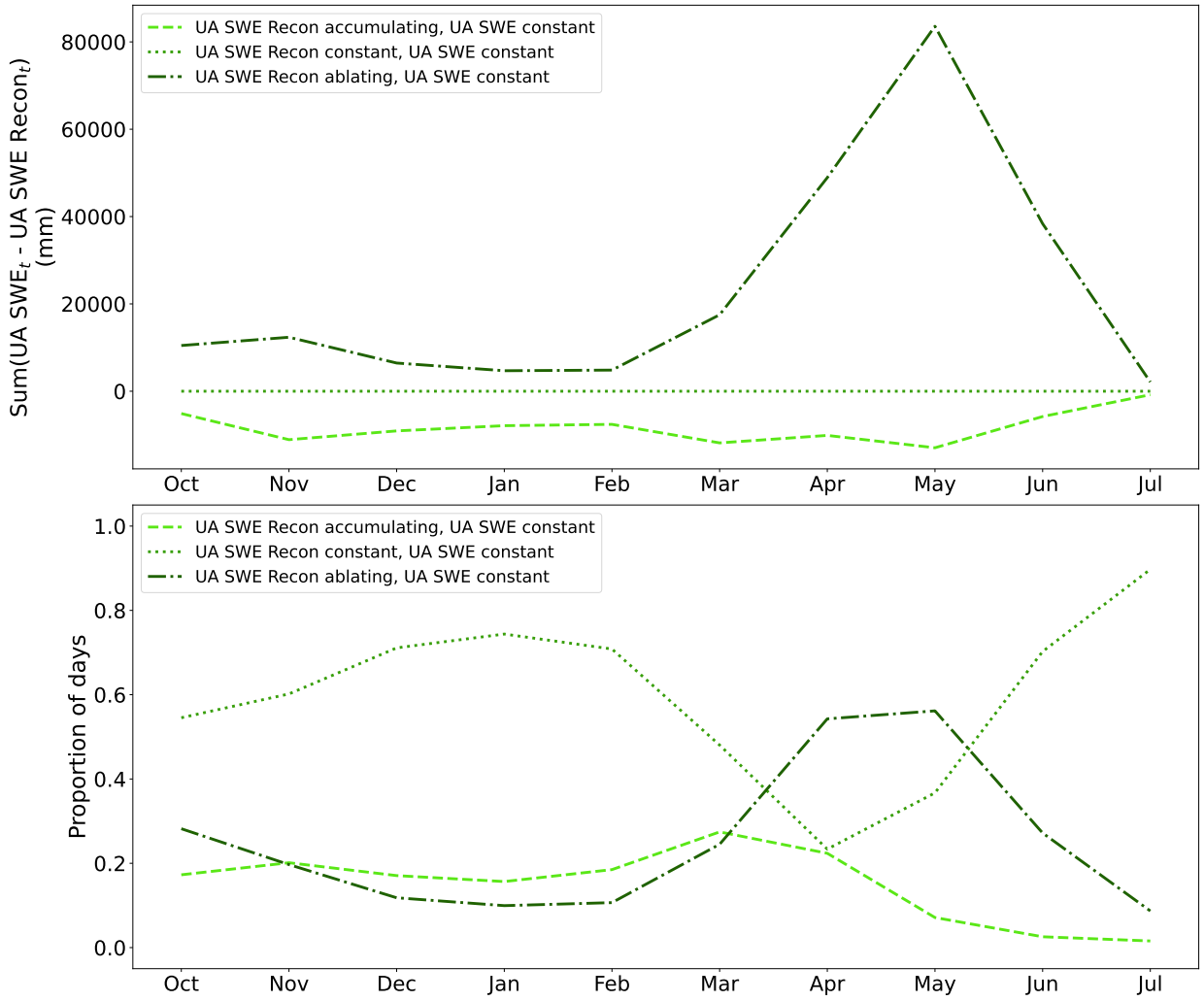


Figure 1.11: Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences no change. Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month.

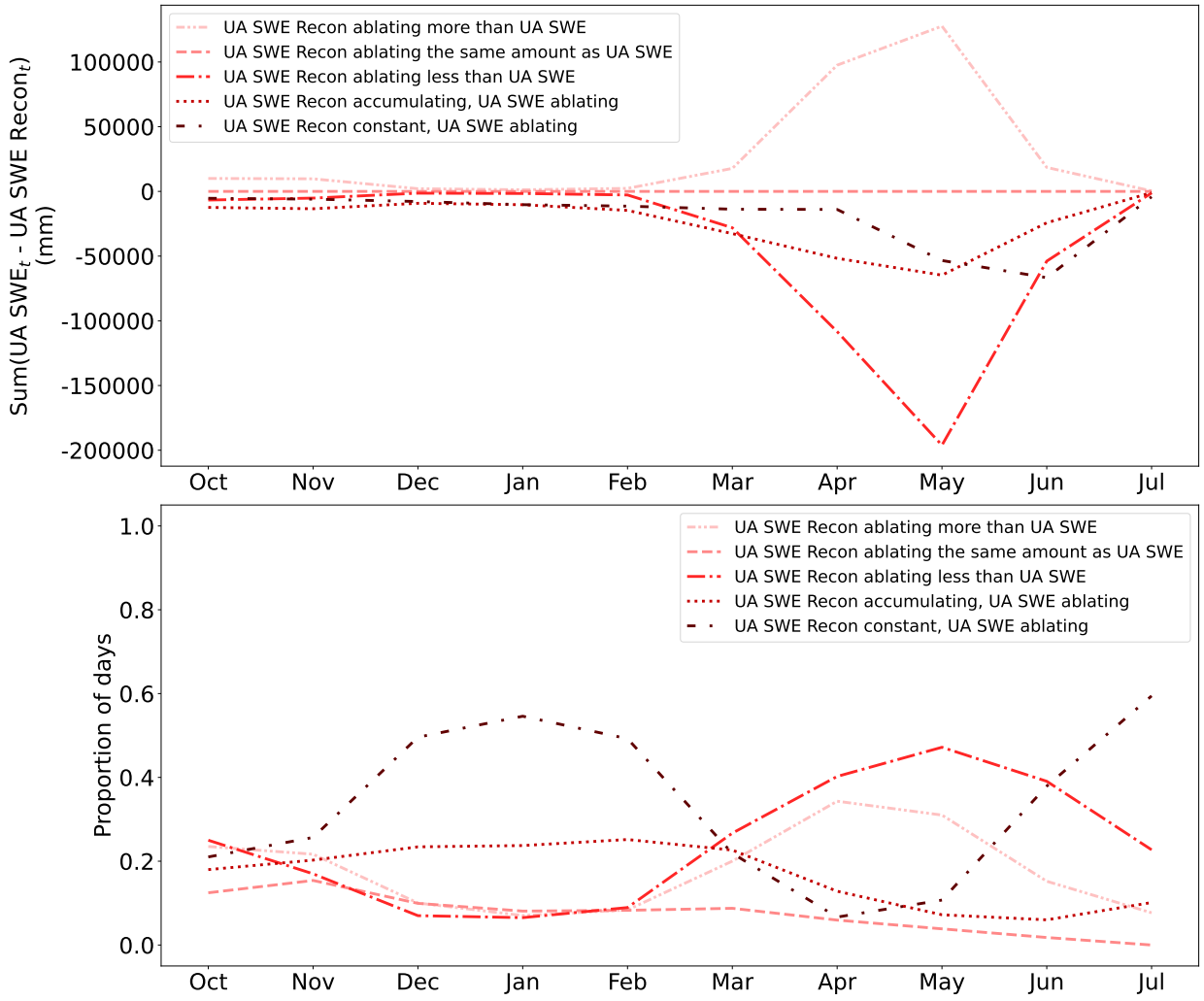


Figure 1.12: Stratification of the UA SWE Reconstruction behavior on days that UA SWE experiences negative change (ablation). Top panel shows, for each stratum, the total difference in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) across all sites and years for each month. Bottom panel shows proportion of days falling within each stratum for each month.

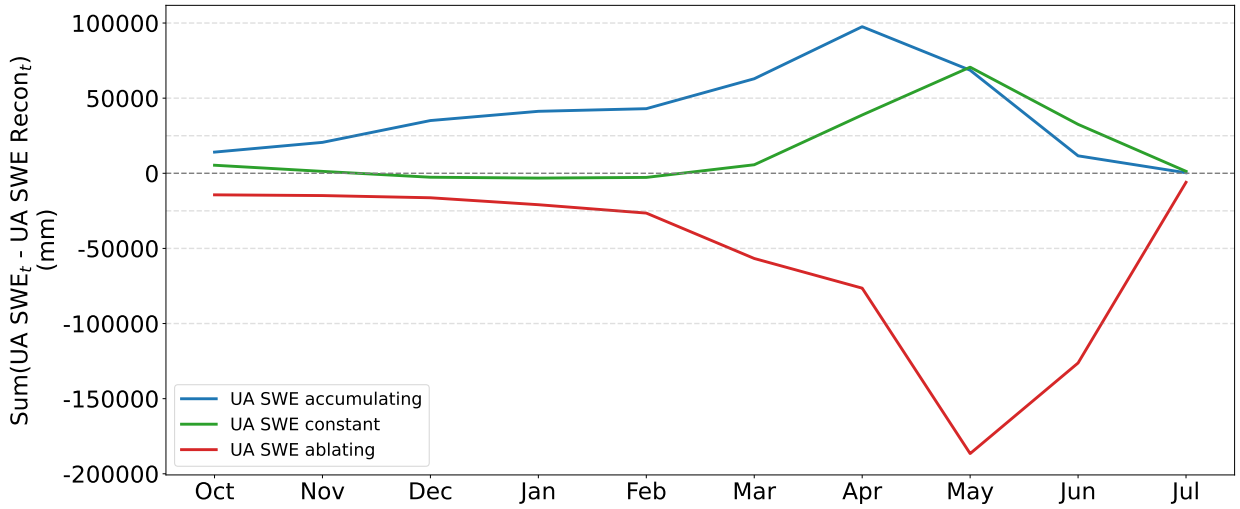


Figure 1.13: Total magnitude of differences in SWE predicted by UA SWE and UA SWE Reconstruction (UA SWE - UA SWE Reconstruction) for each category of UA SWE behavior (accumulation, no change, ablation), across all sites and years for each month.

We focus our discussion and analysis primarily on the accumulation season because water resource managers are most concerned with water quantity at peak SWE timing. Additionally, as the snow season progresses, the cumulative nature of UA SWE estimates introduces increasing confounding errors, making it more challenging to diagnose and disentangle behaviors.

During the accumulation season, specifically between November and April, the predominant behavior of UA SWE is accumulation (positive daily change), as the magnitude of total daily accumulation exceeds that of total daily constant and total daily ablation (Figure 1.13). Figure 1.10 shows that during this period, when UA SWE predicts accumulation, our UA SWE Reconstruction frequently predicts accumulation, but the quantity is lower. On days where UA SWE predicts no change in SWE, UA SWE Reconstruction most commonly also predicts no change in SWE (Figure 1.11, specifically bottom panel). When UA SWE predicts ablation, UA SWE Reconstruction most commonly predicts no change in SWE, with some instances of accumulation (Figure 1.12). However, because total UA SWE accumulation is considerably greater than ablation, especially between November and February, the overestimation of ablation by UA SWE Reconstruction has less impact on the discrepancy between UA SWE and UA SWE Reconstruction than the underestimation of UA SWE accumulation by UA SWE Reconstruction. Thus, a key reason for UA SWE Reconstruction's underestimation of UA SWE is its tendency to predict lower snow accumulation during the snowy months.

Besides the main patterns discussed in the previous paragraph, there are some discrepancies between UA SWE and UA SWE Reconstruction during the accumulation season that are puzzling. On the surface, it is confusing (1) why there are days where UA SWE Recon-

struction predicts ablation or no change in SWE when UA SWE predicts accumulation; (2) why the magnitude of accumulation predicted by UA SWE and UA SWE Reconstruction at times differ drastically, and (3) how UA SWE Reconstruction could estimate accumulation or no change in SWE when UA SWE predicts ablation. Some of these differences are due to kriging. For example, in Case (1) where UA SWE Reconstruction ablates when UA SWE accumulates, kriging is often the cause for the discrepancy when SNOTEL net snowfall and PRISM net snowfall both show accumulation. However, in other situations, it is much less clear whether the differences are due to differences in kriging or driven by differences in decisions made in data preprocessing, the net snowfall computation, and/or handling of numerical issues. In Case (1) where UA SWE Reconstruction ablates when UA SWE accumulates, for example, when SNOTEL net snowfall and PRISM net snowfall both exhibit zero or negative daily change, it is ambiguous as to why UA SWE accumulates. It could be due to kriging, but it could also point to aforementioned data and/or computational drivers. As another example, Case (3) where UA SWE Reconstruction shows zero change but UA SWE ablates often corresponds to days that experience zero SNOTEL and PRISM precipitation and negative SNOTEL and PRISM temperature. It is likely that negative daily change in UA SWE on these days is a result of their kriging; however, the possibility that data preprocessing steps and/or differences in the true UA SWE approach to computing net snowfall contribute to this negative change in UA SWE cannot be ruled out.

During the ablation season, UA SWE Reconstruction predicts large amounts of ablation when UA SWE predicts accumulation or no daily change in SWE. When UA SWE predicts ablation — which is the most dominant category (Figure 1.13) — UA SWE Reconstruction typically predicts lower ablation, though there are also notable instances where we predict higher ablation. The total amount of ablation overestimated by UA SWE Reconstruction across the three categories (sum across Figures 1.10, 1.11, and 1.12), exceeds the amount of ablation underestimated by UA SWE Reconstruction during May and June. This suggests that UA SWE Reconstruction may generally overestimate ablation, contributing to the overall underestimation compared to UA SWE. However, drawing a definitive conclusion is challenging due to the cumulative nature of the estimates and the influence of snow accumulation discrepancies from earlier in the year.

Examining the data, we identify some large magnitude differences between UA SWE and UA SWE Reconstruction during the ablation season, specifically in relation to our reconstruction's prediction of ablation. Investigation reveals that large negative daily changes in UA SWE Reconstruction, which contribute at least in part to the discrepancy with UA SWE, are sometimes directly attributable to kriging. However, there are also a notable number of days where SNOTEL SWE, SNOTEL net snowfall, and PRISM net snowfall values would suggest ablation, but UA SWE predicts accumulation (or no change). It is unclear if these discrepancies are due to kriging, data, and/or computation differences on the UA SWE side.

In sum, these findings suggest that underestimation of our UA SWE reconstruction to the published product is at least in part due to lower snow accumulation predicted throughout the accumulation season and possibly higher ablation predicted during the ablation season.

Discrepancies in the daily changes in SWE estimates can be attributed to one or more of the following factors. First, undocumented data cleaning or preprocessing steps in UA SWE, including possible regridding of the PRISM grid and methods for handling outliers or missing data, for which we made flawed assumptions. Second, differences in the interpolation (kriging) of the SNOTEL ratio, such as variations in the geographical extent of the interpolation. Third, the computation of net snowfall might differ due to modeling decisions and assumptions made in dealing with data intricacies discussed earlier, such as the rain-snow partitioning scheme, ablation regression, UA SWE's treatment of zero SNOTEL net snowfall, and how UA SWE dealt with occurrences where cumulative ablation exceeded snow accumulation.

To conclude, the UA SWE algorithm itself is relatively straightforward, particularly in terms of accumulation, which is where some of the major discrepancies occur between our reconstruction and their values. Given this, our inability to more closely match their values emphasizes the importance of reproducibility for public data products, in particular making code open source so that other researchers can fully reproduce a methodology in order to understand, assess and build upon the methodology.

Code Availability

Code associated with research in this dissertation can be found here.

1.7 Overview

In this chapter, we introduced key scientific terms, concepts, and datasets that will be referenced and leveraged throughout the dissertation. The remainder of this dissertation is organized as follows. In Chapter 2, we assess the sensitivity of UA SWE estimates to varied input data and modeling decisions and systematically decompose errors underlying the product. Taking inspiration from UA SWE, Chapter 3 proposes a novel method for estimating SWE that addresses some of the limitations in the UA SWE algorithm. Chapter 4 examines issues related to point-scale evaluation of gridded data and discusses possible factors contributing to misalignment between gridded SWE data and the true SWE surface, which is only known theoretically.

Chapter 2

Quantifying Variability and Error of a Gridded SWE Product

2.1 Introduction

Given the scientific and socioeconomic importance of snow and associated data, extensive understanding of the nature and magnitude of errors underlying gridded data products on snowpack properties such as SWE is essential for enhancing their reliability and utility. However, few studies have juxtaposed uncertainties in gridded SWE products. The studies that do exist found high uncertainty, variability, and bias across products in their estimates of SWE [46, 79, 81]. For example, [79] reported that interpolation methods employed by gridded snow products produce 40%-66% of the differences in WUS-wide peak SWE estimates [64].

In this chapter, we quantify errors underlying the UA SWE algorithm. Our approach encompasses a sensitivity analysis and an error decomposition to disaggregate total errors into individual components to disentangle the complex interplay of factors influencing predicted SWE variability and accuracy. By thoroughly examining errors and uncertainties underlying empirical SWE prediction, we aim to enhance understanding of limitations inherent in current gridded SWE data products and guide future development toward more robust and reliable estimates.

This chapter is organized as follows. Section 2.3 outlines the methodology employed for sensitivity analysis and error decomposition. Section 2.4 presents the results of our analyses, including the identification and quantification of major error sources. Finally, Section 2.5 summarizes our key conclusions and outlines recommendations for further investigation.

2.2 Datasets

A list of SWE datasets was outlined and described in Sections 1.3 and 1.5. This chapter utilizes SNOTEL SWE and UA SWE.

2.3 Methods

Our investigation of error in UA SWE is composed of a sensitivity analysis and an error decomposition. The sensitivity analysis quantifies how UA SWE predictions vary based on different input datasets and modeling assumptions. The error decomposition quantifies the incremental error contributed by different configurations of UA SWE. In this section, we first identify the major drivers of the UA SWE algorithm and introduce alternative modeling decisions for each driver. These default and alternative modeling decisions are used directly in the methodology underlying the two analyses, which we will then describe in detail.

Main Drivers of the UA SWE Algorithm

To understand the relative importance and impact of the different components that make up the SWE estimates derived from the UA SWE algorithm, we first identify the major drivers or sources of uncertainty in the UA SWE algorithm. These are:

1. Precipitation (P) in the PRISM net snowfall component
2. Average temperature (T) in the PRISM net snowfall component
3. Rain-snow partitioning threshold (Th) in the PRISM net snowfall component
4. Interpolation technique (ϕ) used to smooth the SNOTEL ratio

Other potential sources of uncertainty arise from assumptions in the modeling of cumulative snow ablation, and combining or comparing site (point) values and grid averages.

To analyze the variability and error in UA SWE, we consider alternative modeling decisions for each of these main drivers. We regenerate UA SWE by substituting alternative datasets for precipitation, average temperature, and the rain-snow partitioning threshold and alternative statistical approaches for interpolation. As summarized in Table 2.1, we substitute PRISM precipitation and temperature with GridMET and Livneh precipitation and temperature, respectively, the fixed 2°C rain-snow partitioning temperature threshold with a dynamic temperature threshold [29], and the default ordinary kriging (OK) approach with inverse distance weighting and with OK with a nonzero nugget.

To assess the variability and uncertainty in UA SWE produced by precipitation, temperature, and the rain-snow partitioning threshold, we omit interpolation from UA SWE and evaluate SWE uncertainty at the point-level at SNOTEL sites. To assess the variability in UA SWE due to interpolation, we include interpolation in UA SWE. In essence, our analysis is split into two separate cases: one of UA SWE at the site-level (hereafter, "ignoring interpolation") and one of UA SWE as is (hereafter, "including interpolation").

Key UA SWE Drivers	Modeling Decisions
Temperature	PRISM*
	GridMET
	Livneh
Precipitation	PRISM*
	GridMET
	Livneh
Rain-Snow Partitioning Threshold	Fixed 2°C*
	Jennings
Interpolation (ϕ)	OK with no nugget*
	OK with nugget = 0.04
	IDW

Table 2.1: The four key drivers of UA SWE and their associated datasets or statistical approaches considered in our sensitivity analysis and error decomposition. *indicates the default dataset or approach employed by the UA SWE algorithm.

Alternative Modeling Decisions

We describe each dataset and statistical approach outlined in the previous section.

Gridded Precipitation and Temperature Datasets

By default, PRISM net snowfall uses gridded precipitation and temperature data from the 4 km × 4 km PRISM product. PRISM uses advanced regression-based interpolation techniques to combine meteorological station data, including from the SNOTEL network, and topographic information to generate physically-informed gridded estimates of a range of climate variables, including precipitation and temperature [59].

While the choice of PRISM as the data source for precipitation and temperature is logical, other standard precipitation and temperature gridded datasets exist that could serve as alternatives to PRISM. We discuss the methodologies underlying GridMET and Livneh precipitation and temperature data in the following subsections.

GridMET P and T GridMET precipitation and temperature data (∼4 km spatial resolution) combine PRISM data with regional reanalysis (NLDAS) data [1].

GridMET precipitation is derived by scaling daily NLDAS precipitation with a ratio of monthly PRISM precipitation to monthly NLDAS P.

For minimum and maximum temperature, GridMET is derived by adding a correction factor of the difference between monthly average PRISM temperature and monthly average NLDAS temperature to daily NLDAS temperature. To obtain daily mean temperature for

use in our analyses, we take the average of the minimum and maximum temperature on each day.

Livneh P and T The generation of gridded precipitation and minimum and maximum temperature was described in Section 1.3. Livneh precipitation combines station and PRISM data, whereas Livneh temperature is derived from applying a constant lapse rate to station-interpolated temperature data [40, 41]. To obtain daily mean temperature for use in our analyses, we take the average of the minimum and maximum temperature on each day.

Since Livneh and UA SWE are not at the same spatial resolution (∼6 km compared to 4 km), nearest neighbor resampling was applied to match the Livneh grid to the UA SWE resolution when Livneh is used as input to UA SWE including interpolation.

Rain-Snow Partitioning

The UA SWE algorithm uses a fixed, or uniform, 2°C temperature threshold to partition precipitation into rainfall and snowfall in the accumulation of snow. Despite the utilization of a uniform rain-snow partitioning threshold being common in land surface and snow models, the air temperature at which rain and snow fall in equal frequency varies significantly across larger regions [29]. Given this, we consider an alternative approach to rain-snow partitioning using a dynamic thresholding approach.

Jennings Dynamic Rain-Snow Partitioning Product Using land-based meteorological station data, binary logistic regression models were optimized to predict daily snow occurrence based on combinations of temperature, relative humidity, and surface pressure variables. These models were trained and validated on 6- and 3-hourly synoptic weather reports, spanning 1978 to 2007. The trained models are then applied to daily summaries from a gridded reanalysis dataset to produce daily snowfall probabilities that are used to derive a rain-snow partitioning temperature threshold for each grid cell [29]. Figure 2.1 shows the Jennings rain-snow partitioning temperature thresholds across the UCRB.

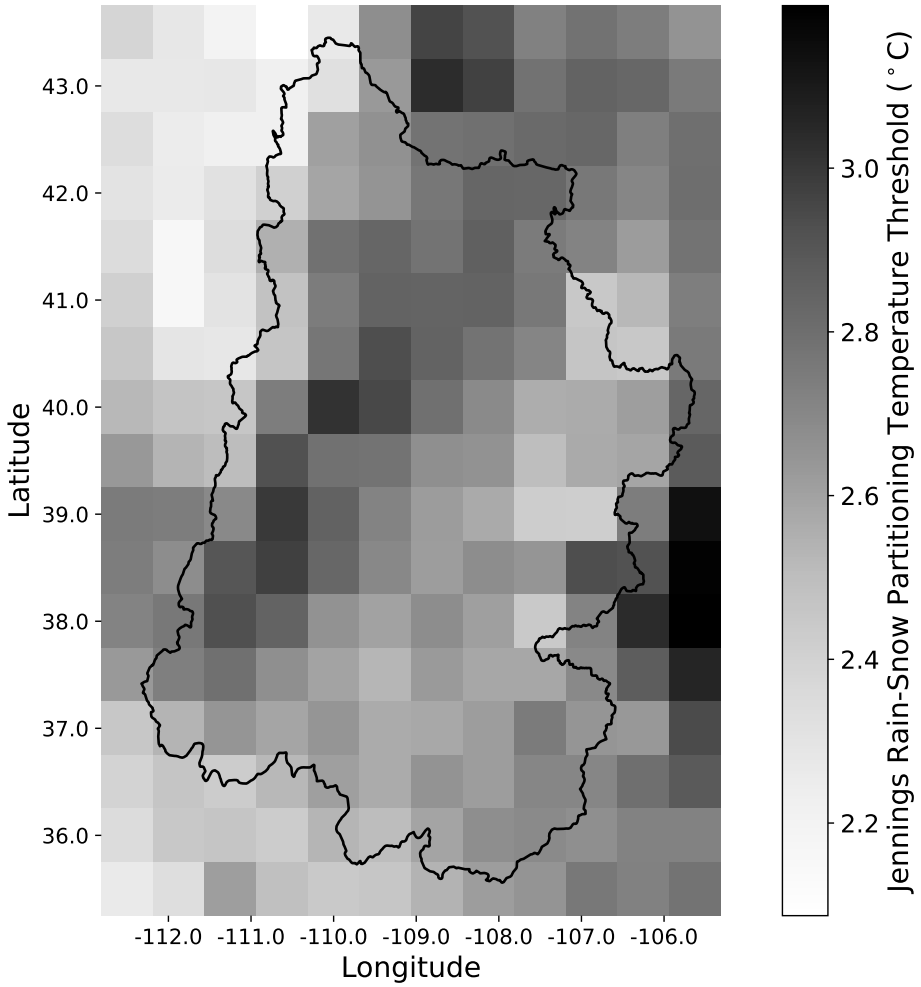


Figure 2.1: Jennings [29] dynamic rain-snow partitioning temperature threshold estimates, clipped to the UCRB region.

Interpolation Approaches

UA SWE employs OK with selected parameters. Since spatial interpolation may affect the accuracy of the resulting continuous field, part of our study strives to understand the sensitivity of UA SWE to the interpolation method used. In the next two subsections, the default kriging approach, along with alternative interpolation techniques considered in our analyses, are discussed.

Ordinary Kriging (OK) Kriging estimates values at unsampled locations by modeling spatial autocorrelation in the data. Works such as [28], [15], and [14] provide comprehensive overviews of the methodology. Only a summary of the general principles is included here.

In OK, the interpolated value $\hat{Z}(s_0)$ at a target location s_0 is calculated by:

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

where $Z(s_i)$ is the observed value at sampled location i and n is the total number of observations. The weight λ_i is informed by the spatial correlation structure described by a variogram model fit to the observed data, which helps determine how much influence each observation should have on the estimation at the target location.

By default, the interpolation method of choice in UA SWE is ordinary kriging with a spherical variogram with no nugget effect. Zero nugget means that no variation is assumed for infinitesimally small separation distances, leading to the estimated surface interpolating precisely through the observed data points. One of the alternative approaches we consider is the same ordinary kriging with a spherical variogram with a nugget effect, so that some variation is assumed for infinitesimally small separation distances. This nugget can be attributed to a combination of measurement errors and natural variation at distances smaller than the sampling interval.

Inverse Distance Weighting (IDW) Inverse distance weighting (IDW) estimates values at unsampled locations based on the weighted average of observed values at nearby locations.

In IDW, the interpolated value $\hat{Z}(s_0)$ at a target location s_0 is calculated by:

$$\hat{Z}(s_0) = \frac{\sum_{i=1}^{n} w_i Z(s_i)}{\sum_{i=1}^{n} w_i}$$

where n is the total number of observations (sampled locations). The weight w_i assigned to each observation $Z(s_i)$ is inversely proportional to its Euclidean distance, d_i, from the target location s_0:

$$w_i = \frac{1}{d_i^p}$$

where p is a user-defined parameter.

Sensitivity Analysis

We use analysis of variance (ANOVA) to partition the total variance in SWE estimated by the UA SWE algorithm. In other words, ANOVA is used to evaluate the contribution of each source of uncertainty to the total predicted uncertainty in UA SWE. As discussed in the previous section (Section 2.3), the major sources of uncertainty in SWE estimation are temperature, precipitation, the rain-snow partitioning threshold, and interpolation. As SWE naturally varies across space and time, spatial and temporal effects also play a significant role in generating variability in SWE estimation. These factors that drive the major sources of uncertainty in estimating SWE as well as the alternative modeling datasets or approaches substituted for each corresponding source, as applicable, are summarized in Table 2.2.

Factors	Estimates
Temperature (T)	PRISM GridMET Livneh
Precipitation (P)	PRISM GridMET Livneh
Rain-Snow Partitioning Threshold (Th)	Fixed 2°C Jennings
Interpolation (ϕ)	OK with no nugget OK with nugget = 0.04 IDW
Location	First Site/Grid Cell ⋮ Last Site/Grid Cell
Year	WY 1986 ⋮ WY 2016

Table 2.2: Estimates corresponding to each factor in our ANOVA setup.

We carry out two variance decomposition analyses: the first is an ANOVA analysis on UA SWE ignoring interpolation and the second is an ANOVA analysis on UA SWE including interpolation. In both cases, an ANOVA analysis is independently conducted for each day, enabling understanding of how the sources of uncertainty influencing UA SWE evolves throughout the water year.

The ANOVA analysis on UA SWE ignoring interpolation is conducted at SNOTEL site locations, where the generation of UA SWE estimates takes place at grid cells containing SNOTEL sites and therefore the interpolation of the SNOTEL ratio is omitted. This setup explicitly highlights the variability produced by precipitation, temperature, and the rain-snow partitioning threshold. Note that since no gridding is involved in this analysis, data from PRISM, GridMET, Livneh, and Jennings grid cells containing SNOTEL sites are first extracted and then used as input to the UA SWE algorithm.

The ANOVA analysis on UA SWE including interpolation is conducted for two sets of grid cells:

1. Grid cells containing SNOTEL sites.

2. Snow covered grid cells, defined via the snow seasonality metric (SSM) [56, 57]:

$$\text{SSM} = \frac{\text{Days}_{\text{Seasonal Snow}} - \text{Days}_{\text{Ephemeral Snow}}}{\text{Days with Snow}}$$

where $\text{Days}_{\text{Seasonal Snow}}$ is the number of days with at least 60 days of continuous non-zero SWE and $\text{Days}_{\text{Ephemeral Snow}}$ is the number of remaining days with non-zero SWE. By definition, completely ephemeral snowpacks receive an SSM value of -1 whereas the SSM value of completely seasonal snowpacks is +1. The SSM is calculated over each individual water year. We define "snow covered grid cells" as those whose average SSM across the water years in the study period exceeds 0.

The analysis on the first set of grid cells provides a useful comparison to the site-level setup, while the analysis on the second set of snow covered grid cells provides a more comprehensive picture of the behavior of UA SWE across the region at large. These grid-level analyses inform on the variability produced by interpolation and the potential effect of interpolation on precipitation, temperature, and the rain-snow partitioning threshold.

Mathematically, ANOVA splits the total sum of squares (SST) into sums of squares due to the individual effects. In the case of UA SWE ignoring interpolation, the SST can be decomposed as:

$$\begin{aligned} SST &= SS_P + SS_T + SS_{Th} + SS_{\text{loc}} + SS_{\text{year}} &+ SS_I &+ SS_{resid} \\ &= \text{main effects} &+ \text{interactions} &+ \text{residuals} \end{aligned} \tag{2.1}$$

where SS_I encompasses the interactions between the main effects of interest: P, T, and Th. SS_{resid} explains the remaining variation, including interactions between location and year with the main effects of interest. A detailed mathematical formulation of the decomposition in Eq. 2.1 is available in Appendix A.1.

Since the individual interaction we are particularly interested in is the precipitation and temperature interaction, we can rewrite the above equation as follows:

$$\begin{aligned} SST &= SS_P + SS_T + SS_{Th} + SS_{\text{loc}} + SS_{\text{year}} &+ SS_{P:T} + SS_{I^*} &+ SS_{resid} \\ &= \text{main effects} &+ \text{interactions} &+ \text{residuals} \end{aligned}$$

In the case of UA SWE including interpolation, the SST can be decomposed as:

$$\begin{aligned} SST &= SS_P + SS_T + SS_{Th} + SS_\phi + SS_{\text{loc}} + SS_{\text{year}} &+ SS_I &+ SS_{resid} \\ &= \text{main effects} &+ \text{interactions} &+ \text{residuals} \end{aligned} \tag{2.2}$$

Similarly, to highlight the precipitation and temperature interaction, we can once again re-organize the terms:

$$\begin{aligned} SST &= SS_P + SS_T + SS_{Th} + SS_\phi + SS_{\text{loc}} + SS_{\text{year}} &+ SS_{P:T} + SS_{I^*} &+ SS_{resid} \\ &= \text{main effects} &+ \text{interactions} &+ \text{residuals} \end{aligned}$$

If the impact of certain other interaction terms is desired, we can rearrange the interaction term in the same way.

The formulation above in Eq. 2.1 and Eq. 2.2, denoted "Comprehensive ANOVA," is general in that the individual sums of squares are computed based on SWE estimates from all locations (sites or grid cells) and years. However, since we are more interested in the amount of variability associated with the various modeling configurations, we focus on an alternative approach to aggregate sums of squares. Instead of a single ANOVA conducted for SWE estimates across all locations and years, a separate ANOVA is conducted at each location across years and then the sums of squares are added across locations. Concretely, for d sites or grid cells, when interpolation is ignored in UA SWE:

$$SST = \sum_{l=1}^{d} \left(SS_{P_l} + SS_{T_l} + SS_{Th_l} + SS_{\text{year}_l} + SS_{I_l} + SS_{resid_l} \right)$$

and when interpolation is included:

$$SST = \sum_{l=1}^{d} \left(SS_{P_l} + SS_{T_l} + SS_{Th_l} + SS_{\phi_l} + SS_{\text{year}_l} + SS_{I_l} + SS_{resid_l} \right)$$

This formulation removes spatial (location) effects in the computation of SST, more effectively isolating the main effects of interest. Since spatial effects are no longer present, the residuals in this formulation encapsulate the interactions between year and the P, T, and Th main effects.

Error Decomposition

In the error decomposition analysis, we pull apart and systematically decompose errors in the UA SWE algorithm to understand the incremental error injected into UA SWE estimates by the various modeling decisions, assumptions, and input datasets employed by the algorithm, as well as how the errors change should alternative modeling decisions or input datasets be used.

We decompose errors following two setups: one of UA SWE ignoring interpolation and one of UA SWE including interpolation. Errors from both are computed at the point-level at SNOTEL sites (so that model-derived SWE estimates can be directly compared to groundtruth SNOTEL SWE), but the generation of SWE estimates in the latter analysis incorporates kriging. Decomposing errors when excluding interpolation from UA SWE allows for direct evaluation of errors contributed by input datasets and modeling decisions aside from interpolation, whereas including interpolation in the assessment of UA SWE error informs on the effect of smoothing methods as well as their interactions with input datasets and other modeling decisions, such as the rain-snow partitioning approach.

Note that there are subtle differences in the ordering of point-to-grid alignment between the error decomposition of UA SWE ignoring interpolation and UA SWE including interpolation. In the case of UA SWE ignoring interpolation, the grid cells containing the SNOTEL

sites are first extracted from PRISM, GridMET, Livneh, and/or Jennings data then the UA SWE algorithm is run based on these inputs, followed by evaluation of predicted SWE against SNOTEL SWE at the sites. In the case of UA SWE including interpolation, the UA SWE algorithm is first executed with gridded PRISM, GridMET, Livneh, and/or Jennings data for the PRISM net snowfall component. Then grid cells containing SNOTEL sites are extracted for evaluation of predicted SWE against SNOTEL SWE at the sites. Hence, the difference between the two types of runs lies in which variable(s) the point-to-grid comparison takes place. In UA SWE ignoring interpolation, the point-to-grid comparison manifests in the effect of SNOTEL T and/or P versus PRISM T and/or P through the algorithm. In UA SWE including interpolation, the point-to-grid comparison only manifests in the evaluation stage when comparing SNOTEL SWE to the UA SWE prediction.

While the error decomposition analyses are characterized as understanding the error contributed by various components of the algorithm, part of the error is natural because UA SWE is a gridded product using gridded inputs and is being evaluated against point level data.

Error in UA SWE

First, we dissect the UA SWE algorithm and quantify errors contributed by each modeling decision in the algorithm. To do this, we incrementally modify a component in UA SWE, starting from SNOTEL SWE. The relative magnitude contributed by each UA SWE configuration C when compared to SNOTEL SWE is:

$$\epsilon_C = \frac{\text{SWE}_C}{\text{SWE}_{\text{SNOTEL}}} \tag{2.3}$$

In the case of UA SWE ignoring interpolation, the alternative configurations of UA SWE consist of substituting SNOTEL P for PRISM P and then SNOTEL T for PRISM T in the PRISM net snowfall component. In the case of UA SWE including interpolation, the alternative configurations of UA SWE consist of adding incremental complexity to the smoothing of the SNOTEL ratio.

The percentage difference, or error, between SWE_C and $\text{SWE}_{\text{SNOTEL}}$ (i.e., $\frac{\text{SWE}_C - \text{SWE}_{\text{SNOTEL}}}{\text{SWE}_{\text{SNOTEL}}}$) can subsequently be approximated by:

$$\delta_C = \log(\epsilon_C)$$

since $\log \epsilon_C \approx \epsilon_C - 1$ for $\epsilon_C = 1$.

Error from Alternative Modeling Decisions

To quantify the error contributed by each source of uncertainty, the SNOTEL ratio in the UA SWE algorithm is kept fixed while the other components are modified. The error decomposition setup can be thought of as building blocks adding steps of error components (or complexity). The relative magnitude contributed by each alternative modeling decision

in each step can be quantified in the following manner based on comparing predicted SWE in the current step to predicted SWE in the previous step:

$$\epsilon_X = \frac{\text{SWE}_X}{\text{SWE}_{X-1}} \tag{2.4}$$

for $X = 2, 3$ in the case of UA SWE ignoring interpolation and $X = 2, 3, 4$ in the case of UA SWE including interpolation. SWE_1 corresponds to SNOTEL observed SWE (equivalent to $\text{SWE}_{\text{SNOTEL}}$ in Eq. (2.3)).

As before, the error contributed by each modeling decision to the peak SWE value is the percentage difference between SWE_X and SWE_{X-1} (i.e., $\frac{\text{SWE}_X - \text{SWE}_{X-1}}{\text{SWE}_{X-1}}$), which can be approximated by:

$$\delta_X = \log(\epsilon_X) \tag{2.5}$$

since $\log \epsilon_X \approx \epsilon_X - 1$ for $\epsilon_X = 1$.

In the case of UA SWE ignoring interpolation, only the PRISM net snowfall component is modified through the substitution of alternative P, T, and Th datasets (Table 2.1). The error decomposition steps can therefore be thought of in the following manner, with the first step corresponding to the baseline ground truth SNOTEL SWE:

1. SNOTEL observed SWE, which equates to using SNOTEL T and P data in place of PRISM T and P to compute the PRISM net snowfall component of UA SWE when interpolation is ignored.

2. UA SWE ignoring interpolation. The error from this configuration, corresponding to δ_2 in Eq. (2.5), is calculated by comparing it to SNOTEL SWE (Step 1), which serves as the reference.

3. SWE estimated by the UA SWE algorithm ignoring interpolation, with modified input data for computing PRISM net snowfall. Here, the PRISM net snowfall component is computed using either GridMET/Livneh T and/or P data in place of PRISM T and/or P or the dynamic Jennings Th in place of the default fixed 2°C Th. The error from each of these alternative configurations is summarized in δ_3 in Eq. (2.5); each is calculated by comparing the alternatively derived UA SWE prediction to default UA SWE ignoring interpolation (Step 2), which serves as the reference.

   To compare against SNOTEL SWE as the reference (if desired), this can be achieved mathematically by $\epsilon_3 \cdot \epsilon_2 = \frac{\text{SWE}_3}{\text{SWE}_2} \cdot \frac{\text{SWE}_2}{\text{SWE}_1} = \frac{\text{SWE}_3}{\text{SWE}_1}$. Subsequently, for error approximation, a logarithmic transformation can be applied to $\epsilon_3 \cdot \epsilon_2$, as in Eq. (2.5).

In the case of UA SWE including interpolation, both the PRISM net snowfall component and the interpolation is modified. The error decomposition steps are outlined in the steps below. We employ spatial cross validation ($k = 4$ splits). Note that the errors are computed at grid cells corresponding to the SNOTEL sites rather than grid cells spanning the entire geographic region of the UCRB. This is to allow for direct validation against SWE measured at the SNOTEL sites, since the true SWE is unknown at the grid-level.

1. SNOTEL observed SWE, which equates to using SNOTEL T and P data in place of PRISM T and P to compute the PRISM net snowfall component of UA SWE and when interpolation is ignored (equivalently, when interpolation does not influence the SNOTEL ratio).

2. UA SWE including interpolation (default OK with no nugget). Here, SNOTEL T and P data continue to be used in place of PRISM T and P to compute the PRISM net snowfall component. The error from this configuration, corresponding to δ_2 in Eq. (2.5), is calculated by comparing it to SNOTEL SWE (Step 1), which serves as the reference.

3. SWE as estimated directly by UA SWE, where PRISM T and P are used to compute the PRISM net snowfall component. The error from this configuration, corresponding to δ_3 in Eq. (2.5), is calculated by comparing it to predicted SWE from Step 2, which serves as the reference.

   If desired, comparing against SNOTEL SWE as the reference can be achieved mathematically by $\epsilon_3 \cdot \epsilon_2 = \frac{\text{SWE}_3}{\text{SWE}_2} \cdot \frac{\text{SWE}_2}{\text{SWE}_1} = \frac{\text{SWE}_3}{\text{SWE}_1}$. Subsequently, for error approximation, as in Eq. (2.5), a logarithmic transformation can be applied.

4. SWE estimated by the UA SWE algorithm, with modified input datasets or interpolation method. Here, either GridMET/Livneh T and/or P data is used in place of PRISM T and/or P to compute the PRISM net snowfall component, the dynamic Jennings Th is used in place of the default fixed 2°C Th to compute the PRISM net snowfall component, or IDW/OK with nugget is applied in place of the default OK with no nugget for the interpolation of the SNOTEL ratio. The error from this configuration, corresponding to δ_4 in Eq. (2.5), is calculated by comparing it to UA SWE (Step 3), which serves as the reference.

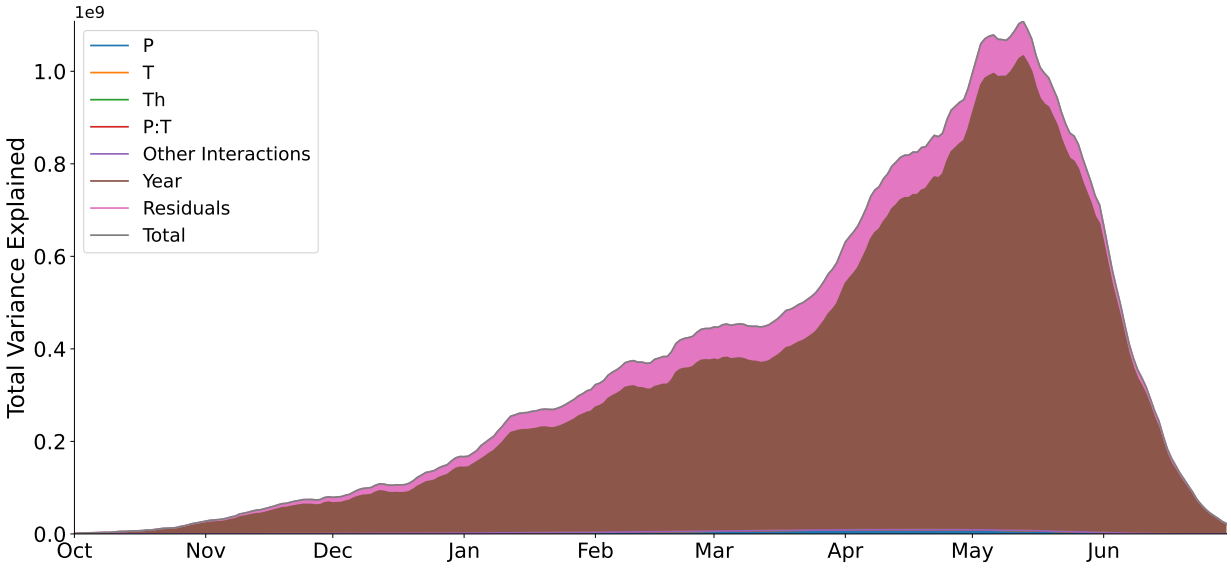
   Once again, to compare against SNOTEL SWE as the reference (if desired), this can be achieved mathematically by $\epsilon_4 \cdot \epsilon_3 \cdot \epsilon_2 = \frac{\text{SWE}_4}{\text{SWE}_3} \cdot \frac{\text{SWE}_3}{\text{SWE}_2} \cdot \frac{\text{SWE}_2}{\text{SWE}_1} = \frac{\text{SWE}_4}{\text{SWE}_1}$. Subsequently, for error approximation, a logarithmic transformation can be applied on $\epsilon_4 \cdot \epsilon_3 \cdot \epsilon_2$, as in Eq. (2.5).

2.4 Results

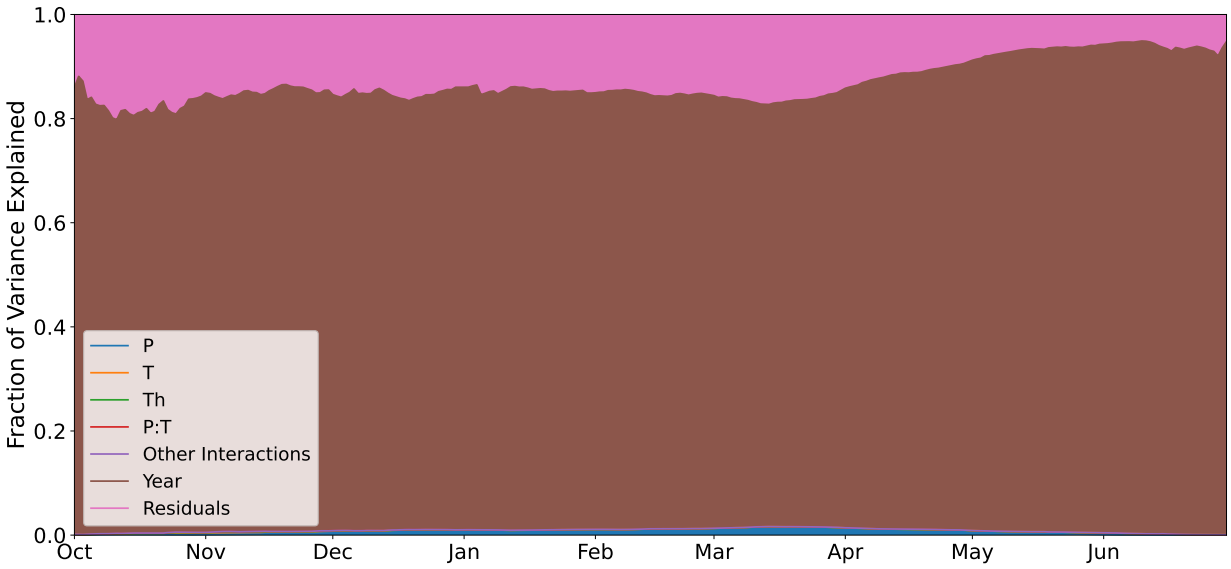
Sensitivity Analysis

Ignoring Interpolation (Point-Level)

Ignoring interpolation in the UA SWE algorithm allows us to assess the variability in SWE estimates produced by the algorithm formulation at the site locations directly.



(a) Total variance.



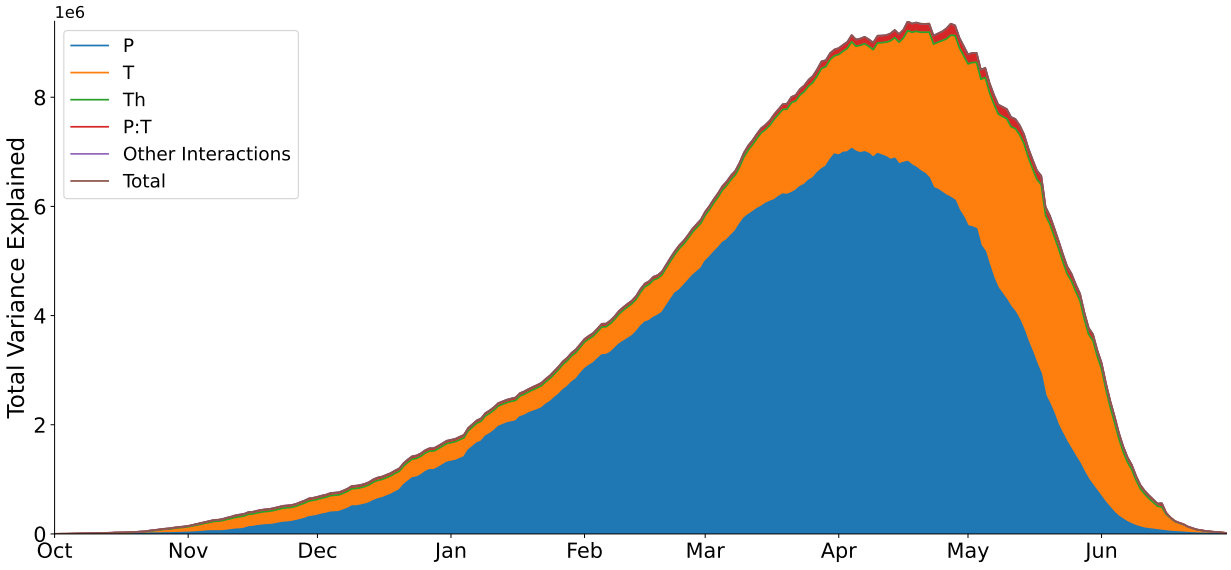
(b) Fractional variance.

Figure 2.2: Aggregated site-specific ANOVAs of fractional and total variance explained by each factor on each day at SNOTEL sites.

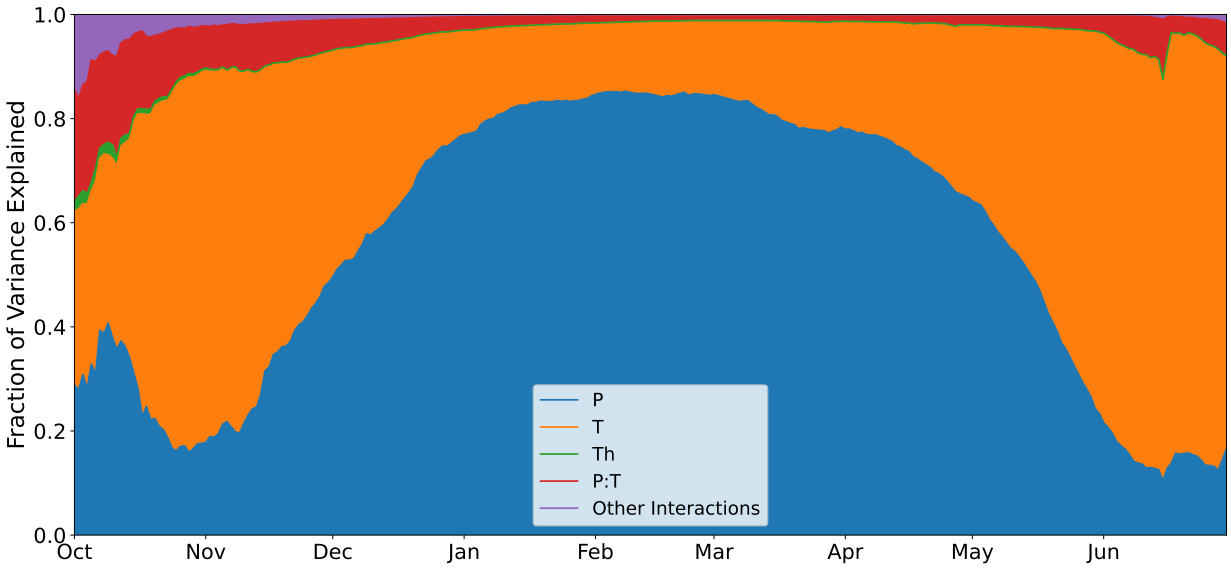
Figure 2.2 shows the fractional and numerical contribution of each factor to the total variance in UA SWE on each day through the snow season. It is apparent that the majority ($> 98\%$) of the variability in UA SWE estimates is in fact due to temporal variability across years and residual variability (which encapsulates temporal interactions with other factors),

with temporal effects being the most dominant, explaining over 80% of the total UA SWE variability across the snow season. The variance explained by the main effects of interest (precipitation, temperature, rain-snow partitioning threshold) is barely visible in Figures 2.2b and 2.2a.

To obtain a clearer view of the factors of interest, we remove the temporal and residual effects from the illustration, zooming in on the variability explained by the main effects of interest. This is shown in Figure 2.3.



(a) Total variance.



(b) Fractional variance.

Figure 2.3: Aggregated site-specific ANOVAs of fractional and total variance explained by each factor on each day at SNOTEL sites. Temporal and residual effects are excluded to highlight main effects of interest.

Figure 2.3a shows that there is low variability in UA SWE estimates at the beginning and end of the snow season (October and June, respectively). During the snow season, variability in the UA SWE algorithm gradually increases as the snow season progresses, peaking in

April when SWE is often at a maxima. Throughout the accumulation and ablation seasons, variability in the UA SWE algorithm is predominantly driven by precipitation and temperature, accounting for over 90% of the total variability contributed by modeling decisions, as seen in Figure 2.3b. In October, precipitation contributes more variability than temperature. Between mid-October and mid-December, temperature contributes more variability than precipitation. Between mid-December and early May, precipitation plays the dominant role in contributing to the total variability. From May onwards (through the end of the snow season in late June), temperature gradually contributes more to the total variability as SWE fully ablates.

Since snow falls and ablates during most of the snow season, the effects of both precipitation and temperature are present and significant in explaining UA SWE variability throughout the snow season. As the accumulation season progresses during the year, the effect of precipitation on UA SWE variability gradually increases and becomes the most dominant driver of SWE estimation, with its influence peaking around the peak SWE period (late winter to early spring, typically around March and April). This is expected since the computation of snow accumulation depends on precipitation data and SWE quantity is heavily driven by falling snow during these months. Additionally, temperature is typically low during these months, which means there is little effect on the rain-snow partitioning and no considerable ablation. As the ablation season unfolds following the peak SWE period, temperature gradually explains a more considerable proportion of the total variance in UA SWE than precipitation. This is also expected since the computation of snow ablation depends strictly on temperature data and the occurrence of ablation is stronger during these months. Nevertheless, the cumulative effect of snow accumulation throughout the water year is still evident via precipitation, albeit to a smaller degree, even in these summer months.

The rain-snow partitioning scheme, on the other hand, contributes very little variability to the algorithm's SWE estimates throughout the entire snow season. A possible explanation is that although the dynamic Jennings thresholds across the UCRB categorically exceed the fixed 2°C threshold (Figure 2.1), the difference in their partitioning of precipitation is minimal in this geographic region. The effect of the partitioning scheme does not exceed 4% on any day (Figure 2.4).

The interaction between precipitation and temperature, exhibited in the net snowfall computation, contributes some, but not substantially, to UA SWE variability.

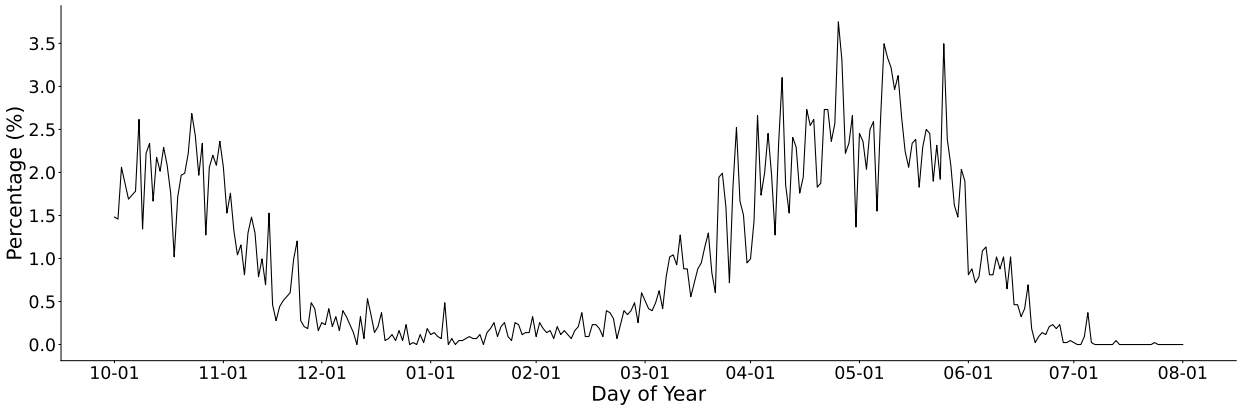


Figure 2.4: Percentage of sites and years that precipitation is partitioned differently by the default 2°C threshold and the dynamic Jennings threshold.

Including Interpolation

Including interpolation in the UA SWE algorithm allows us to assess the full algorithm on the grid cells, as the product is intended.

At grids corresponding to SNOTEL sites We first discuss results for grid cells containing SNOTEL sites. The spatial distribution of these grid cells is shown in Figure 2.5.

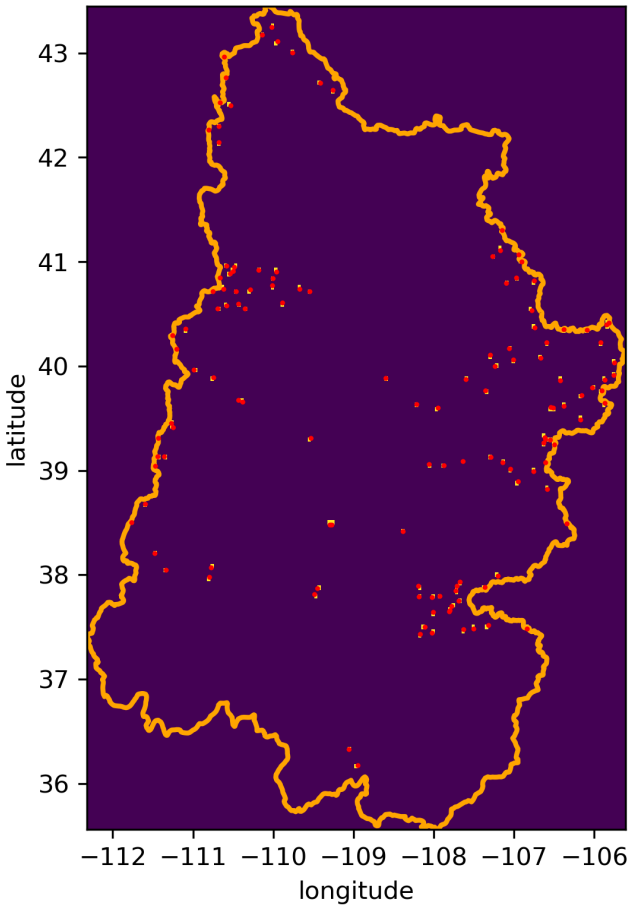
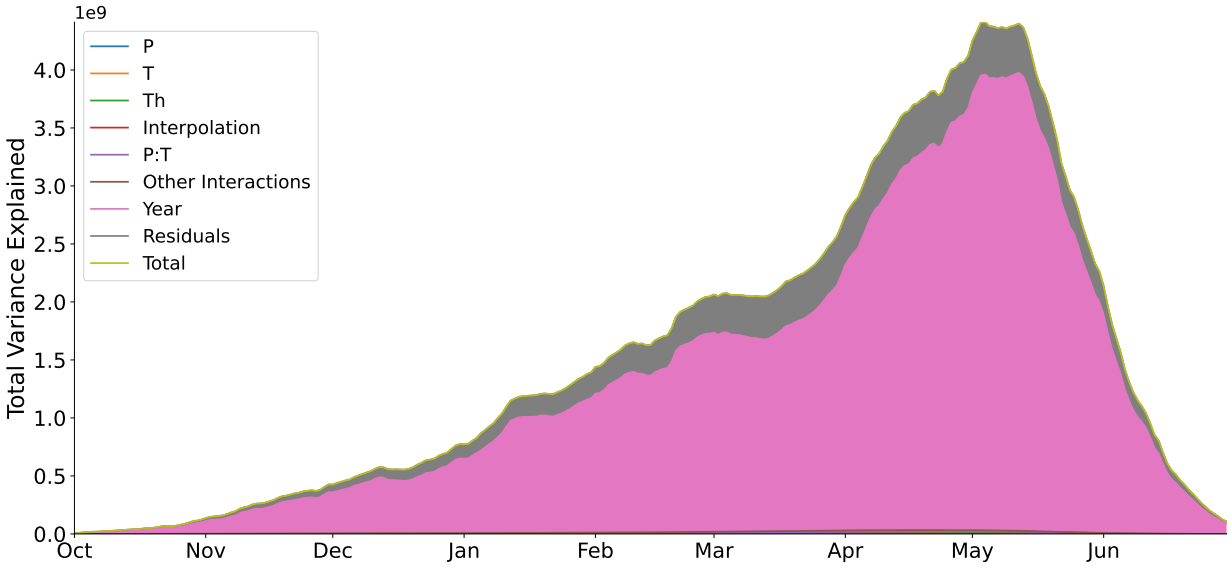
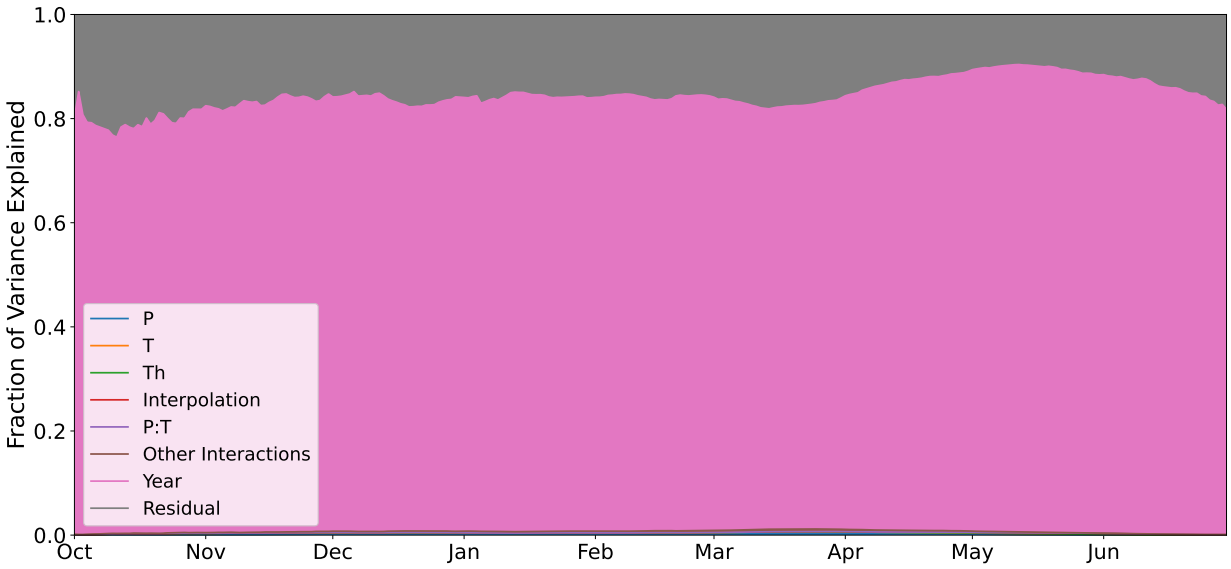


Figure 2.5: Distribution of grid cells in the UCRB corresponding to SNOTEL sites (dots or squares) for an arbitrary day (April 1, 2000).



(a) Total variance.



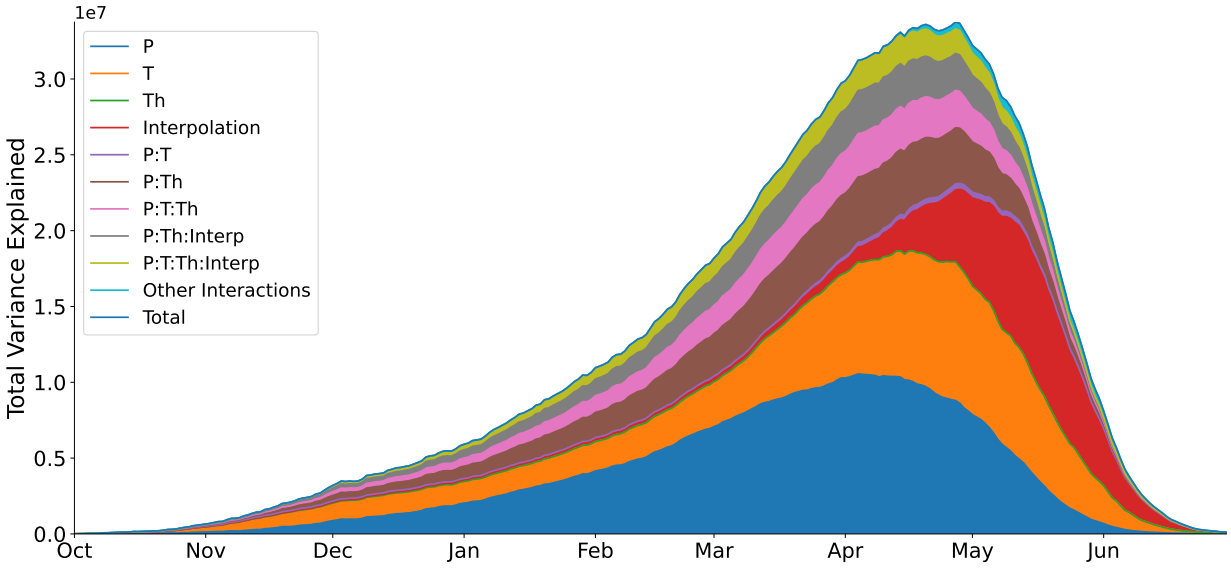
(b) Fractional variance.

Figure 2.6: Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at grid cells containing SNOTEL sites.

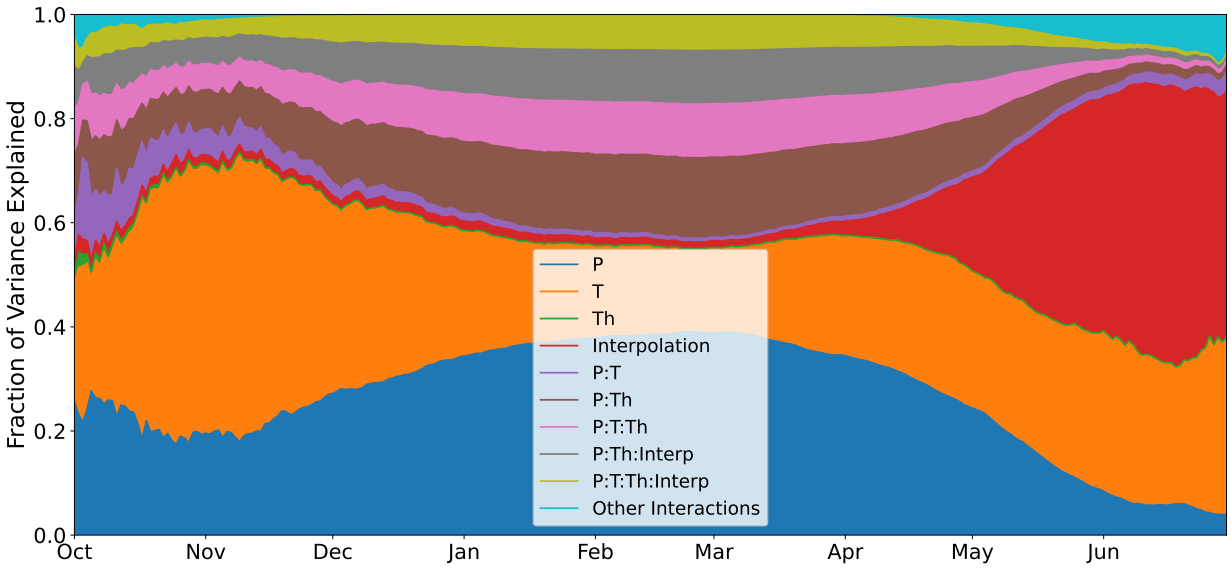
Including interpolation in the modeling process significantly increases the total variance in UA SWE, with the total variance being approximately 4.4×10^9 when interpolation is included, compared to approximately 1.1×10^9 when interpolation is ignored, indicating that interpolation amplifies the variability in UA SWE estimates at SNOTEL site locations

by approximately four times. Nevertheless, similar to results from the point-level analysis, Figure 2.6 shows that $> 98\%$ of the total variance in UA SWE is attributed to temporal and residual variability, with around 80% of the total variance attributed to year-to-year temporal variability, on average, and the remainder to temporal interactions.

To better assess the variance explained by the main effects of interest, we once again remove the temporal and residual effects in the illustration. This is shown in Figure 2.7.



(a) Total variance.



(b) Fractional variance.

Figure 2.7: Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at grid cells corresponding to SNOTEL sites. Temporal and residual effects are excluded for clarity.

From October through April, precipitation and temperature explain around 50%-70% of the variability in UA SWE. Precipitation explains more variability during the snowy season between December and April and temperature explains more variability during the bulk of

the ablation season after April.

While the main effect of interpolation contributes relatively little to the total variability in UA SWE in the accumulation season, its effect increases beginning in April and becomes dominant by June, during the main ablation season. This is likely due to inconsistencies in the timing of full ablation (i.e., zero SWE) across sites and years. Different sites will reach full ablation on different days during the ablation season, resulting in days where interpolation is conducted on a wider range of ratio values, contributing to heightened variability.

Similar to in the point-level results, the main effect of the rain-snow partitioning scheme explains a trivial amount of variability in UA SWE. This can similarly be attributed to the categorical exceedance of the Jennings dynamics thresholds to the fixed 2°C threshold (Figure 2.1) across the UCRB, where the partitioning outcomes differ in less than 4% of cell-years on any day (Figure 2.4). However, when interpolation is included in the UA SWE algorithm, interactions that involve precipitation and the rain-snow partitioning scheme (namely, P:Th, P:T:Th, P:Th:Interpolation, and P:T:Th:Interpolation) become key drivers in UA SWE variability. These interactive effects are especially prevalent during the accumulation season and begin to decrease as the ablation season approaches. This is in line with physical understanding because the computation of snow accumulation in the UA SWE algorithm is based on applying the thresholding scheme directly to precipitation data, whereas the computation of cumulative ablation does not rely on precipitation or the thresholding scheme, and snow accumulates very little during the ablation season. These are the most significant interaction terms out of all combinations of the main effects and are brought out by the incorporation of interpolation in the algorithm. Comparing the total variance plots in the point-level analysis to the present analysis (Figures 2.3a and 2.7a), we observe that the inclusion of interpolation leads to these interactive effects contributing as additional variance (rather than replacing any proportion of the total variance observed when interpolation is ignored).

While the interactions involving precipitation and the rain-snow partitioning scheme produce the most dominant interactive effects, the interaction between precipitation and temperature on UA SWE variability is present, but relatively small after December. All other interactive effects are considerably smaller, with their combined effect contributing up to 10% of the total variability in the ablation season and almost 0% during most of the accumulation season.

At grids in snow covered cells Here, we discuss results for snow covered grid cells only, as defined by the SSM. SSM is used to distinguish grid cells where seasonal (or more persistent) snow exists throughout the water year, rather than ephemeral (or short lived) snow. The map of seasonal snow covered cells in the UCRB is shown in Figure 2.8.

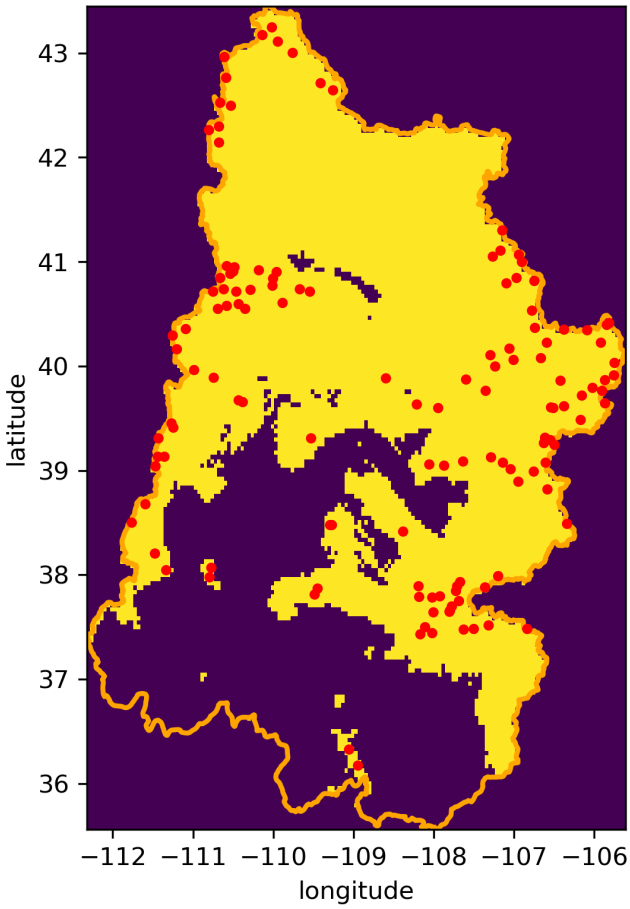
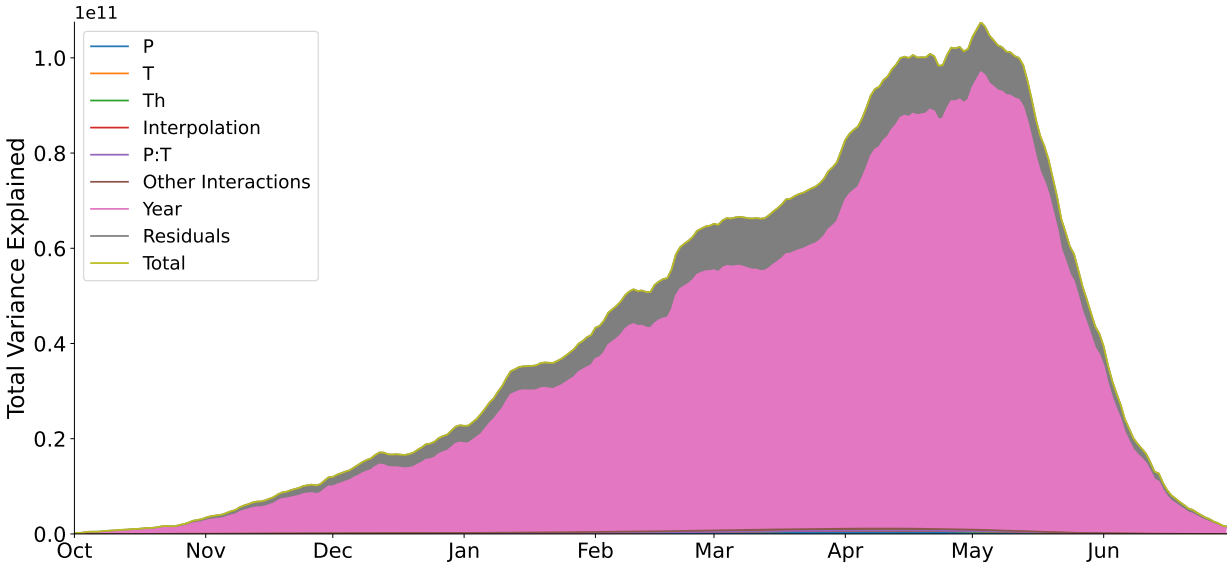
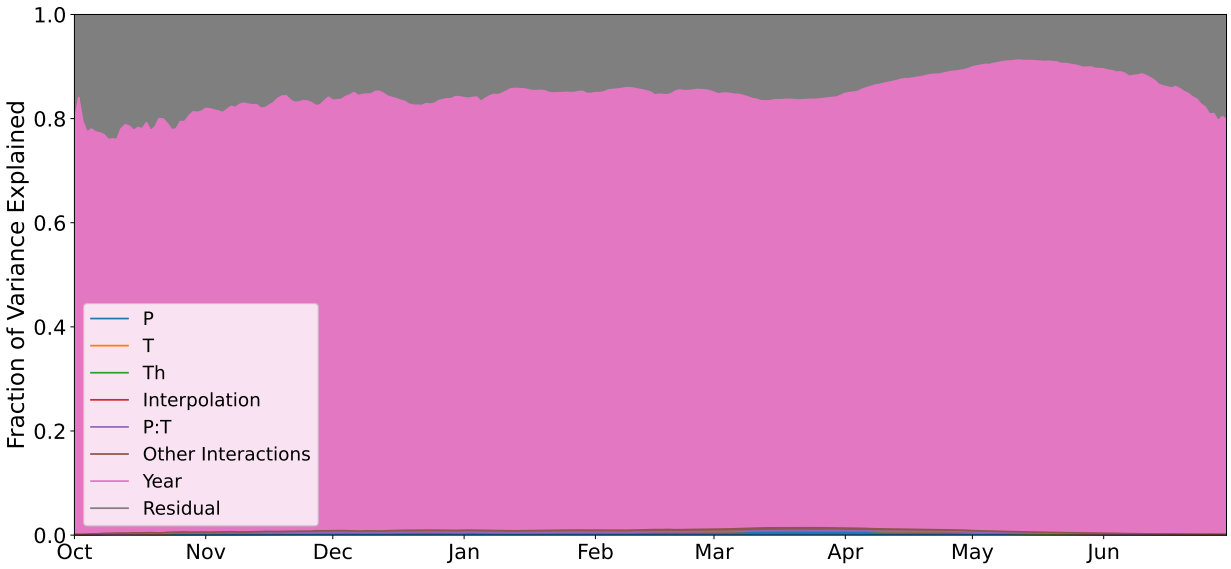


Figure 2.8: Distribution of snow covered grid cells in the UCRB, as defined by the SSM, for an arbitrary day (April 1, 2000). Snow covered grid cells are indicated in yellow. Red dots indicate SNOTEL sites.



(a) Total variance.

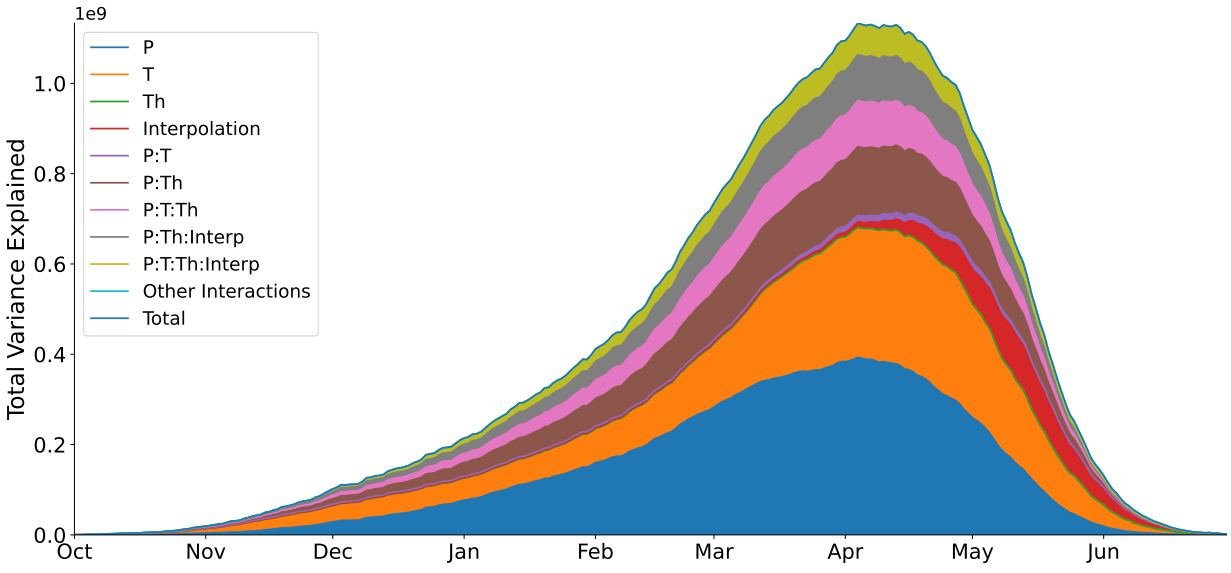


(b) Fractional variance.

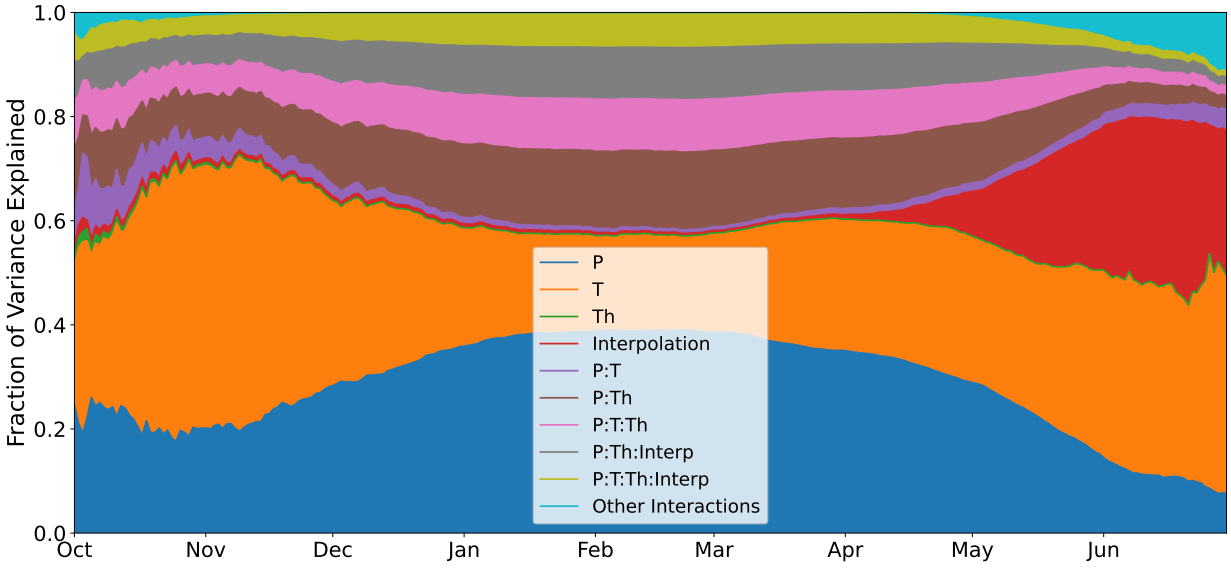
Figure 2.9: Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at snow covered grid cells.

As in the point-level analysis and the grid-level analysis at grid cells corresponding to site locations, the total variance in UA SWE is predominately attributed to temporal and residual variability ($> 98\%$), with around 80% of the total variance attributed to year-to-year temporal variability, on average, and the remainder to temporal interactions.

To better assess the variance explained by the main effects of interest, we once again remove the temporal and residual effects and focus on the main effects in the illustration in Figure 2.10.



(a) Total variance.



(b) Fractional variance.

Figure 2.10: Aggregated cell-specific ANOVAs of fractional and total variance explained by each factor on each day at snow covered grid cells. Temporal and residual effects excluded for clarity.

The partitioning of variance computed across all snow covered grid cells behaves very similarly to the partitioning of variance computed only at grid cells containing SNOTEL sites. Specifically, at snow covered grid cells, precipitation and temperature continue to explain a significant amount of total variability in UA SWE, the interactive effects containing precipitation and the rain-snow partitioning scheme continue to be significant, and the non-interactive effect of the rain-snow partitioning scheme is negligible. While interpolation also plays an important role in explaining UA SWE variability during the ablation season, its fractional effect is smaller than that seen at grid cells corresponding to site locations.

Error Decomposition

Ignoring Interpolation (Point-Level)

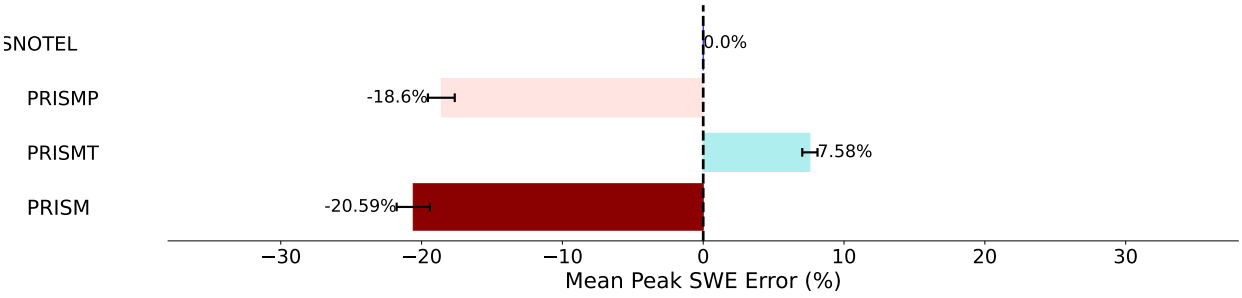


Figure 2.11: Error decomposition of UA SWE components, when interpolation is ignored in the algorithm. Solid colored bars indicate the average of the δ_C errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference (SNOTEL SWE in this case). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the error of the reference (SNOTEL SWE), which is 0% by definition.

In Figure 2.11, the first row labeled "SNOTEL" refers to utilizing SNOTEL P and T in place of PRISM P and T in the PRISM net snowfall component of the UA SWE algorithm when ignoring interpolation. By definition, there is zero error. The next two rows "PRISMP" and "PRISMT" are a decomposition to quantify the effect of each modeling component that make up UA SWE. "PRISMP" and "PRISMT" refer to using PRISM P and PRISM T, respectively, to compute the PRISM net snowfall component, keeping SNOTEL as the data source for T and P, respectively. These represent the individual effects of PRISM P and PRISM T on SWE estimation via the UA SWE algorithm. The amalgamation of these modeling decisions is captured in the last row, "PRISM", which is exactly the UA SWE algorithm ignoring interpolation. 95% confidence intervals around the mean errors of each

row are included. The computation of standard errors in the confidence intervals assumes independence. This is an approximation because repeated samples at each site introduces dependence.

The utilization of PRISM P in the UA SWE algorithm produces negative error, while the utilization of PRISM T in the UA SWE algorithm produces positive error, although at a lower magnitude, on average. Using both PRISM P and T together leads to UA SWE underestimation compared to SNOTEL SWE at SNOTEL sites. Notably, the magnitude of underestimation is larger than the negative error exhibited by only using PRISM P. These behaviors indicate that the input data used for the gridded net snowfall component (i.e., the PRISM net snowfall component) is a catalyst for biasing SWE estimates away from SNOTEL SWE. This may be attributed to differences in the precipitation and temperature series from gridded products compared to SNOTEL. In particular, the discrepancy between SNOTEL and PRISM precipitation starting in November, as seen in Figure 2.13, drives PRISM underestimation of accumulated snow, contributing to its underestimation of SWE at peak SWE time.

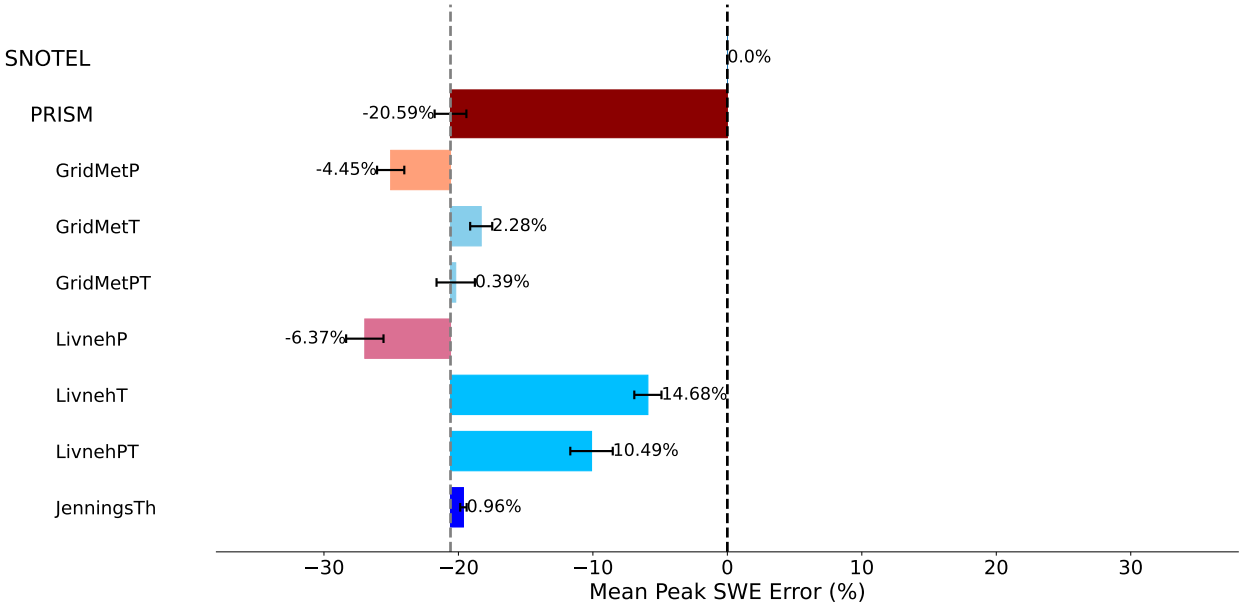


Figure 2.12: Error decomposition of alternative UA SWE modeling decisions, when interpolation is ignored in the algorithm. Solid colored bars indicate the average of the δ_2 and δ_3 errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference ("SNOTEL" and "PRISM" for δ_2 and δ_3, respectively). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the mean SNOTEL SWE error (0% by definition) and the grey dashed vertical line marks the mean error of "PRISM" (i.e., UA SWE ignoring interpolation).

The rows in Figure 2.12 capture the errors associated with alternative modeling decisions that could have been made in the generation of UA SWE. Substituting GridMET P or Livneh P for PRISM P in the UA SWE algorithm adds negative error to the UA SWE estimates, on average. Since the generation of both GridMET P and Livneh P depend on PRISM P and their average time series (across all sites and years) are comparable (Figure 2.13), the difference in errors exhibited by using GridMET P or Livneh P is likely a result of the scaling of PRISM P underlying the generation of GridMET P and Livneh P.

On the other hand, substituting GridMET T or Livneh T for PRISM temperature in the UA SWE algorithm reduces the SWE error, offsetting some of the error exhibited by UA SWE. On average, GridMET T corrects for some of the UA SWE error. The small differences in errors on UA SWE produced by using GridMET T over the default setting is likely a result of the generation procedure of GridMET T, which is a scaling, or adjustment, of PRISM T. On the other hand, using Livneh T in place of PRISM T in the UA SWE algorithm offsets the error considerably. This may be explained by Livneh generally predicting lower

temperature than PRISM, especially during the snow season, as seen in Figure 2.14. Lower temperature predictions from Livneh leads to lower cumulative degree days, resulting in lower predicted cumulative ablation and therefore higher predicted SWE, explaining the reduction in SWE error when Livneh T is used, when compared to SNOTEL SWE. It is notable that Livneh T underestimates SNOTEL T, which is similar to PRISM T and Gridmet T. Livneh T is generated based on the GHCN temperature network and is completely independent of SNOTEL T, unlike PRISM T and GridMET T, which incorporate SNOTEL T into their data generation procedure. Some studies have suggested a warm bias in SNOTEL T measurements in colder conditions, in part due to lack of homogenization of data collected by old and new sensors [54, 42]. The inclusion of a lapse rate correction in Livneh T may also explain its underestimation compared to SNOTEL T.

For GridMET, the interaction between the P and T products shows some evidence of offsetting errors in the UA SWE estimates. Given that the confidence interval around the mean incremental error with respect to PRISM spans both negative and positive values, there is substantial uncertainty about whether using GridMET P and T leads to fewer or more erroneous SWE estimates compared to using PRISM P and T, on average. Meanwhile, for Livneh, the interaction between the P and T products consistently offsets approximately half of the errors underlying UA SWE estimates.

Finally, employing the dynamic Jennings temperature threshold instead of the fixed 2°C threshold for partitioning rain and snow in the estimation of snow accumulation has a trivial effect on UA SWE prediction at SNOTEL sites. This is explained by Figure 2.4, which shows that the partitioning of rain and snow by Jennings compared to the fixed temperature threshold differs at less than 3% of sites and years before and on April 1. Thus, at the majority of stations in most years, the fixed and Jennings thresholds partition precipitation into the same category of rain and snow. As a result, the fixed threshold introduces minimal uncertainty. This result might be specific to the UCRB and warrants further research in other mountainous regions.

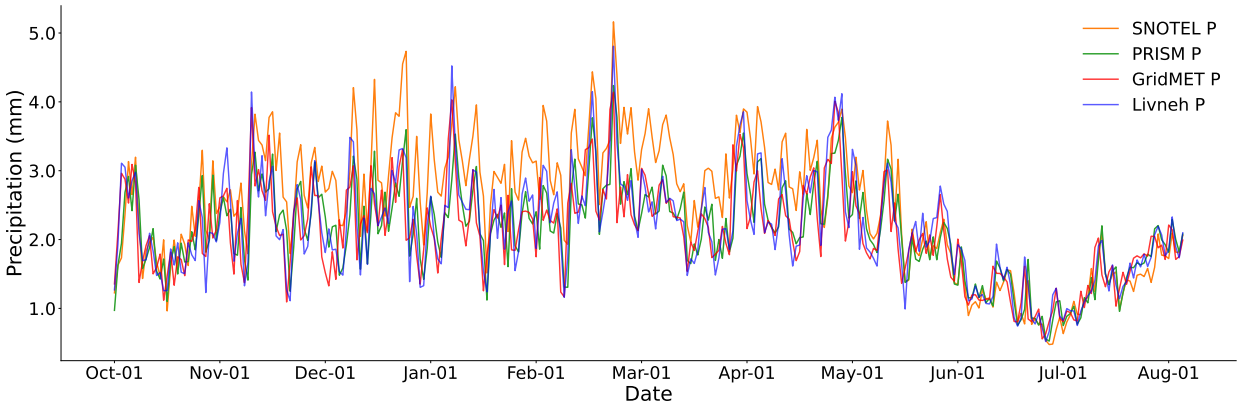


Figure 2.13: Average precipitation across sites and years on each day from the SNOTEL, PRISM, GridMET, and Livneh data sources.

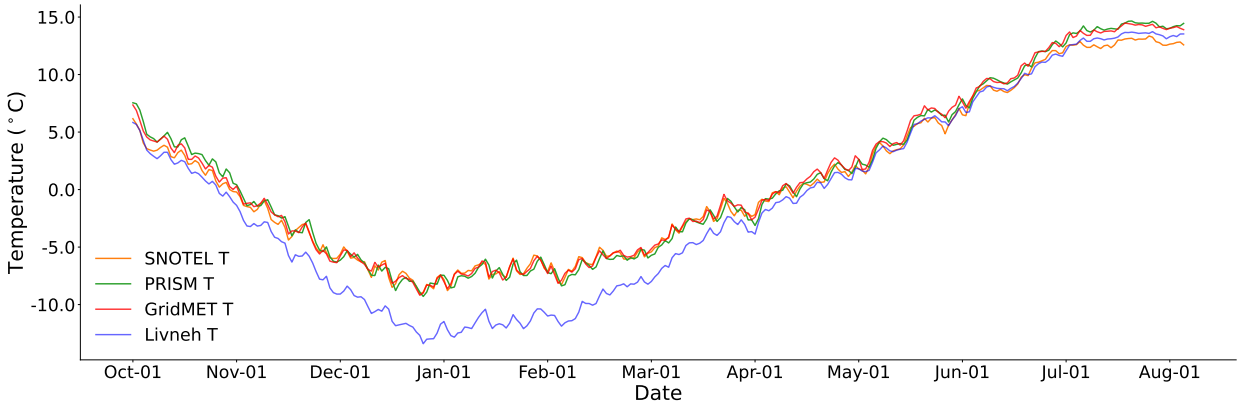


Figure 2.14: Average temperature across sites and years on each day from the SNOTEL, PRISM, GridMET, and Livneh data sources.

Including Interpolation

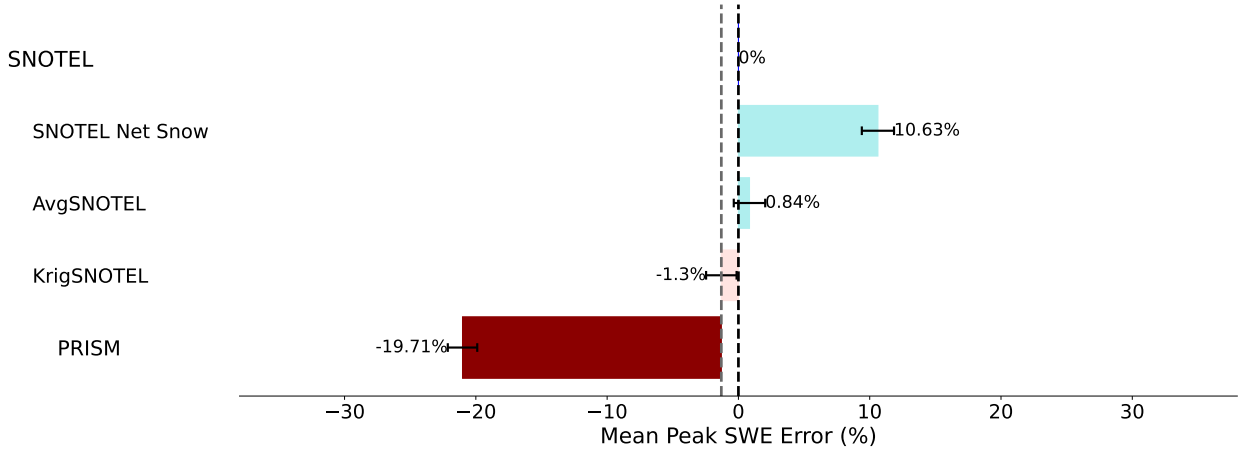


Figure 2.15: Error decomposition of UA SWE components, when interpolation is included in the algorithm. Solid colored bars indicate the average of the δ_C errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference (SNOTEL SWE and subsequently "krigSNOTEL"). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the error of the SNOTEL SWE reference, which is 0% by definition and the grey dashed vertical line marks the error of the "krigSNOTEL" reference.

In Figure 2.15, the first row "SNOTEL" refers to using SNOTEL precipitation and temperature in place of PRISM precipitation and temperature at the SNOTEL site-level in the UA SWE algorithm (which includes interpolation by default). By definition, there is zero error. The fifth row "PRISM" refers to the UA SWE algorithm with its default configuration.

The sequence of "SNOTEL Net Snow", "AvgSNOTEL", and "KrigSNOTEL" experiments (corresponding to the second, third, and fourth rows) systematically decompose and quantify the errors produced by the interpolation in the UA SWE algorithm. In all three steps, the PRISM net snowfall component of the UA SWE algorithm is computed using SNOTEL precipitation and temperature rather than PRISM. More specifically, since net snowfall is treated as a proxy for SWE, "SNOTEL Net Snow" quantifies the difference between SNOTEL net snowfall and SNOTEL SWE. In other words, it quantifies the error of UA SWE estimates if the SNOTEL ratio is constant at 1.0 across all dates and sites. "AvgSNOTEL" quantifies the error of UA SWE estimates if the SNOTEL ratio is fixed at a more informed constant value (rather than assuming 1.0). The average SNOTEL ratio across all dates and sites is used. "KrigSNOTEL" quantifies the error of UA SWE estimates if a more sophisticated interpolation scheme (i.e., OK, where the SNOTEL ratio varies in space and time) is used to scale net snowfall.

SNOTEL net snowfall (i.e., assuming a net snowfall scaling factor of 1.0) overestimates SNOTEL SWE. Using a constant average SNOTEL ratio to scale net snowfall in the UA SWE algorithm reduces this error substantially, resulting in minimal difference when compared to SNOTEL SWE at the SNOTEL sites, on average. The confidence interval spans the mean error of SNOTEL SWE (i.e., 0%), indicating uncertainty about whether there is significant difference from SNOTEL SWE. Applying a more intricate interpolation scheme, such as kriging, to the SNOTEL ratio to scale net snowfall reduces the average error in net snowfall significantly and consistently. The small negative mean error and its associated confidence interval not containing the baseline SNOTEL SWE mean error provides evidence of a significant (albeit small) difference when kriging is used.

The fifth row, "PRISM," is exactly UA SWE, which exhibits negative error when compared to SNOTEL SWE at SNOTEL sites. The incorporation of PRISM data in the UA SWE algorithm results in a significant underestimation of UA SWE relative to SNOTEL SWE. The small errors contributed by smoothing and the ablation modeling, in conjunction with large negative errors presented when using PRISM P and/or T in the estimation of net snowfall, suggest that the mathematical formulation of the UA SWE algorithm is generally sound and reinforces the predominant effect of input data in inducing bias in UA SWE estimates, at the very minimum at SNOTEL sites in the UCRB.

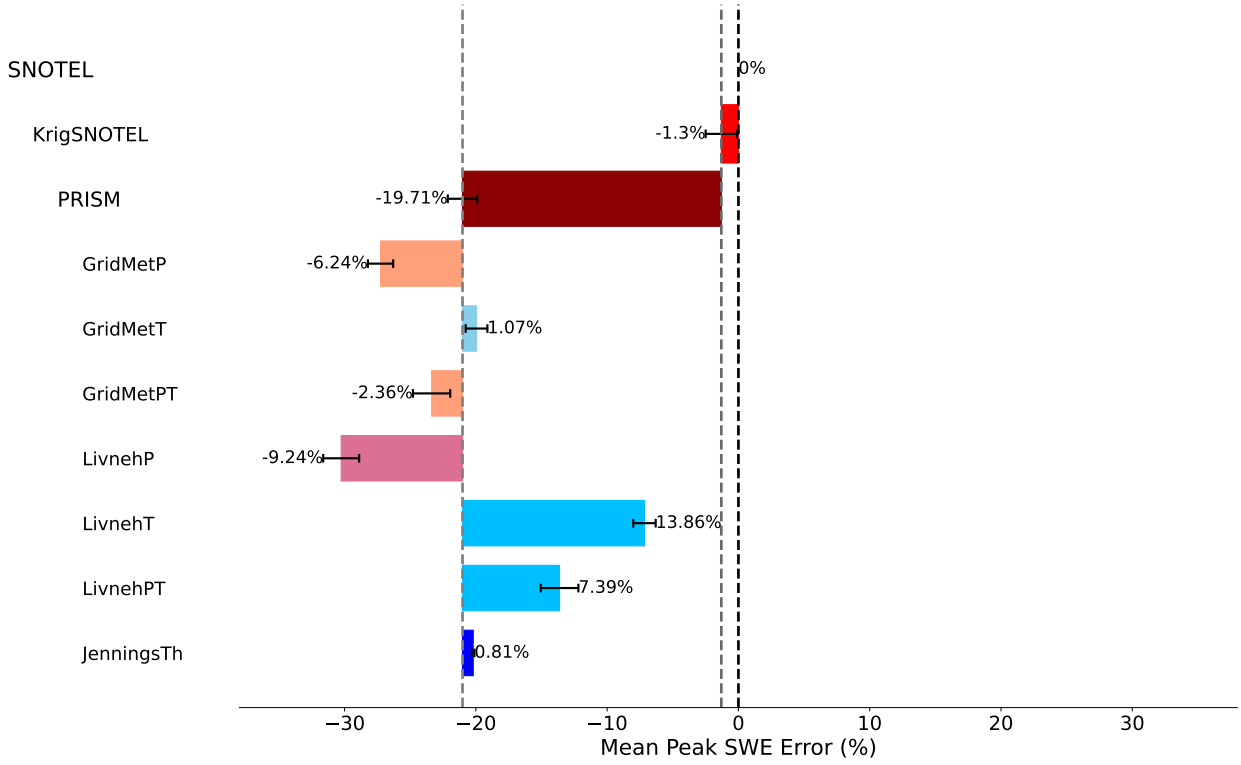


Figure 2.16: Error decomposition of alternative UA SWE modeling decisions, when interpolation is included in the algorithm. Solid colored bars indicate the average of the δ_2, δ_3, and δ_4 errors on April 1, computed across sites and years. Bars with red hue signify underestimation and bars with blue hue signify overestimation relative to the reference ("SNOTEL", "krigSNOTEL", and "PRISM" for δ_2, δ_3, and δ_4, respectively). The horizontal black error bars represent 95% confidence intervals (mean $\pm$ 1.96 $\times$ standard error) with respect to the reference. The black dashed vertical line marks the mean SNOTEL SWE error (0% by definition) and the grey dashed vertical lines mark the mean error of "krigSNOTEL" and "PRISM" (i.e., UA SWE).

The rows in Figure 2.16 capture the errors associated with alternative modeling decisions that could have been made in the generation of UA SWE. Substituting GridMET and Livneh precipitation for PRISM precipitation in the UA SWE algorithm adds negative error to the UA SWE estimates, on average. On the other hand, on average, substituting GridMET and Livneh temperature for PRISM temperature in the UA SWE algorithm offsets some of the errors in the UA SWE estimates. In particular, using the Livneh temperature series in place of PRISM temperature in the UA SWE algorithm offsets the error significantly more frequently. While using GridMET or Livneh precipitation in place of PRISM precipitation produces similar amounts of added negative error to UA SWE, using Livneh temperature in place of PRISM temperature produces significantly larger positive error (i.e., reduces

overall UA SWE error) than using GridMET temperature in place of PRISM temperature. The dynamic Jennings rain-snow partitioning threshold in place of the default fixed 2°C temperature threshold once again exhibited very minor effects on UA SWE estimates.

Overall, with the exception of GridMET P and T whose interaction consistently adds negative error on top of UA SWE (as evidenced by its confidence interval that is entirely below the mean error of the UA SWE reference), the incremental errors and associated confidence intervals in this analysis including interpolation are similar to those found in the analysis where interpolation was ignored (Figure 2.12). This agreement suggests that interpolation has a minor effect on differences between predicted SWE and SNOTEL SWE.

2.5 Conclusion

In this study, we proposed a suite of methods for identifying and quantifying compensating errors in gridded datasets that can be easily concealed when conducting bulk error analyses. We applied our proposed sensitivity analysis and error decomposition approaches to understand variability and errors underlying the UA SWE gridded product.

Our results suggest that precipitation and temperature are the most significant drivers influencing variability and bias in the UA SWE algorithm. The impact of precipitation and temperature lies in both the natural variability of these variables as well as in differences between datasets providing estimates for these variables. We found that PRISM (and GridMET and Livneh) underestimation of SNOTEL precipitation propagates errors into UA SWE estimation. This is because SWE estimation depends on precipitation and it therefore inevitably inherits some of the errors in the underlying precipitation input data. In terms of temperature, we found that the Livneh product, whose generation procedure, unlike PRISM and GridMET, involves a lapse rate correction and does not depend on SNOTEL temperature data, offsets negative errors produced by the default UA SWE formulation that depends on PRISM substantially more than GridMET. While this seems to suggest that using the Livneh temperature product in place of PRISM temperature is favorable for reducing UA SWE error, Livneh temperature diverges considerably from SNOTEL temperature during the snow season. Thus, the promising performance of Livneh temperature on UA SWE prediction may either be misleading due to inaccurate Livneh predictions that happen to cancel out negative errors in UA SWE or it could be true, in which case reported biases in SNOTEL temperature records are a concern for producing gridded datasets, such as PRISM and GridMET, that integrate SNOTEL temperature observations into their data generation process. In order to draw definite conclusions on which input datasets to employ in algorithms such as UA SWE, a thorough analysis of observed and gridded input datasets, for variables such as precipitation and temperature, as well as model assumptions, is recommended. Furthermore, careful consideration is required for generalizing findings presented here to other or larger geographic regions, as results may differ in other areas.

By quantifying uncertainty contributed by model components, our analysis supports improved SWE model development and helps guide decisions on areas that require additional

attention.

Chapter 3

Advancing SWE Prediction through Statistical Modeling

3.1 Introduction

While gridded products such as UA SWE provide significant utility to the community, assumptions underlying the generation of most gridded SWE products have not been sufficiently investigated. In this chapter, we closely examine methodological decisions, especially around the prediction of snow ablation, in the UA SWE algorithm and introduce a novel approach for predicting SWE.

At present, UA SWE estimates snow ablation on a cumulative basis depending strictly on average air temperature. However, snow ablation, the process of snowpack reduction due to melting, sublimation, and evaporation, is complex and influenced by a multitude of factors beyond air temperature, including energy balance dynamics (e.g., longwave radiation, sensible heat flux, and latent heat flux), meteorological conditions (e.g., wind speed, humidity, and solar radiation), terrain effects (e.g., slope orientation, shading, and local topography), and snowpack properties (e.g., albedo, density, and grain size) [58, 34].

Following an in-depth discussion on the limitations of the present ablation modeling approach employed by the UA SWE product development team, we propose an alternative data-driven approach for modeling ablation. Our proposed approach models ablation on the daily rather than cumulative scale and takes advantage of a wide range of relevant physical variables to improve prediction accuracy. We further extend our methodology to produce a second empirical model that predicts daily SWE directly. Our versatile method can be easily applied to predict SWE at the point-level as well as extended to generate gridded fields of SWE across geographical areas of interest.

Our daily SWE modeling approach shares some similarities to a recently published paper [76], which models daily change in SWE using a support vector regression (SVR). In addition to differences in covariate sets employed by ours and their method, our daily SWE model makes no assumptions about the snow state while their approach incorporates various manual

checks (as does our daily ablation modeling approach). For example, they only uses the pretrained daily SWE SVR to predict change in SWE on warm days that experience average temperature exceeding $-1°C$; on all other days (i.e., cold days), the change is SWE is simply taken as the precipitation on that day.

This chapter is structured as follows. Section 3.2 outlines the limitations associated with the UA SWE ablation estimation approach. Section 3.3 details our proposed ablation modeling approach, including the data and covariates used. A full assessment of the performance of our proposed ablation modeling approaches in estimating ablation and SWE is provided in Section 3.4. The implications of our findings are discussed and concluding remarks are made in Section 3.5.

3.2 Datasets

Datasets Used

A list of SWE datasets was outlined and described in Sections 1.3 and 1.5. This chapter leverages SNOTEL SWE and UA SWE.

Limitation in UA SWE

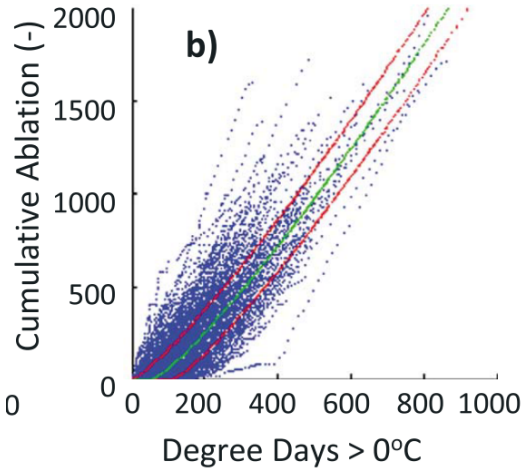


Figure 3.1: The UA SWE ablation regression as depicted in Figure 2b of [10]. Each blue dot represents a day at a SNOTEL site located in a $10° \times 10°$ longitude-latitude box in the WUS during the 2004-2007 water years. The green line is the best fit regression line that is used to estimate cumulative ablation on each day for both SNOTEL and PRISM in the net snowfall components in the UA SWE algorithm.

Figure 3.1 illustrates the regression used to estimate ablation in the net snowfall component in the UA SWE algorithm, taken directly from [10]. The values are taken from SNOTEL sites within a $10^\circ \times 10^\circ$ longitude-latitude box in a region of the WUS over the 2004-2007 water years [10]. The quadratic regression utilizes cumulative degree days > 0°C to predict cumulative ablation. The cumulative degree days variable is the total degrees across days in a water year up to the target date for which temperature > 0°C at the SNOTEL sites. The outcome variable (cumulative ablation) is defined as the difference between accumulated snowfall and SWE at the SNOTEL sites.

As is apparent in Figure 3.1, there are large deviations in ablation quantities from the regression line, implying large errors in the model predicted ablation values. Moreover, ablation is modeled based solely on average temperature data, through degree days. We hypothesize that including information from different moments of the diurnal cycle of air temperature (e.g., minimum and maximum) and additional physical variables could improve the model's ability to predict ablation. However, it is not straightforward to include physical variables in the modeling pipeline when it is on a cumulative scale.

Assumptions made in the UA SWE algorithm also raise some concerns. First, the use of a rain-snow partitioning threshold leads to some erroneous categorization of precipitation as rain or snow. Decisions on rain-snow partitioning thresholds are often made in modeling exercises, however, studies have shown that the chosen threshold does not always yield accurate predictions of rain versus snow, especially when applied over larger regions [29]. The sensitivity of the estimation of snow accumulation to a fixed rain-snow partitioning threshold is illustrated empirically in Figure 3.2. The deviation in climatological snow accumulation across all days derived from five different partitioning thresholds, ranging from -4°C to +4°C, demonstrates the strong influence the selected threshold value has on estimating snow accumulation, and therefore SWE. Second, the ablation model assumes that ablation never occurs on cold days (days where the average temperature falls below 0°C), even though it is physically possible through processes such as sublimation. Third, the effects of rainfall on persisting snow on the ground, or rain-on-snow, on both snow accumulation and ablation are not able to be explicitly and fully captured in their approach.

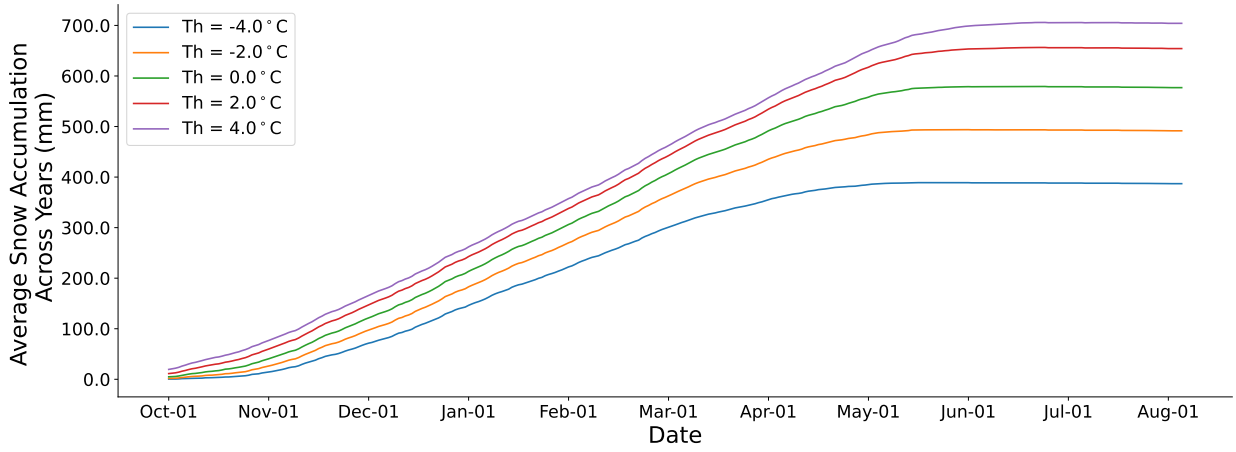


Figure 3.2: Climatological snow accumulation derived from five different fixed rain-snow partitioning thresholds, from -4°C to +4°C. The snow accumulation quantity on each day is the average snow accumulation across all sites and years in the study area and period.

For these reasons, we seek an alternative ablation modeling approach to estimating snow ablation that uses daily ablation as the target variable, rather than cumulative ablation.

3.3 Methods

In this section, we describe our proposed daily modeling approaches, the derivation of cumulative SWE across water years from our daily model predictions, and the validation schemes we employ.

Figure 3.3 provides a high-level overview of our overall approach. After collecting and cleaning all relevant data, we train two daily models: (1) a daily ablation model whose ablation predictions are used to derive SWE; (2) a daily SWE model. Feature importance experiments are carried out during the initial model training stage under the daily ablation model, in order to narrow down the set of desired input variables to use for the remainder of the study. Once both the daily ablation and daily SWE models are trained with the desired covariate set, the daily model predictions are combined mathematically to generate continuous predictions on the cumulative scale (time series), for SWE as well as intermediate variables (such as cumulative ablation) when applicable. After the SWE time series are validated, a best performing model is selected and applied to gridded input data to produce a gridded field of SWE.

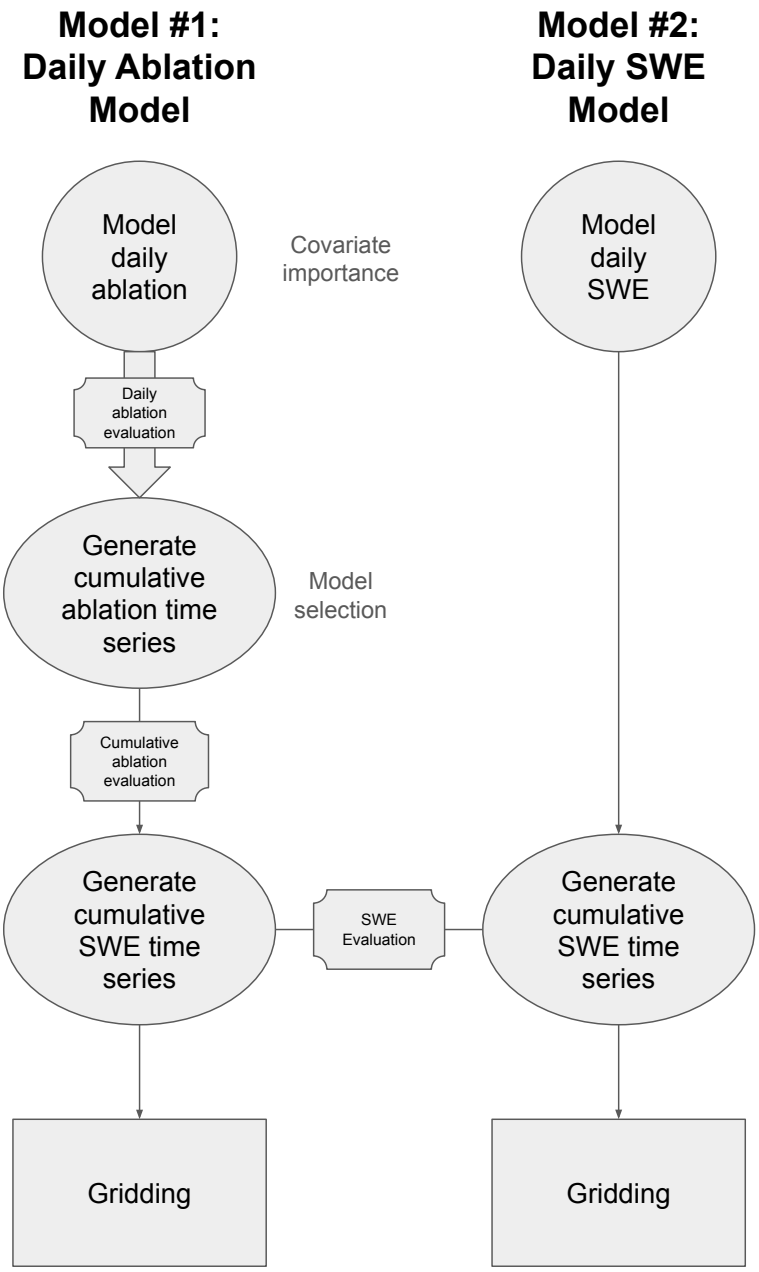


Figure 3.3: Overview of study workflow, including model training, evaluation, and application.

Modeling Approaches

We investigate two modeling approaches based on the daily scale:

1. **Daily ablation modeling**, where we train models to predict daily ablation, then predict SWE on each day in a water year by combining independently estimated snowfall and ablation components. This approach follows the intuition underlying the UA SWE algorithm.

2. **Daily SWE modeling**, where we predict the daily change in SWE and then accumulate the predicted SWE across days for each water year. This approach eliminates computational intricacies involved in balancing the snow accumulation and snow ablation components, as well as any assumptions on an optimal rain-snow partitioning threshold.

In the next two subsections, we discuss these two approaches in greater detail.

Modeling Daily Snow Ablation

This approach predicts SWE by estimating snow accumulation and ablation separately and then taking their difference, as in the UA SWE algorithm. Our focus here is on estimating the snow ablation component via daily ablation modeling.

Following the notation introduced in Section 1.5 and defining the true daily ablation (A_t) on each day as:

$$A_t = S_t - (\text{SWE}_t - \text{SWE}_{t-1}) \tag{3.1}$$

where snowfall S_t is the total precipitation observed on day t if mean temperature falls below the 2°C rain-snow partitioning threshold and zero if mean temperature exceeds 2°C. $(\text{SWE}_t - \text{SWE}_{t-1})$ is the change in SWE from day $t-1$ to day t.

True daily ablation is defined as above, rather than simply the negative change in SWE on each day (i.e., $-(\text{SWE}_t - \text{SWE}_{t-1})$), for several reasons. First, UA SWE defines true cumulative ablation on each day t as the difference between accumulated snow and cumulative SWE. The daily equivalent on each day t is therefore the difference between snowfall and the change in SWE, which is exactly Eq. (3.1). Second, there exist discrepancies between precipitation and SWE quantities recorded at SNOTEL sites on cold days. This phenomenon is clear when plotting the distribution of differences between observed snowfall (precipitation on cold days) and change in SWE (Figure 3.4) — there is a non-negligible number of cold days on which a discrepancy between precipitation and change in SWE is observed. These differences in the records are largely attributable to instrumental error at the sites (precipitation and SWE data are collected by precipiation gauges and snow pillows, respectively) and since it is unclear whether to interpret them as ablation, defining daily ablation the way we have here mitigates this challenge and eliminates erroneous estimates caused by instrumental differences. Finally, this definition also offers the added benefit of allowing ablation to occur on snow that falls on the same day.

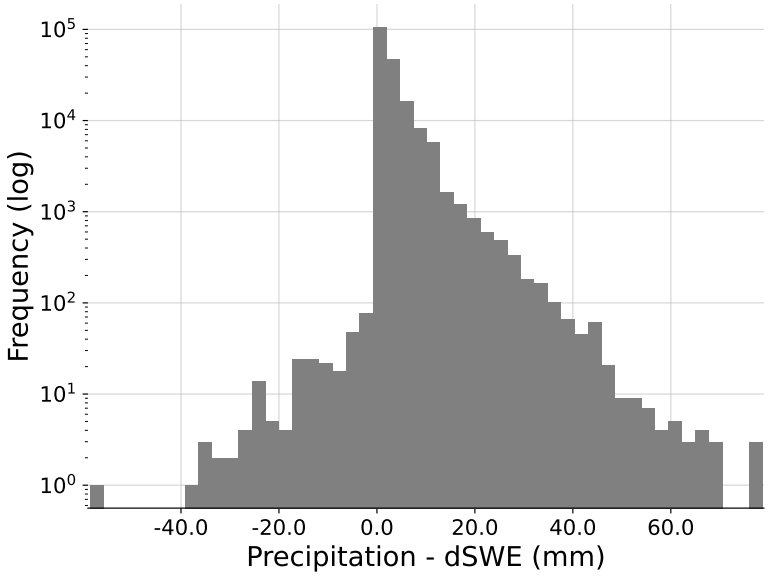


Figure 3.4: Distribution of non-zero differences between precipitation on cold days (snowfall) and change in SWE (dSWE).

We model daily ablation as a function of relevant climate variables (Section 3.3) on days where ablation may occur. A_t in Eq. (3.1) is the outcome variable used in the model training, as well as the ground truth against which daily ablation predictions from our trained models are validated. Given ablation defined as A_t in Eq. (3.1), we determined our training datasets as follows.

We consider three types of daily ablation models, distinguished by differences in the preprocessing of their training set and by the days on which our models are used to predict possible ablation:

1. **UA SWE-like daily ablation model**: This model setup aligns exactly with the intuition underlying UA SWE. According to the UA SWE algorithm, ablation on a day can only occur when temperature is greater than zero (cumulative degree days is based on degree days exceeding 0°C), regardless of the amount of precipitation. So, we allow the model to predict daily ablation only on days where T $>$ 0°C, as long as there is already snow present on the ground. All other days are filtered out of the training set and assumed to experience zero ablation when the model-predicted daily ablation quantities are accumulated to produce cumulative ablation time series and cumulative SWE time series.

2. **Physically-constrained daily ablation model**: This model is constrained based on physical intuition. Physically, on days where snow is already present, we expect SWE

may decrease because of ablation on days with high temperature or on days where no snow falls (through melt or sublimation). Based on this reasoning, we allow our model to predict daily ablation on days with snow on the ground, either when temperature exceeds the rain-snow threshold (melt) or when temperature falls below the rain-snow partitioning threshold and no precipitation occurs (sublimation). All other days are filtered out of the training set and assumed to experience zero ablation when the model predicted daily ablation quantities are accumulated to produce cumulative ablation time series and when generating cumulative SWE time series.

3. **Full daily ablation model**: This model learns and predicts daily ablation on all days except when no snow is present on the ground and no precipitation occurs (as there is nothing to ablate). The idea is that the model will determine on its own, based on the information we provide it, whether a day should or should not experience ablation. Unlike in the configuration of the physically-constrained daily ablation model, the configuration of the full daily ablation model allows ablation to occur on days that experience snowfall, which corresponds to days where there is snow falling for only part of the day.

A tabular summary of these three model configurations is included in Table B.1.

Predicting SWE from Modeled Daily Ablation From our model-predicted daily ablation ($\hat{A}_t$), cumulative ablation ($\hat{A}_{1:t}$) can be computed in a straightforward manner by summing across the water year:

$$\hat{A}_{1:t} = \hat{A}_{1:t-1} + \hat{A}_t \tag{3.2}$$

Continuous estimates of cumulative SWE can also be derived across an entire water year following a straightforward procedure similar to that of the UA SWE algorithm:

$$\widehat{\text{SWE}}_t = \widehat{\text{SWE}}_{t-1} + (\hat{S}_t - \hat{A}_t) \tag{3.3}$$

where $\widehat{\text{SWE}}_{t-1}$ is the cumulative SWE predicted from the previous day, snowfall $\hat{S}_t$ is the total precipitation on day t if temperature falls below the 2°C rain-snow partitioning threshold and zero otherwise, and $\hat{A}_t$ is daily ablation predicted directly by our model.

Modeling Daily SWE

This approach models SWE directly, bypassing the need for any arbitrary decision-making in the modeling process.

We define the true daily SWE on each day as the change in SWE (ΔSWE) from the previous day:

$$\Delta\text{SWE}_t = \text{SWE}_t - \text{SWE}_{t-1} \tag{3.4}$$

Just like in the daily ablation model, we model daily SWE as a function of relevant climate covariates on days where we expect SWE may be present. We utilize the same set of

climate covariates as inputs to the daily SWE (dSWE) model, except ΔSWE_t in Eq. (3.4) is the target variable used for model training and evaluation in place of A_t in Eq. (3.1).

When there is no snow already on the ground (i.e., when $\text{SWE}_{t-1} = 0$ mm) and no precipitation recorded (i.e., $\text{P}_t = 0$ mm), we expect there to be no change in SWE (i.e., $\Delta\text{SWE}_t = 0$ mm). Under all other conditions, SWE has the potential to increase or decrease. As a result, we allow the dSWE model to learn and predict on all days except those that experience zero SWE and zero precipitation. These days are filtered out of the training set and assumed to be zero when accumulating the daily SWE quantities.

Predicting SWE from Modeled Daily SWE A continuous time series of SWE over a water year can be easily derived from our dSWE model by accumulating predicted daily SWE across time:

$$\widehat{\text{SWE}}_t = \widehat{\text{SWE}}_{t-1} + \widehat{\Delta\text{SWE}}_t \tag{3.5}$$

where $\widehat{\Delta\text{SWE}}_t$ is the change in SWE prediction on day t taken directly from the dSWE model and $\widehat{\text{SWE}}_t$ and $\widehat{\text{SWE}}_{t-1}$ are cumulative SWE quantities on day t and day $t-1$, respectively.

Modeling Decisions

Input Variables

UA SWE depends solely on average temperature to predict snow ablation, and solely on precipitation and average temperature to predict SWE. Precipitation also directly affects melt rates through processes such as rain-on-snow [39]. As part of this study, we investigate whether the inclusion of different moments of daily air temperature (e.g., minimum and maximum) and additional variables including precipitation itself improves the predictive performance of an ablation model in the estimation of SWE. Table 3.1 outlines the full set of input variables considered in our ablation and SWE modeling exercises, as well as their associated data sources.

Variable	Units	Data Source
Precipitation	mm	SNOTEL [PRISM]
Mean Temperature (on days t up to $t-l$, $l \in \{0,1,3,7\}$)	°C	SNOTEL [PRISM]
Min. Temperature (on days t up to $t-l$, $l \in \{0,1,3,7\}$)	°C	SNOTEL [PRISM]
Max. Temperature (on days t up to $t-l$, $l \in \{0,1,3,7\}$)	°C	SNOTEL [PRISM]
Elevation	m	SNOTEL [GTOPO30]
Slope	deg	GTOPO30
Aspect	deg	GTOPO30
Diurnal Anisotropic Index (DAH)	-	GTOPO30
Topographic Solar Radiation Aspect Index (TRASP)	-	GTOPO30
Snow Water Equivalent on previous day (SWE_{t-1})	mm	SNOTEL [Predicted]
Snow Depth on previous day (SD_{t-1})	mm	SNOTEL

Table 3.1: Full list of covariates considered in daily ablation and daily SWE modeling. Variables for which site-level data exists, the corresponding data source at the grid scale, if available, is indicated in brackets.

Slope and aspect were computed using the Global 30 Arc-Second Elevation (GTOPO30) Digital Elevation Model (DEM) dataset. DAH and TRASP are indices that account for topographic aspect and/or slope that help to determine solar radiation loading on the snowpack (a key driver of dSWE on non-precipitation days). DAH is defined by

$$\text{DAH} = \cos(\alpha_{\max} - \alpha) * \arctan(\beta),$$

where $\alpha_{\max}$ is the aspect receiving the maximum amount of heat, α is the aspect and β is the angle of the slope [5].

TRASP is defined by

$$\text{TRASP} = \tfrac{1}{2}\left[1 - \cos\left(\left(\tfrac{\pi}{180}\right)(\alpha - 30)\right)\right].$$

Beyond precipitation and average temperature, each of the other covariates in Table 3.1 falls into one of the following groups: 1) static terrain characteristics (elevation, slope, aspect, DAH, and TRASP), which provide spatial context; 2) different moments of the diurnal cycle of air temperature (minimum and maximum temperature), which provide additional context about temperature conditions that the snowpack experiences (e.g., overnight refreezing potential); 3) historical weather conditions that influence the current snowpack state in the form of temporal lags (temperature and precipitation at lagged values), which informs recent meteorological conditions that may have lingering effects; 4) existing snowpack conditions (previous day SWE and SD).

Terrain attributes, such as local topography, slope orientation, and shading, strongly influence snow dynamics, including ablation patterns, particularly in heterogeneous mountainous landscapes. Higher elevations generally experience colder temperatures, which are

conducive to snow formation, accumulation, and persistence [3]. Steeper slopes generally accumulate less snow and experience increased ablation rates, due to gravitational effects and wind redistribution [45]. South-facing slopes receive more solar radiation and experience faster snowmelt compared to north-facing slopes [51]. Shaded areas may retain snowpack longer due to reduced solar exposure [45].

In addition to daily average temperature, daily minimum and maximum temperature are useful because, together, they capture diurnal temperature variations that are critical for understanding freezing and thawing cycles within snowpack, which influences its stability and melting rates [33]. Additionally, information on the range of temperatures experienced by the snowpack provides more comprehensive insight into the surface energy balance that governs ablation processes like snowmelt and sublimation [58, 33].

Precipitation and temperature temporal lags (indicated by lag days l in Table 3.1) also influence snow dynamics, including snow ablation, because they capture the delayed response of snowpack to changes in weather conditions. The influence of changing weather conditions (e.g., temperature, precipitation, humidity, and solar loading) can take time to accumulate and melt snow, while temperature changes affect snowmelt rates with a lag due to the thermal inertia of the snowpack [8].

Previous day SWE reflects the state of the existing snowpack. The combination of previous day SD and SWE provides additional context on the condition and density of the current snowpack, which directly impacts how quickly snow melts and how much water is released during the ablation process.

Each group of covariates has key physical connections to the evolution of snow dynamics. As a result, one may expect that including all or a subset of these covariate groups could aid a model in more accurately predicting snow ablation and SWE. We first assess, from an empirical standpoint, the relative importance of these individual groups of SWE covariates and their combination on affecting snow ablation modeling. Then, using the most promising sets of variables, we carry out the entire modeling exercise and the remainder of our analyses.

In model training, we utilize SNOTEL data for variables that are observed by SNOTEL at the site-level (i.e., precipitation, temperature variables, SD, SWE, and elevation). Variables that cannot be derived from SNOTEL data alone due to the network's spatial discontinuity, yet are important sources of spatial heterogeneity that influence snow (i.e., slope, aspect, DAH, TRASP), are computed from the GTOPO30 grid and the nearest grid cells containing the corresponding SNOTEL sites are then used in the model training. Once the models are trained on site-level data, the same or different site-level data can be fed into the trained model to predict SWE at certain stations over time or grid-level data can be used as input variables (indicated in brackets in Table 3.1) if a gridded field or product of SWE is desired.

It is important to note that while SNOTEL SWE_{t-1} is listed as a covariate in Table 3.1, it is not a feasible model input since SWE data is not readily available at the grid-level and SWE is also the variable we are ultimately interested in predicting. Thus, when generating continuous time series from our trained daily models, the previous day SWE predicted by our model is used in place of previous day SNOTEL SWE. SD_{t-1}, on the other hand, is only included and considered in the covariate set to gauge the influence of snow density, since

it is a key physical factor affecting ablation. SD is not used in the main models leading to the final results due to gridded data limitations, but a dedicated discussion is included in Section 3.4 to provide a note on the utility of snow depth in informing ablation, in the event that snow depth data becomes available at a more consistent spatial and temporal scale in the future.

Statistical and Machine Learning Models

For each daily ablation and daily SWE modeling approach and each covariate set, we examined several statistical and machine learning approaches. Since we expect the effects of the physical variables to be non-additive, we investigated linear and non-linear models of varying complexities. Linear models are more interpretable and computationally efficient and provide a baseline understanding of relationships between variables. Non-linear models, on the other hand, are better able to capture interactions and intricacies in non-additive and non-multiplicative relationships. Specifically, we considered regression, random forest, and neural network models. In the following subsections, we provide a brief overview of each of these model types.

Regression Linear regression is a widely used modeling technique that aims to establish a linear relationship between the input variables X (Table 3.1) and the continuous target variable Y (daily ablation or dSWE). The model can be summarized by:

$$Y = X\beta \tag{3.6}$$

where $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$, and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$. for p input variables. The model assumes a linear combination of the input covariates to predict the output variable.

While linear regression is simple to implement and interpret, real world data are usually complex and nonlinear. Polynomial regression, an extension of linear regression that capture non-linear relationships between the input and target variables, can therefore be considered. The fundamental idea is to fit a polynomial equation of degree k to the data, allowing for a more flexible representation of the underlying relationship. The general form of a polynomial regression can be expressed exactly as Eq. (3.6),

with $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{12}^2 & \dots & x_{1m}^k \\ 1 & x_{21} & x_{22}^2 & \dots & x_{2m}^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2}^2 & \dots & x_{nm}^k \end{bmatrix}$, and $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$.

where $m = k \cdot p$. The coefficients of the input variables, β, are estimated using optimization algorithms, such as ordinary least squares, that minimize the difference between the predicted values from the equation and the actual observed values.

In our modeling efforts, we consider polynomial regression with linear, quadratic, and cubic terms. It is imperative to acknowledge that incorporating additional terms and non-linear extensions to the simple linear regression can improve the model's fitting capacity, but the resulting model may be more complex, less interpretable, and more susceptible to overfitting.

Random Forest (RF) Random forest (RF) regression [7] is an ensemble learning technique. It trains a large number of decision trees on random subsets (bootstrap samples) of the training data and randomly selects covariates at each node for splitting. Each decision tree makes independent predictions and the final prediction is the average prediction across all trees. The randomness in both sample selection and covariate choice reduces the possibility of overfitting and improves the model's robustness. RF can handle both linear and non-linear relationships and is equipped to handle high-dimensional data with large number of covariates that may involve complex interactions. In our study, we build RFs consisting of 100 trees.

Neural Network (NN) An artificial neural network (NN) [37] is a deep learning model that consists of interconnected nodes, or neurons, organized into layers that process information hierarchically. During training, NNs use backpropagation to iteratively adjust the weights and biases of the connections between the neurons and refine the network's parameters, by minimizing the error between the predicted and actual output. This iterative process helps the model capture complex non-linear relationships between input and output variables. In our study, we employ NNs with a single hidden layer consisting of 1024 nodes and we train the networks for 1500 epochs using a learning rate of 0.0001 and batch size of 256.

Cross-Validation

To train our daily ablation and dSWE models, we divide the 30 water years using an 80%-20% train-test split, resulting in 24 water years for training and 6 for testing. We employ k-fold cross-validation with k = 4 to validate our models. In k-fold cross-validation, the 24 water years in the training set are randomly partitioned into four folds of six water years each. During each iteration, one fold is designated as the validation set while the remaining three folds constitute the training set. Daily ablation and dSWE models are trained on the training set and evaluated on the validation set, with performance metrics computed on the validation set and retained. This process iterates through all folds, ensuring each serves as the validation set exactly once. After completing the cross-validation procedure, performance metrics such as those outlined in Section 3.3 are averaged across all folds to

obtain a final performance metric. Averaging these metrics across folds provides a robust measure of model performance, accounting for variations in training-validation splits.

Evaluation Schemes

We assess our models following three validation schemes, illustrated in Figure 3.3. First, we compare our daily ablation models amongst themselves in a straightforward approach on the daily scale. Then, we assess the performance of our daily ablation models compared to the UA SWE quadratic regression in estimating cumulative ablation. Finally, we compare the effect of our daily ablation model, dSWE model, and a point-level UA SWE approach in predicting cumulative SWE across time and space. In the following subsections, we describe each of the validation schemes in detail.

Daily Ablation

First, we directly evaluate the performance of our daily ablation models. We carry out this evaluation scheme by preprocessing the validation and test sets following the same procedure as the preprocessing of the training set, as outlined in Section 3.3, and then comparing our daily ablation model predictions on the remaining dates. This approach allows a straightforward one-to-one comparison of our daily ablation predictions to observed daily ablation without any influence from decisions in the time series generation. The directness of this approach also allows us to investigate the importance of various sets of covariates (Section 3.3) under this evaluation scheme. Only the most promising covariate sets determined by this evaluation are used in subsequent modeling to generate cumulative ablation and SWE time series.

Cumulative Ablation Time Series

Next, we evaluate the performance of our cumulative ablation predictions. In this evaluation approach, we simultaneously examine the effect of our proposed time series generation procedure and compare cumulative ablation predictions derived from our modeling approach directly against cumulative ablation generated by the UA SWE algorithm.

The true cumulative ablation against which we validate our predictions and UA SWE predictions follows UA SWE's definition of cumulative ablation in their original estimation of the ablation regression coefficients:

$$A_{1:t} = \hat{S}_{1:t} - \text{SWE}_t \tag{3.7}$$

where $\hat{S}_{1:t}$ is accumulated snow, estimated by summing daily snowfall, which is determined by partitioning daily precipitation with the 2°C rain-snow threshold.

SWE

Finally, we evaluate the performance of SWE predictions — the ultimate quantity of interest — derived from both our daily ablation and dSWE modeling approaches, in comparison to SNOTEL net snowfall from our reconstruction of UA SWE and a bias-corrected version of SNOTEL net snowfall (since these are point-level estimates in the UA SWE algorithm). All of these model-predicted quantities are validated against SNOTEL SWE, the true SWE.

Performance Metrics

We evaluated the performance of all the models under all evaluation schemes using the coefficient of determination (R^2) and root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (\text{observed}_t - \text{predicted}_t)^2}{\sum_{t=1}^{n} (\text{observed}_t - \overline{\text{observed}})^2} \tag{3.8}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (\text{observed}_t - \text{predicted}_t)^2}{n}} \tag{3.9}$$

where n is the total number of data points, i.e., days across sites and years, and $\overline{\text{observed}}$ denotes the mean of the observations. R^2 typically takes values between 0 and 1, with 1 indicating perfect alignment of the predictions to the observations; however a negative R^2 is possible if the model performs worse than the mean. This is similar to the Nash-Sutcliffe Efficiency (NSE), which ranges from $-\infty$ to 1, where 1 denotes perfect prediction. RMSE takes values greater than or equal to 0, where 0 indicates perfect alignment of the predictions with the observations.

As we employed cross-validation, all metrics reported in Section 3.4 are metrics averaged over the validation folds.

3.4 Results

Evaluation of Daily Ablation: Covariate Selection

To establish baselines for evaluating covariate importance, we first assess the predictive performance of the full daily ablation model with average temperature as the sole input variable, and then assess the predictive performance of the full daily ablation model with average temperature and precipitation as the only two input variables. We specifically employ the full daily ablation model for this evaluation as it has the fewest constraints and therefore contains the largest number of data points encompassing the widest range of conditions. We identify the models with average temperature and average temperature and precipitation inputs as baselines because average temperature is the sole variable on which UA SWE depends for predicting ablation and precipitation is also an important factor influencing ablation, through processes such as rain-on-snow [71]. The R^2 values of models trained on

these two covariate sets are represented by the black and grey bars, respectively, in Figures 3.5 (through 3.10).

When average temperature is the sole input variable, model performance increases with model complexity, with the machine learning models (RF and NN) outperforming regression models and additional coefficients in regression leading to improved performance. Including precipitation in addition to average temperature in the covariate set has little effect on the predictive power of the regression models, but improves the performance of both machine learning models.

In the following subsections, we evaluate the efficacy of including additional covariates beyond precipitation and average temperature, specifically by the groups outlined in Section 3.3. As mentioned previously, the goal is to identify the most promising covariate set(s) to best estimate cumulative ablation and SWE.

Terrain Attributes

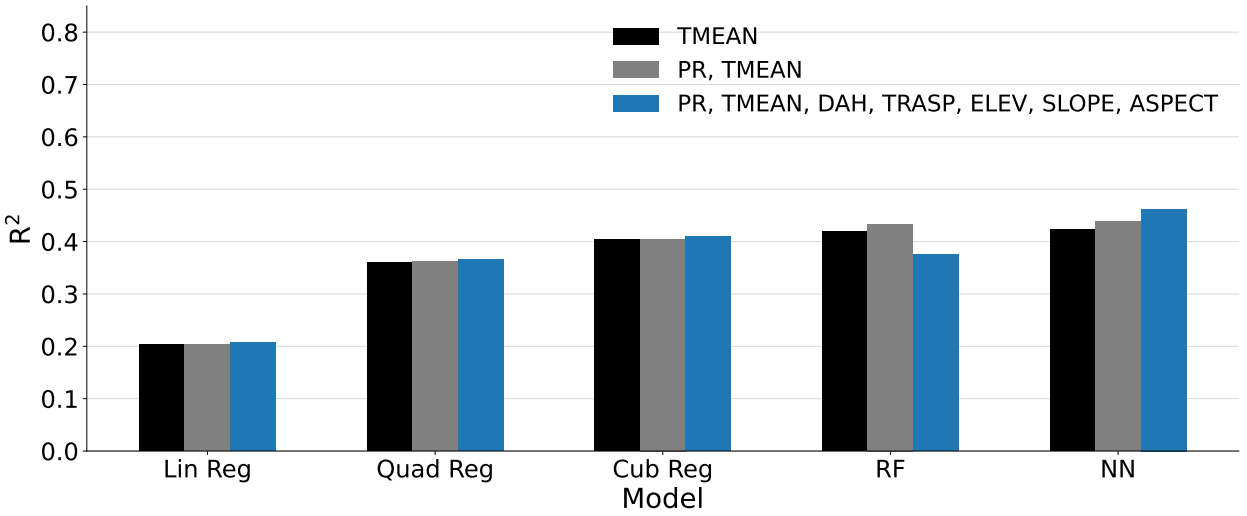


Figure 3.5: Performance of the full daily ablation model trained only with static terrain covariates, compared to the baselines, under the daily ablation evaluation approach.

When including static terrain attributes in addition to the baseline precipitation and average temperature covariates, the performance of the regression models is minimally improved, while the performance of the neural network improves more noticeably. However, the performance of the RF decreases.

Granular Temperature

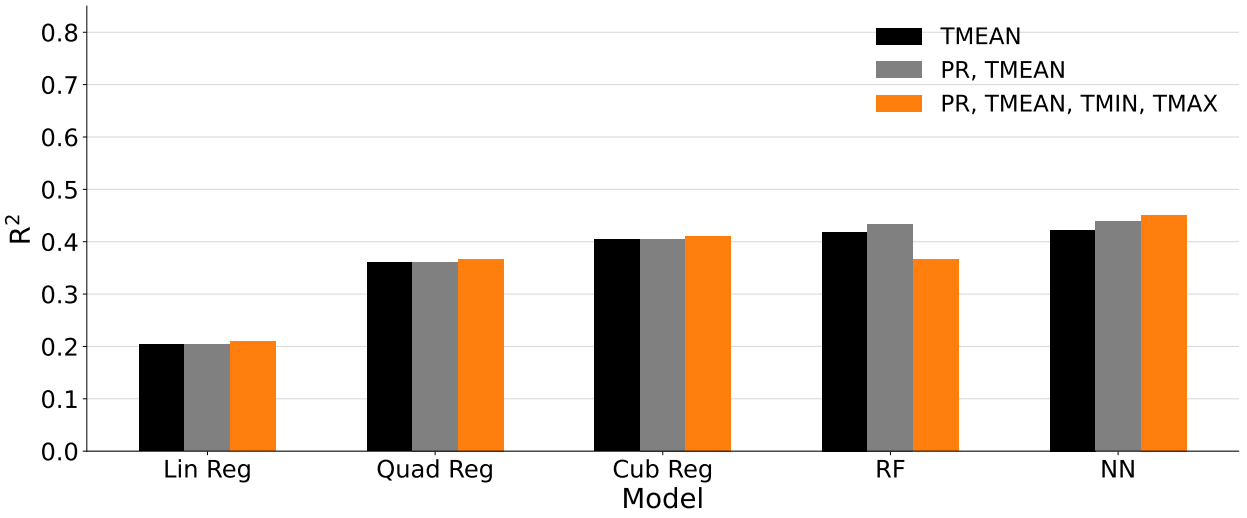


Figure 3.6: Performance of the full daily ablation model trained only with granular temperature covariates, compared to the baselines, under the daily ablation evaluation approach.

When including granular temperature variables in addition to the baseline covariates, the behavior in performance across statistical models is very similar to that when adding only static terrain attributes. Specifically, the performance of the regression models minimally improves and the performance of the neural network more noticeably improves, while the performance of the RF reduces. The magnitude of the improvement in neural network against the baseline neural networks is smaller than the magnitude of the improvement in the neural network adding only static terrain attributes to the baseline covariates. Similarly, the magnitude of the reduction in performance of the random forest against the baseline neural networks is larger.

Historical Weather Conditions

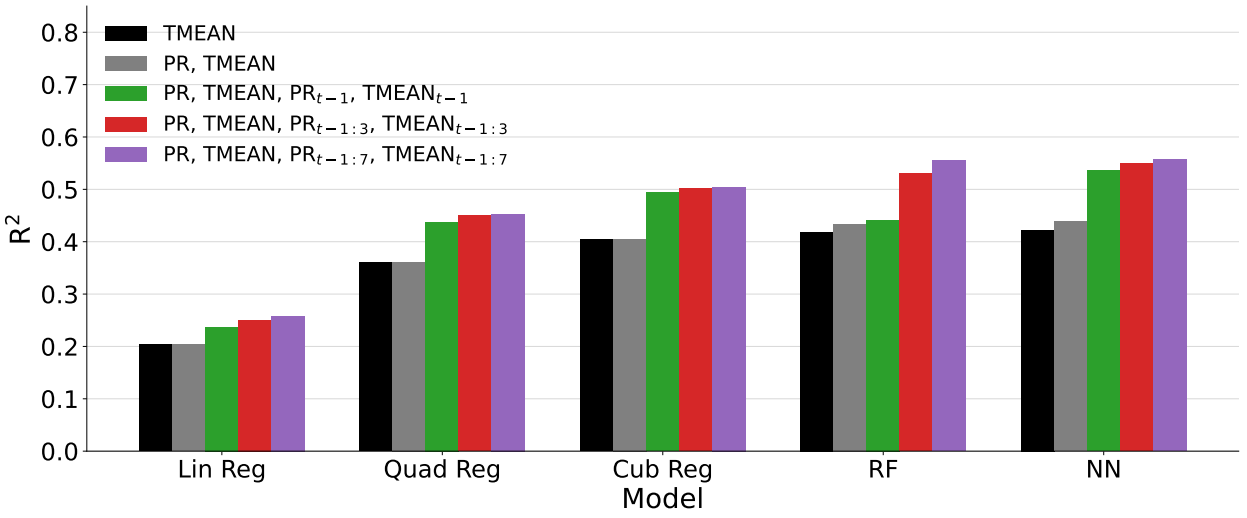


Figure 3.7: Performance of the full daily ablation model trained with historical weather condition covariates of differing temporal lags, compared to the baselines, under the daily ablation evaluation approach.

Including a one day lag of precipitation and average temperature improves the performance of all models against the baseline — significantly in the regression and NN models and marginally in the RF.

Including longer precipitation and average temperature lags in the covariate set improves the performance across all models. Notably, including additional lag days significantly improves the performance of RF, moving it from one of the worst performing models (when only a one day lag is included) to the best performing model (when up to seven lag days are included). This significant change in RF performance with the inclusion of additional meteorological lag variables could suggest the existence of complex temporal patterns in the data that the random forest model can't adequately capture unless sufficient lags are explicitly fed into the model. Overall, out of the various maximum lag days considered for precipitation and average temperature, seven days leads to the most favorable performance across models.

Including information on historical weather conditions strongly enhances the performance of all models, in contrast to including information on terrain attributes or granular temperature, both of which exhibit mixed impacts across models, with the improvements relatively minimal.

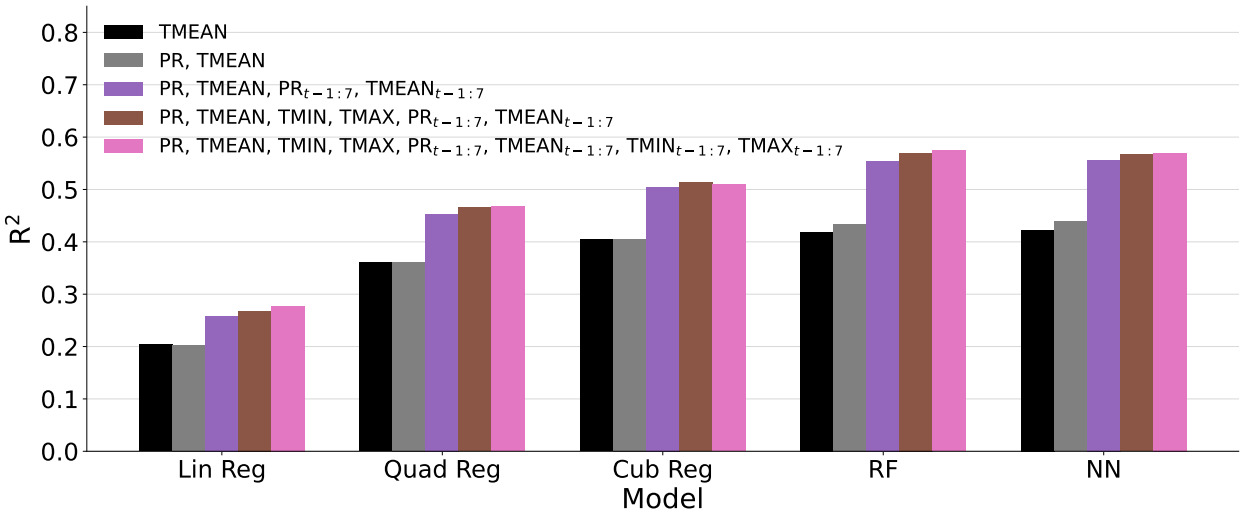


Figure 3.8: Performance of the full daily ablation model trained with static granular temperaturelags, compared to the baselines, under the daily ablation evaluation approach.

Lags for Granular Temperature Given the significant improvements in model predictive performance exhibited when lag days for precipitation and average temperature are included in the covariate set, a question about the potential utility of further incorporating minimum and maximum temperature and their lags naturally arises. In Figure 3.8, we investigate these effects. Since including up to seven days of lags proved to be the most effective (Figure 3.7) out of the values examined, we only show results for up to seven lag days.

When including the granular temperature variables, in addition to the baseline and historical weather variables, the predictive performance increases in all models. Implementing lags of the granular temperature variables leads to marginal improvement in linear regression, quadratic regression, RF, and NN, and a slightly worse performance in cubic regression.

All Covariates

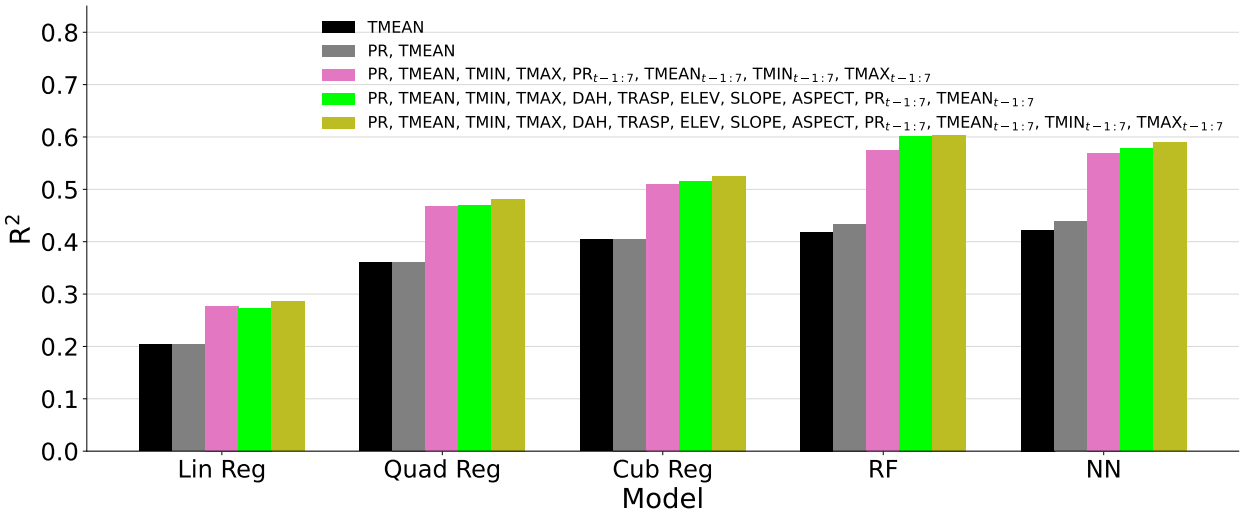


Figure 3.9: Performance of the full daily ablation model trained with all static terrain, granular temperature, and historical weather (seven day lag) covariates, compared to the baselines, under the daily ablation evaluation approach.

As illustrated earlier in Figures 3.5 to 3.8, incorporating information on terrain attributes, granular temperature, and historical weather conditions independently improves the performance of some, if not all, models, by varying magnitudes. Since the interaction between these variables could reveal additional insights that further enhances model learning, we also consider the effect of training the models with all aforementioned covariates.

Leveraging all covariates improves predictive performance across all models (Figure 3.9), in comparison to the baseline models as well as to all previously considered models trained on fewer covariates. Notably, the inclusion of lags on the granular temperature variables in conjunction with all other physical variables leads to improvements in model predictive performance across the board.

SWE

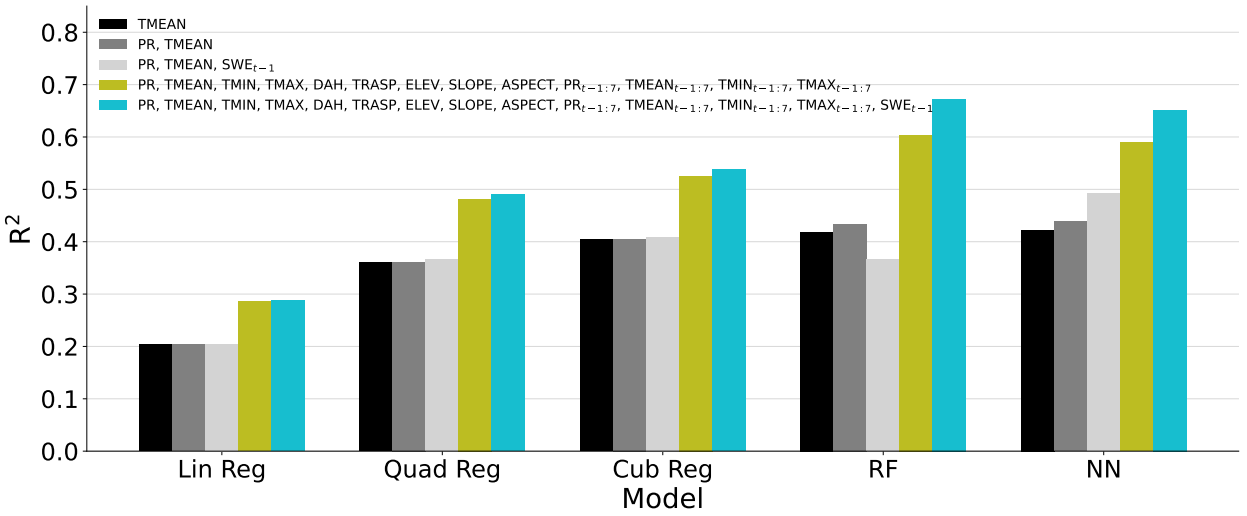


Figure 3.10: Performance of the full daily ablation model trained with covariates sets including previous day SWE, compared to the baselines, under the daily ablation evaluation approach.

Including previous day SWE on top of the baseline precipitation and average temperature input variables marginally improves the performance of the regression models and noticeably improves the performance of the neural network, while reducing the performance of the random forest. The performance pattern is similar to that observed in Figures 3.5 and 3.6.

However, when including previous day SWE in conjunction with all other physical variables, the performance of all models is significantly improved, even compared to the previously best performing covariate set (all aforementioned covariates except SWE$_{t-1}$), with random forest continuing to show the best performance.

SD

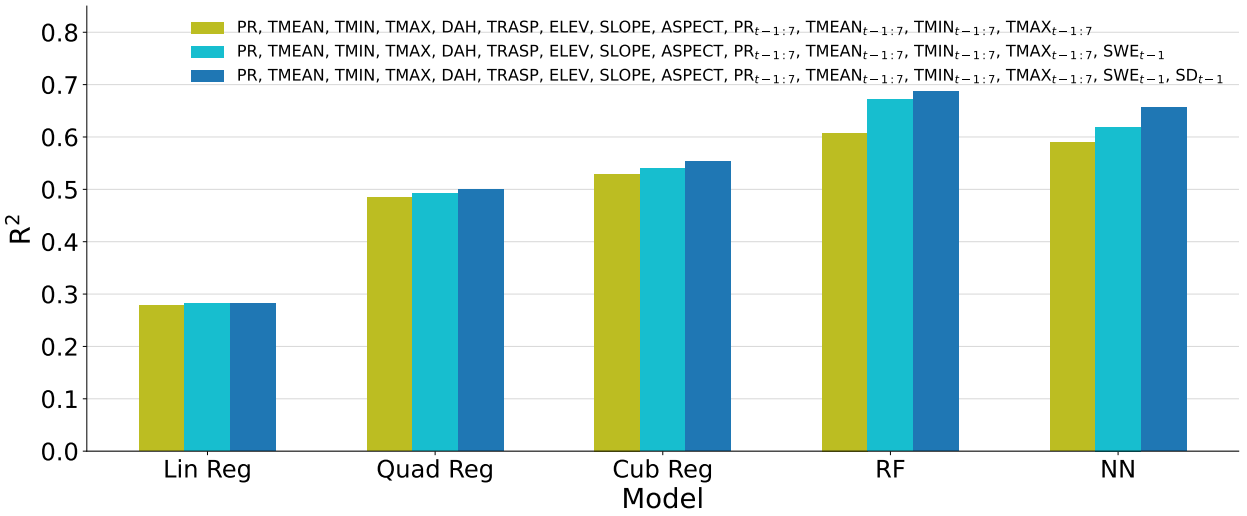


Figure 3.11: Performance of the full daily ablation model trained with previous day SD in addition to previous day SWE, compared to the baselines, under the daily ablation evaluation approach.

When previous day SD is also included in the covariate set along with previous day SWE (and all other covariates), the predictive performance of all models increases further, suggesting that snow density provides additional insight for predicting snow ablation.

Note that due to additional missing data in the SNOTEL SD record, compared to the other SNOTEL meteorological records, the results in Figure 3.11 are based on training and validating models on a subset of site-days from the dataset used in previous subsections. This difference in datasets explains discrepancies between the R^2 values reported for the full covariate sets excluding SD_{t-1} but both excluding and including SWE_{t-1} (olive green and cyan bars in Figures 3.11 and 3.10).

Discussion

When evaluating the covariate groups individually, including terrain attributes, granular temperature or previous day SWE as covariates in addition to the baseline precipitation and average temperature provides marginal benefit in the performance of the regression models, some improvement to the performance of the NN, and decreases the performance of the RF.

However, the inclusion of temporal lags of the meteorological variables improves model performance across the board, especially when more lagged values are included. Including lagged minimum and maximum temperature in addition to lagged precipitation and average temperature further improves the model performances across the board, although the impact

is not particularly large. In both cases, including lagged variables up to seven day is the most effective, followed by including lag variables up to three days, which is in turn more effective than including only one day lag variables. This suggests that incorporating information about past meteorological conditions at larger lags is more beneficial for predicting snow ablation. However, additional experimentation is needed to determine the lag threshold at which this utility may begin to converge or decline.

Including all covariates with seven day lags except SWE_{t-1} and SD_{t-1} improves the predictive performance of all models over the next best performing corresponding models. This suggests that there are interactions and/or non-linear relationships between the individual covariates that are only revealed when covariates are considered altogether.

Injecting information on the previous snow state, via previous day SWE or via previous day SWE and SD, on top of all other terrain, temperature, and historical variables further improves model prediction, with the latter exhibiting the best performance.

Based on the covariate selection exercise carried out on the daily scale here, we select the best covariate set for generating time series of cumulative ablation and SWE. While the model trained on all covariates, including previous day SWE and SD, shows the strongest performance overall, data limitations in gridded SD and significant amount of missing data in the SD SNOTEL measurements makes this covariate set a less feasible option. In the future, if point- and grid-level SD data become more abundant and reliable, a covariate set containing SD is a promising avenue to reconsider.

The second best performing covariate set is the one containing all covariates in Table 3.1 except SD_{t-1}. Recall, however, that because we do not possess SNOTEL SWE data in the validation phase, we instead substitute our model-predicted previous day SWE ($\widehat{\text{SWE}}_{t-1}$), during the prediction phase. As there remains uncertainty around whether and how SWE_{t-1} may differ from $\widehat{\text{SWE}}_{t-1}$, we carry out subsequent analyses of cumulative ablation and SWE using models trained on the full covariate set with up to seven day lags both including and excluding previous day SWE.

Evaluation of Cumulative Ablation

The daily ablation evaluation scheme identified the full set of covariates with seven day lags including and excluding previous day SWE as the practical covariate sets leading to the most promising model performance. The results shown in the following sections are based on these two covariate sets.

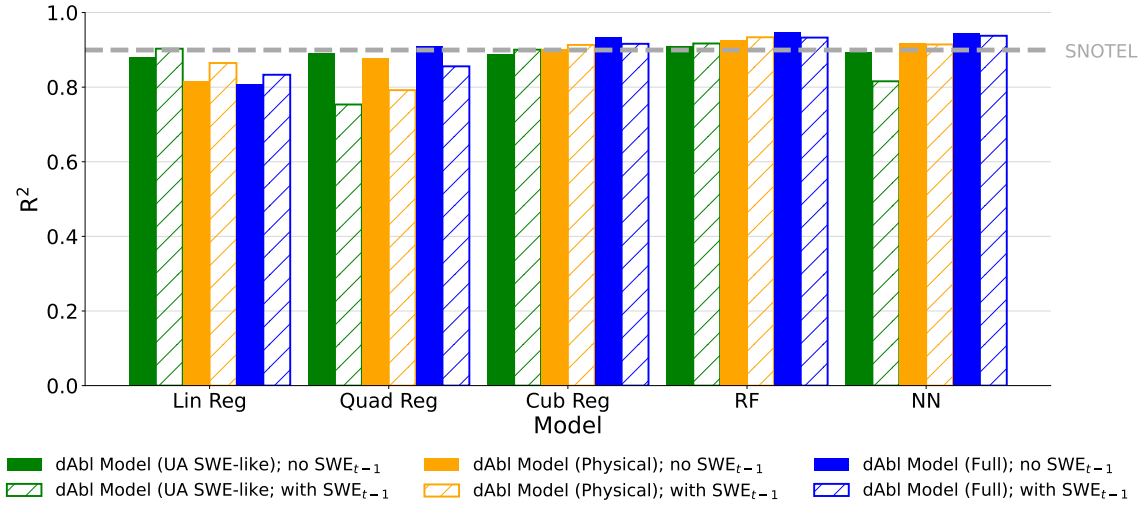


Figure 3.12: Performance from the cumulative ablation evaluation scheme. The green bars correspond to the UA SWE-like daily ablation (dAbl) model, the orange bars correspond to the physical daily ablation model, and the blue bars correspond to the full daily ablation model. For each daily ablation model type, the solid bar corresponds to the model excluding $\widehat{\text{SWE}}_{t-1}$ from the covariate set, and the textured bar corresponds to the model including $\widehat{\text{SWE}}_{t-1}$ in the covariate set.

In the daily ablation evaluation scheme, we focused on comparing our models with different covariate sets and model choices. Now, in the cumulative ablation evaluation scheme, we accumulate daily predictions generated by the trained models over time to construct cumulative ablation time series (as detailed in Section 3.3). Once on the cumulative scale, we can naturally extend the comparison to cumulative ablation predicted by the UA SWE algorithm. For a direct point-level comparison, in Figure 3.12, the performance of cumulative ablation predictions from our models are compared against our implementation of the UA SWE cumulative ablation regression approach applied to SNOTEL input temperature.

Recall that we considered three configurations for the daily ablation modeling setup (Section 3.3). The goal of examining these different settings is in part to construct our approach in a way that closely resembles UA SWE to allow for a more straightforward comparison of UA SWE to the two more physically based daily ablation modeling approaches (i.e., the UA SWE-like daily ablation model acts as a bridge to facilitate comparison of UA SWE to the physical and full daily ablation models).

In Figure 3.12, if we stratify the three modeling configurations by inclusion or exclusion of $\widehat{\text{SWE}}_{t-1}$ from the full covariate set, we observe that cumulative ablation derived from the full daily ablation model generally performs more favorably than cumulative ablation predicted by the UA SWE-like and physical daily ablation models across most model types. This indicates that the full daily ablation modeling approach improves upon the approach

employed by UA SWE, specifically in largely eliminating its decisions surrounding when ablation may occur.

Focusing on the modeling configurations when $\widehat{\text{SWE}}_{t-1}$ is excluded from the covariate set (solid colored bars), the full daily ablation model exhibits the largest variance in performance across models, followed by the physical daily ablation model, while the performance of models under the UA SWE-like daily ablation model configuration is much more similar. This is not surprising given that the full daily ablation model configuration requires the model to predict across the largest number of days, while the UA SWE-like daily ablation model configuration is the most limited in terms of number of days on which the model predicts ablation (versus assuming zero ablation). However, when focusing on the model configurations when $\widehat{\text{SWE}}_{t-1}$ is included in the full covariate set, this pattern is not apparent.

Dissecting the cumulative ablation results further, we can compare the results within each model configuration. For the UA SWE-like daily ablation model configuration (solid and textured green bars), Figure 3.12 shows that including $\widehat{\text{SWE}}_{t-1}$ in the full covariate set improves the performance of all models except quadratic regression and NN. For both covariate sets (including and excluding $\widehat{\text{SWE}}_{t-1}$), RF exhibits the best performance.

A similar pattern is observed within the physical daily ablation model configuration (solid and textured orange bars), where including $\widehat{\text{SWE}}_{t-1}$ in the full covariate set improves the performance of all models except quadratic regression and NN, although the difference in NN performance is smaller than the UA SWE-like daily ablation model configuration. For both covariate sets, RF once again exhibits the best performance.

A different pattern is present within the full daily ablation model configuration (solid and textured blue bars), where including $\widehat{\text{SWE}}_{t-1}$ in the full covariate set only improves the performance of the linear regression. When including $\widehat{\text{SWE}}_{t-1}$ in the full covariate set, NN exhibits the best performance (followed closely by RF). In contrast, in the case of excluding $\widehat{\text{SWE}}_{t-1}$ from the full covariate set, RF has the best performance (followed closely by NN).

When cumulative ablation predicted by our models is compared against SNOTEL cumulative ablation following the UA SWE methodology, our RF models from all configurations outperform SNOTEL cumulative ablation. The physical and full cubic regression and NN models also exhibit comparable or better performance to SNOTEL cumulative. The linear and quadratic regression generally perform worse than SNOTEL cumulative ablation, although the UA SWE-like linear regression model including $\widehat{\text{SWE}}_{t-1}$ in the covariate set and the full cubic regression model model excluding $\widehat{\text{SWE}}_{t-1}$ from the covariate set show comparable performance to SNOTEL cumulative ablation.

Overall, the RF under the full daily ablation model configuration excluding $\widehat{\text{SWE}}_{t-1}$ from the full covariate set exhibits the best performance in cumulative ablation prediction across all model configurations, statistical models, and covariate sets.

Evaluation of SWE at SNOTEL Sites

Since deriving SWE from our daily ablation models depends solely on linear arithmetic operations, the relative performance of our models in predicting SWE under the different daily ablation modeling configurations will be equivalent to the relative performance of our models in predicting cumulative ablation seen in Figure 3.12. In view of this, this section will focus solely on results from the full daily ablation modeling approach, comparing them against results from the dSWE model and SNOTEL net snowfall from UA SWE.

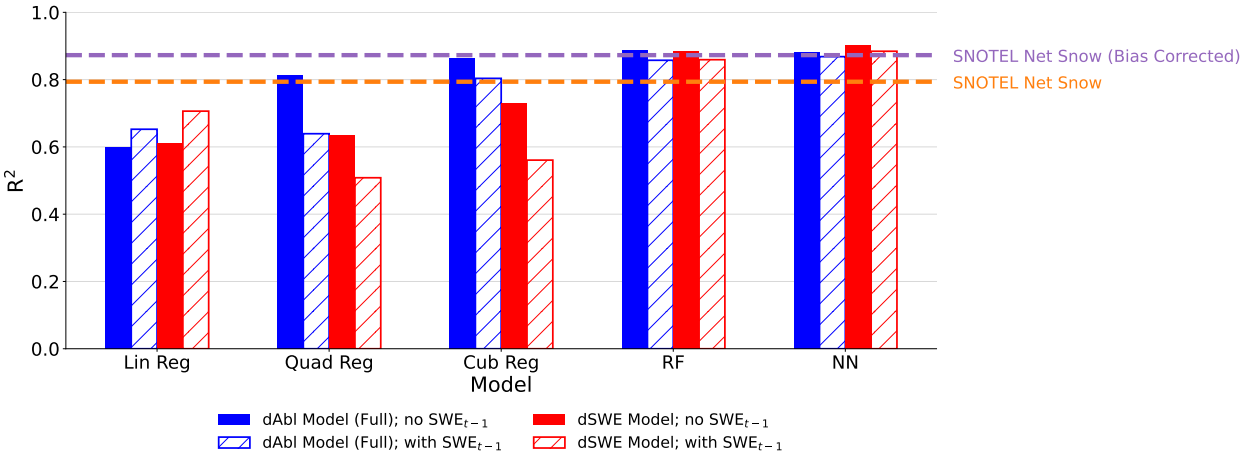


Figure 3.13: Performance from the SWE evaluation scheme. The blue bars correspond to the full daily ablation model, whereas the red bars correspond to the dSWE model. For each model type, the solid bar corresponds to the model excluding $\widehat{\text{SWE}}_{t-1}$ from the covariate set, and the textured bar corresponds to the model including $\widehat{\text{SWE}}_{t-1}$ in the covariate set.

The pattern of behavior across models under the full covariate set including and excluding $\widehat{\text{SWE}}_{t-1}$ is the same under the full daily ablation and dSWE modeling configurations (despite differing in magnitudes), with the covariate set including $\widehat{\text{SWE}}_{t-1}$ outperforming the covariate set excluding $\widehat{\text{SWE}}_{t-1}$ only in linear regression and underperforming across all other models.

Comparing the full daily ablation and the dSWE model, we observe that the dSWE model performs significantly worse than the full daily ablation model in the quadratic and cubic regression models, while showing comparable performance in RF, and better performance in the NN.

When compared to SNOTEL net snowfall (from our reconstruction of UA SWE) at SNOTEL sites, all linear regression models underperform. For the quadratic regression, only the full daily ablation with the covariate set excluding $\widehat{\text{SWE}}_{t-1}$ outperforms. For the cubic regression, the full daily ablation model with both covariate sets outperform, with the set

excluding $\widehat{\text{SWE}}_{t-1}$ exhibiting stronger performance. All RF and NN models substantially outperform SNOTEL net snowfall.

Since SNOTEL net snowfall was previously found to be a biased estimate of SNOTEL SWE (as discussed in Section 1.6, for instance), we also compare our SWE predictions to a bias corrected version of SNOTEL net snowfall (purple dashed line in Figure 3.13). The simple bias correction involves adjusting SNOTEL net snowfall for each site on each day by multiplying it by the average SNOTEL ratio computed from all sites for that day. When comparing to the bias-corrected SNOTEL net snowfall at SNOTEL sites, only our RF and NN models outperform, all with slimmer margins of improvement.

For both covariate sets, the full daily ablation and dSWE machine learning models perform well. For subsequent analyses in this section, we focus on predictions from the four machine learning models driven by the full covariate set excluding $\widehat{\text{SWE}}_{t-1}$ since this covariate set leads to the best performance within each model configuration and is generally more stable across models.

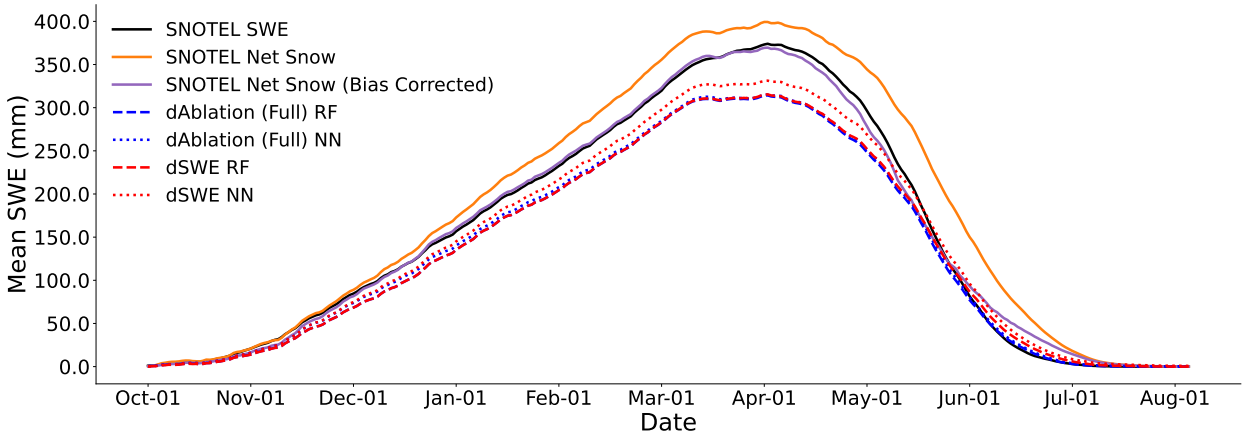


Figure 3.14: Climatological mean of observed and predicted SWE across the water year, comparing SNOTEL SWE, SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation (dAblation) and dSWE ML models.

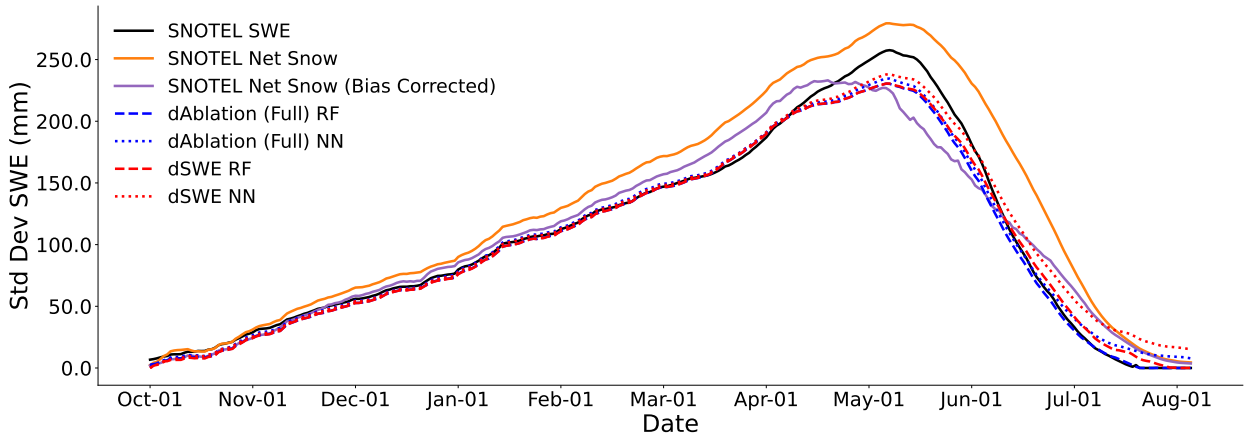


Figure 3.15: Climatological standard deviation of observed and predicted SWE across the water year, comparing SNOTEL SWE, SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation and dSWE ML models.

Figures 3.14 and 3.15 show the climatological mean and standard deviation of SWE values for each day of the water year, comparing SNOTEL against SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our four best performing models. Amongst our models, SWE predicted by the full daily ablation RF, full daily ablation NN, and dSWE RF produce similar climatological time series across the entire water year. On average, all four of our models underestimate SNOTEL SWE at SNOTEL sites during most of the water year, up until mid-May. SWE predicted by our full daily ablation RF, full daily ablation NN, and dSWE RF produce similar climatological mean across the entire water year, all of which are lower than SWE predicted by the dSWE NN around the peak SWE period between February and May. As June approaches, the average predicted SWE from all models and observed SNOTEL SWE begin to converge. Throughout the water year, SNOTEL net snowfall generally overestimates SNOTEL SWE, while the bias correction brings the average net snowfall estimate close to the average SNOTEL SWE.

In terms of spread, all four of our models exhibit approximately the same climatological standard deviation as SNOTEL SWE until around mid-April, when they all begin to underestimate until mid-June. Meanwhile, the climatological standard deviation of SNOTEL net snowfall consistently overestimates that of SNOTEL SWE. The bias correction aligns the spread of values closer to that of SNOTEL SWE; however, bias-corrected SNOTEL net snowfall values still exhibit larger spread than SNOTEL SWE values during the accumulation season and after June and smaller spread between mid-April and June. Beginning in June, the climatological standard deviation of the dSWE NN and to a lesser extent, the dSWE RF and full daily ablation NN, overestimate that of SNOTEL SWE.

Overall, the climatological mean of bias-corrected SNOTEL net snowfall agrees well with SNOTEL SWE, but the climatological standard deviation of bias-corrected SNOTEL net snowfall is less aligned with SNOTEL SWE than that of our models.

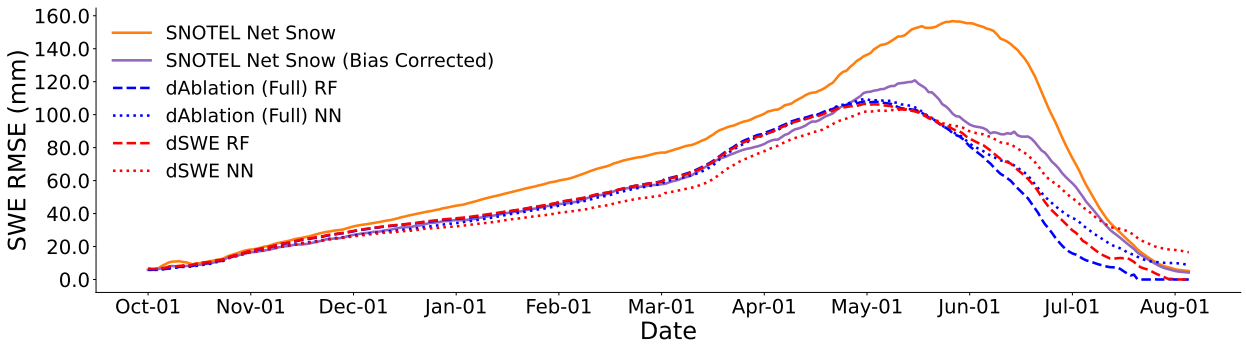
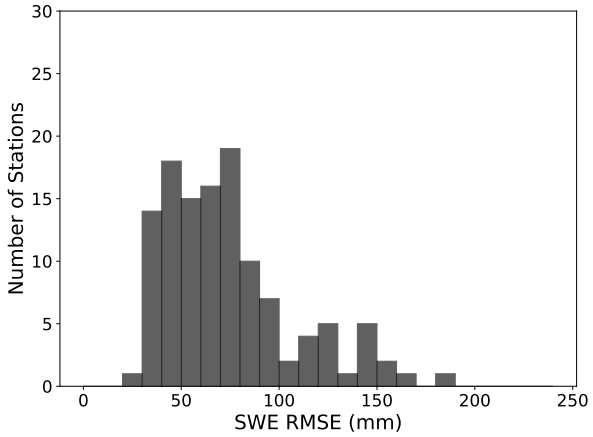


Figure 3.16: Climatological RMSE on each day across the water year for the best performing full daily ablation and dSWE ML models, compared to SNOTEL net snowfall and bias-corrected SNOTEL net snowfall.

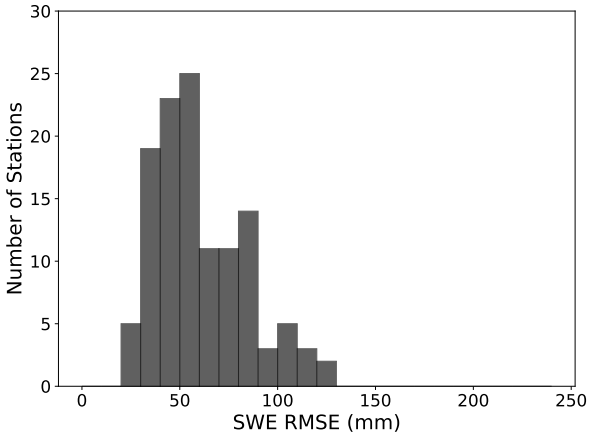
Examining the daily climatological errors (average RMSE against SNOTEL SWE across sites and years for each day) of the four best performing models (Figure 3.16), we can get a better sense of how the errors are distributed across the water year. The RMSE of the full daily ablation RF, full daily ablation NN, and dSWE RF modeling approaches are very similar during most of the snow season, from October to mid-May. Compared to bias-corrected SNOTEL net snowfall, all three models produce comparable error between October and March, demonstrate slightly greater error in April, and reduce the error between May and July. Starting around mid-May, the errors of the full daily ablation RF, full daily ablation NN, and dSWE RF begin to deviate from one another, with the former exhibiting the lowest climatological error through the end of the water year, followed by the dSWE RF and then the full daily ablation RF.

The daily climatological errors exhibited by the dSWE NN model differs from our other three models. Specifically, the dSWE NN produces lower error during the bulk of the snow season, from November to mid-May, and gradually higher error during the summer season, beginning around mid-to-late May.

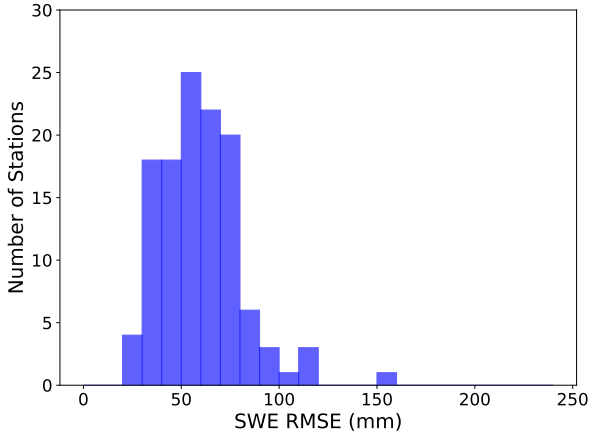
Collectively, the results from Figures 3.13, 3.14, 3.15, and 3.16 suggest that our four ML models outperform SNOTEL net snowfall at SNOTEL sites.



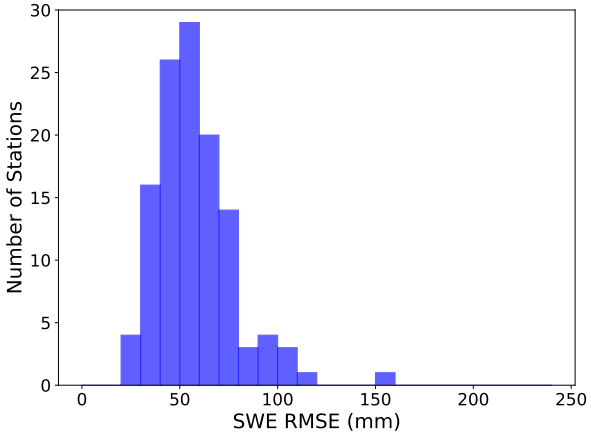
(a) SNOTEL Net Snow.



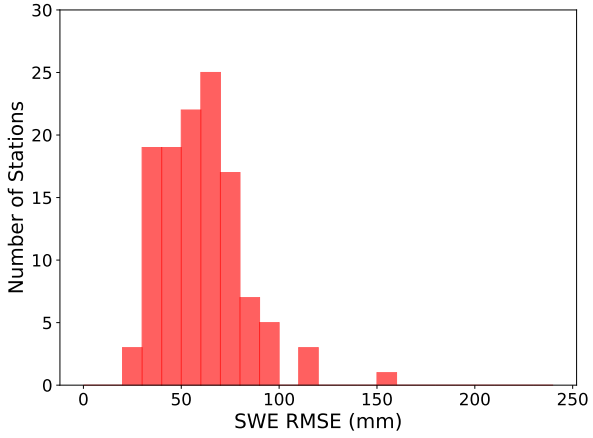
(b) Bias-Corrected SNOTEL Net Snow.



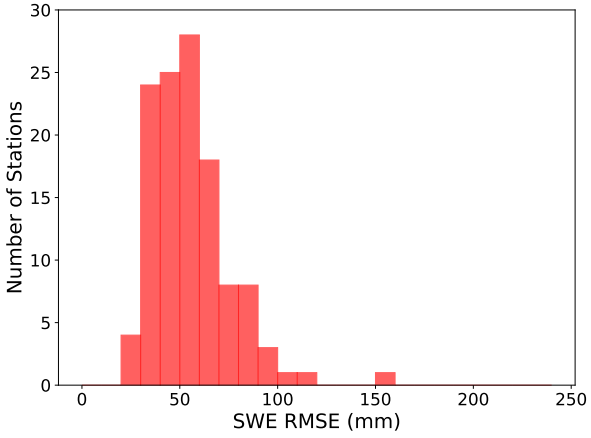
(c) Full daily ablation RF.



(d) Full daily ablation NN.



(e) dSWE RF.



(f) dSWE NN.

Figure 3.17: Distribution of site RMSE (against SNOTEL SWE) of SNOTEL net snowfall, bias-corrected SNOTEL net snowfall, and our full daily ablation and dSWE ML models.

Figure 3.17 shows the distribution of RMSE of SNOTEL net snowfall and bias-corrected SNOTEL net snowfall, and our models by site. The RMSE distribution of SNOTEL net snowfall exhibits a longer skewed tail that the bias-corrected mitigates. The error distributions of the RF model predictions are more consistent with each other, and those from the NN model predictions are more similar to each other.

Gridded SWE

All of our results outlined in the previous sections were at the point level at SNOTEL sites within the UCRB. Our proposed modeling approach can be easily extended to the grid level to produce a gridded SWE product. We can simply apply our pre-trained daily ablation or dSWE models to gridded covariate data (data sources outlined in Table 3.1). Practically, since more detailed information is derived at a higher resolution, we first convert PRISM meteorological covariate data from 4 km to 30 m resolution using nearest neighbor resampling to match the spatial resolution of GTOPO30. We then apply the pre-trained models to the 30 m resolution covariate data and then aggregate the predicted values back to the 4 km $\times$ 4 km spatial resolution to match UA SWE for comparison.

Evaluation of Gridded SWE at Grid Cells Containing SNOTEL Sites

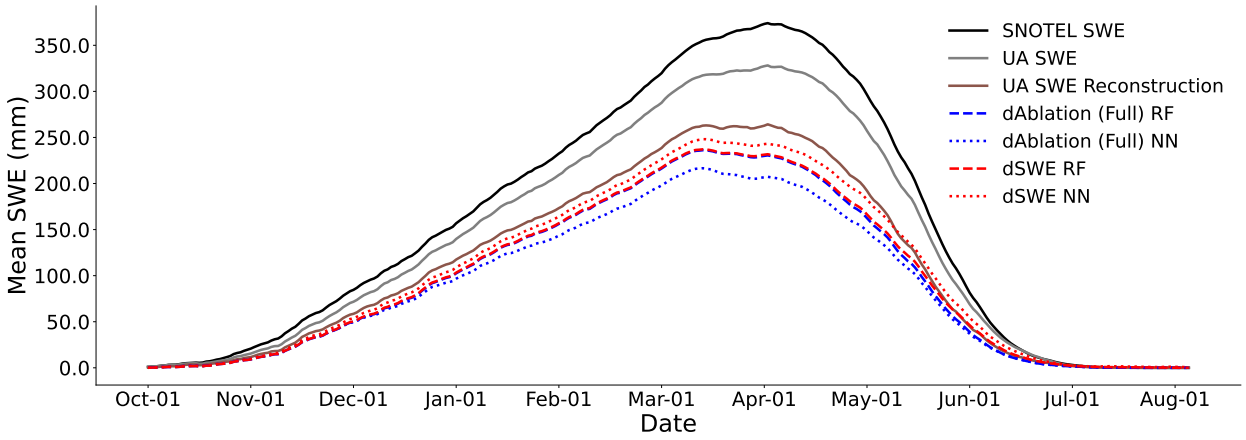


Figure 3.18: Climatological mean of observed and predicted SWE across the water year at grid cells containing the SNOTEL sites, comparing SNOTEL SWE, UA SWE, and our full daily ablation and dSWE ML models when PRISM meteorological inputs are used.

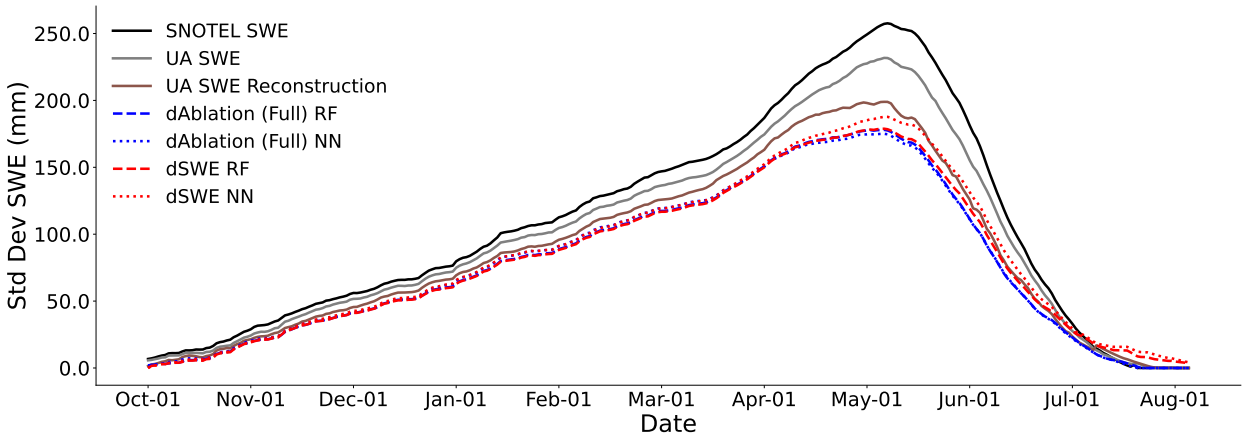


Figure 3.19: Climatological standard deviation of observed and predicted SWE across the water year at grid cells containing the SNOTEL sites, comparing SNOTEL SWE, UA SWE, and our full daily ablation and dSWE ML models when PRISM meteorological inputs are used.

Figures 3.18 and 3.19 show the climatological mean and standard deviation of SWE values for each day of the water year predicted by our four ML models when 4 km gridded PRISM is used as input data for the meteorological covariates. The distributions exhibited are respectively contrasted against Figures 3.14 and 3.15, in which point SNOTEL data was used as input meteorological data to predict SWE. Evidently, when using gridded PRISM input data instead of point SNOTEL input data, the mean and standard deviation of SWE predicted by all four of our ML models, as well as in the estimation of net snowfall, generally decreases. This demonstrates that when gridding with PRISM data, the resulting predicted SWE at grid cells containing SNOTEL sites is much lower, on average, than direct point predictions at sites using SNOTEL meteorological data.

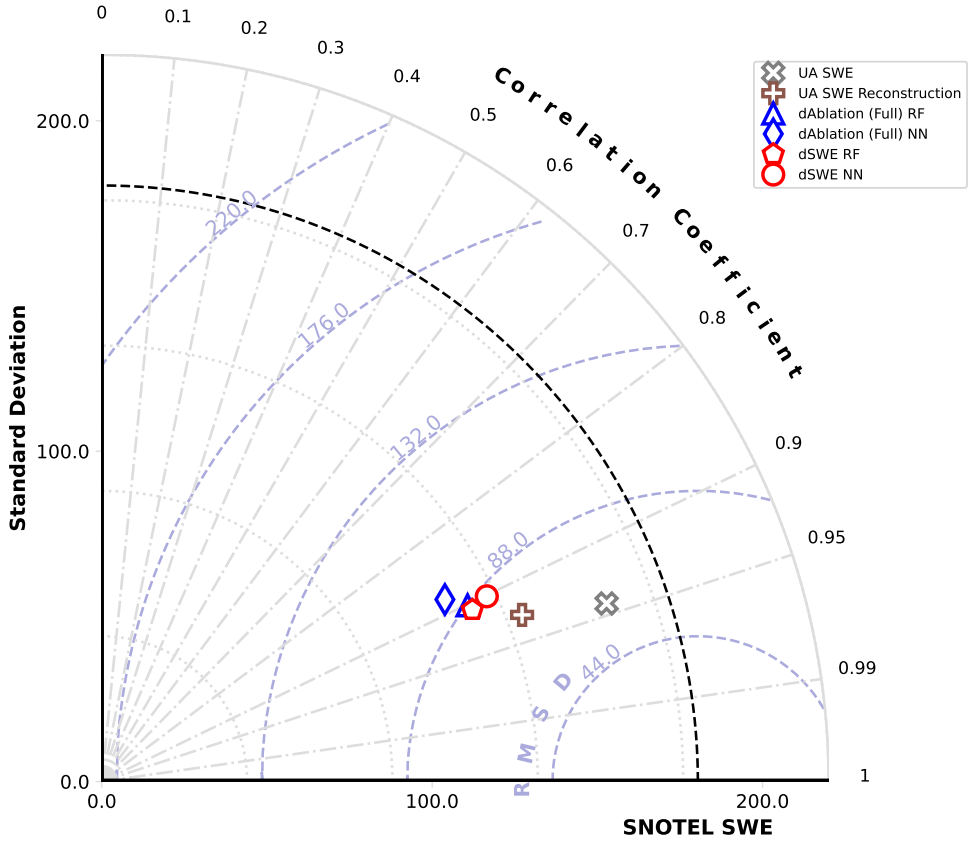


Figure 3.20: Taylor diagram.

Figure 3.20 is a Taylor diagram of the four best performing models. Taylor diagrams summarize three different statistical metrics — correlation coefficient, standard deviation, and centered root mean square difference (RMSD) — in a single graphical representation [72]. The x- and y-axes of the Taylor diagram represent the standard deviation, the curve connecting the x- and y-axes illustrates the correlation coefficient, and the faint violet curves running in the opposite direction indicate contours of RMSD ticks. The correlation coefficient quantifies the strength and direction of the relationship between observed and predicted values, taking values between -1 and 1 (only values between 0 and 1 are shown in the diagram), with 1 being an exact match of SNOTEL observed SWE with model predictions, and the RMSD takes values greater than 0, with 0 indicating exact alignment between SNOTEL observed SWE and model predictions. The standard deviation of SNOTEL SWE is shown with the black dashed curve for reference.

UA SWE outperforms our four models in terms of correlation and RMSD. UA SWE maintains a larger standard deviation that is closer to that of SNOTEL SWE. Amongst our four models, the full daily ablation RF model and dSWE ML models are comparable in correlation and RMSD and perform more favorably than the full daily ablation NN model based on these metrics.

Evaluation of Gridded SWE at the Grid Level

While ground truth SWE data is not readily available at the grid level, we can examine how gridded fields of SWE produced by our method differ from that produced by other gridded SWE products, such as UA SWE. Based on our findings thus far, the full daily ablation RF and dSWE NN models exhibit the best overall performance. In this section, we conduct an analysis of gridded fields produced by the full daily ablation model, delving into the full grid-level results. For a comparative overview, gridded results from the dSWE NN model are included in Appendix B.2. Figure 3.21 shows the 4 km gridded SWE from our full daily ablation RF model compared to UA SWE, and Figure 3.22 shows the difference in climatological average and standard deviation of peak SWE (April 1) values. The SNOTEL sites are included for reference (orange dots in the Figure 3.21 maps and black dots in the Figure 3.22 maps).

In higher elevation mountainous areas that tend to experience more snow during the winter season, UA SWE estimates are on average higher than SWE predicted by our daily ablation RF model on gridded covariate data. The spread of SWE predicted by the UA SWE product also tends to be larger than that predicted by our daily ablation RF model over most of the UCRB to varying magnitudes; however, there do exist pockets of smaller areas in the eastern side of the region that exhibit lower standard deviation.

Figure 3.23 reveals that the pattern of UA SWE estimates being higher on average than SWE predicted by our daily ablation RF model with gridded covariate data does not apply solely to April 1, but it is a trend observed throughout the snow season. Particularly in more mountainous regions, UA SWE tends to predict larger SWE, with the difference between UA SWE and SWE predicted by our model, peaking in April.

This general trend is also evident when comparing the empirical cumulative distribution functions (ECDF) of the climatological mean and standard deviation of April 1 SWE predicted by UA SWE and the full daily ablation RF model (Figure 3.24).

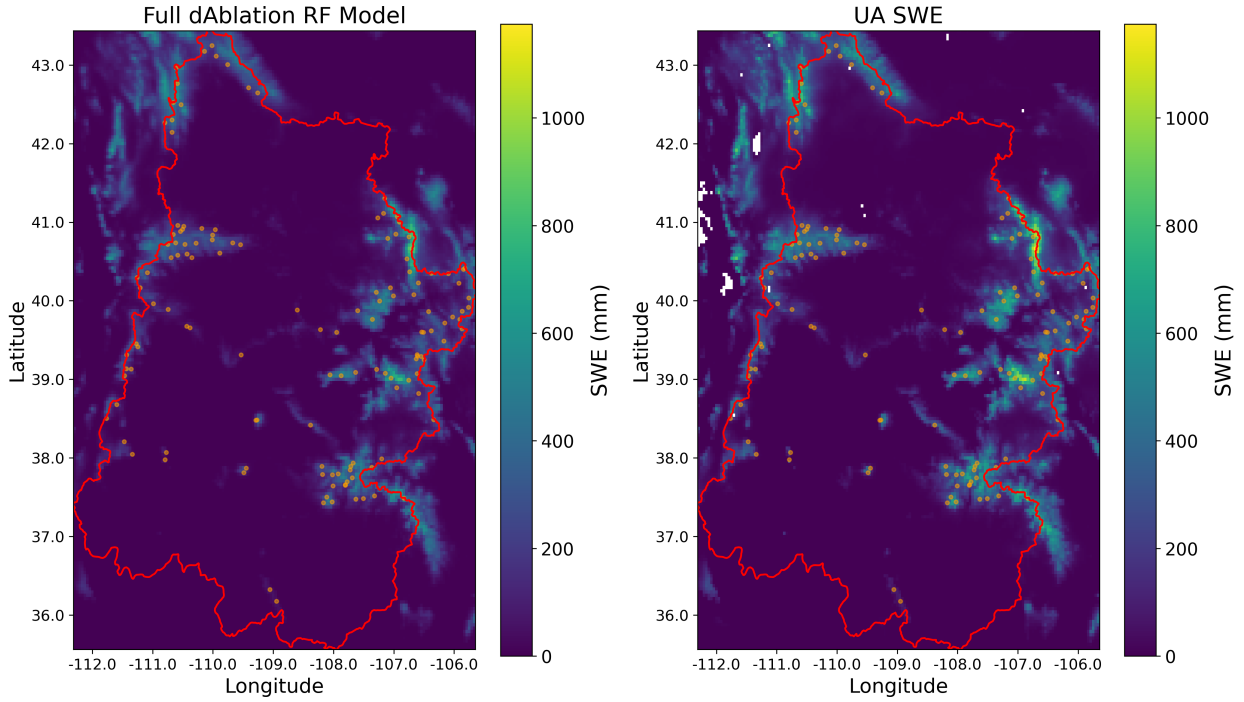
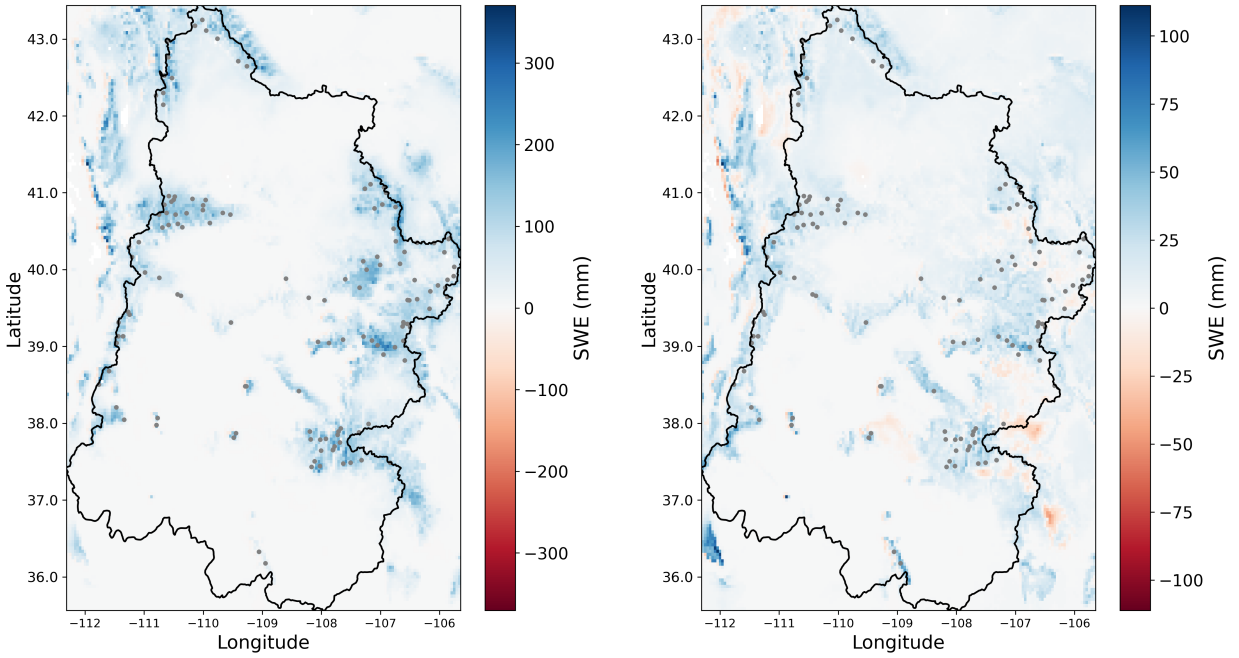


Figure 3.21: Gridded SWE prediction from the full daily ablation RF model compared to UA SWE on April 1.



(a) Differences in mean on April 1.

(b) Differences in standard deviation on April 1.

Figure 3.22: Differences in the mean and standard deviation of our full daily ablation RF model compared to UA SWE (i.e., UA SWE - full daily ablation RF SWE) on April 1, by 4 km $\times$ 4 km grid.

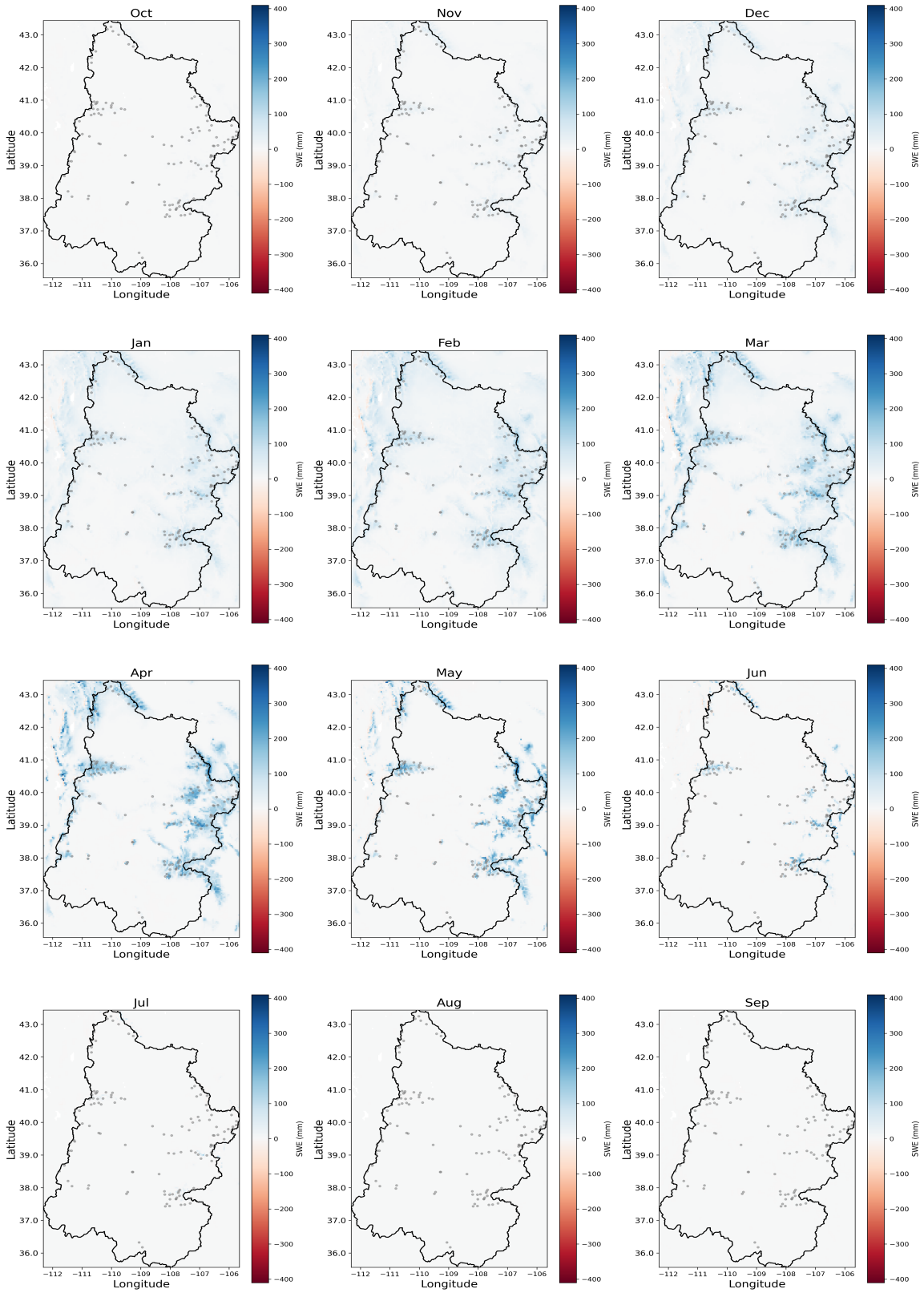
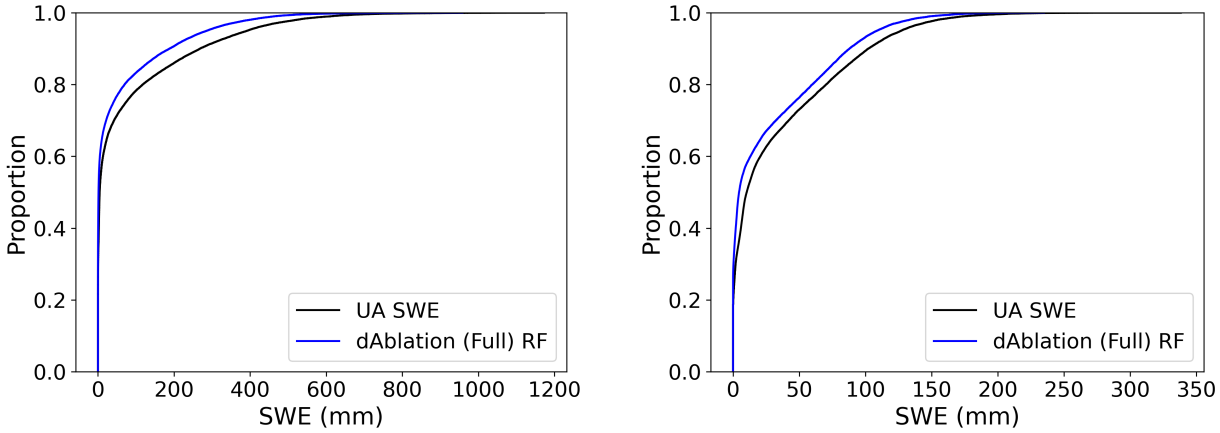


Figure 3.23: Difference in mean of our full daily ablation RF model versus UA SWE by 4 km × 4 km grid for each month.



(a) ECDF of means.

(b) ECDF of standard deviations.

Figure 3.24: Comparison of ECDFs of climatological April 1 means and standard deviations of SWE predicted by UA SWE compared to the full daily ablation RF model at the grid level over the UCRB.

Discussion

Our training and validation approaches share some similarities and differences to a recent paper [76] on machine learning for daily SWE prediction. That study, which also leverages SNOTEL data in model training, was touched upon in this chapter's Introduction. Their final model of choice was the SVR and their model configuration differed in where manual decisions were made (see Section 3.1). While their covariate set differs from ours, there are similarities, including the use of precipitation, average temperature, minimum temperature, maximum temperature, as well as consideration of precipitation and temperature lags and elevation. For model training, their study area and time period differed from ours: they used SNOTEL data from the WUS for the 2017 and 2018 water years, whereas we used SNOTEL data from the UCRB covering 80% of the water years from 1986 to 2016 (based on cross-validation train-test split).

To evaluate model performance, the authors of [76] conducted three main analyses. First, their model was tested using observation data from a few hundred SNOTEL sites in the WUS that were not part of the training dataset, covering the 2011 to 2022 water years. Sites with significant precipitation undercatch were excluded, and only those with complete daily records for at least five water years were included. They found that the model generally performed well in simulating SWE and snow melt-out dates, though variability in snow accumulation predictions was noted. Second, they evaluated model performance at 10 reference stations from the Earth System Model–Snow Model Intercomparison Project (ESM–SnowMIP). The analysis revealed general robust performance, but significant discrepancies were found at some sites, due to factors such as wind redistribution. Lastly, their

model's performance was assessed against gridded snow cover data from remote sensing for the western Pamir and central Andes regions across two consecutive snow seasons between 2014 and 2016. To generate the gridded predictions, their model relied on 1 km resolution meteorological inputs from the CHELSA-W5E5 dataset [31]. The evaluation showed high pixel-wise accuracy in snow/no-snow detection and generally captured the overall trend of snow cover area. However, various spikes in snow cover area were under- or overestimated.

For site-level evaluation, our approach focused mainly on SNOTEL data. Our method employs temporal cross-validation, where all sites in the study area are used, but data is divided based on time periods to ensure that models are trained and validated on different temporal segments. In contrast, their approach trains and tests on distinct sites and years. No cross validation was documented in their publication. For gridding, we used 4 km gridded PRISM data for the meteorological inputs, to match UA SWE. Gridded evaluation against remote sensing observations for total snow cover, such as Moderate Resolution Imaging Spectroradiometer (MODIS), would be beneficial to carry out.

Given that our results show that our approach performs better at the point level than the grid level, there is significant potential for exploring the use of such daily SWE prediction methods to generate gridded products.

3.5 Conclusion

The ablation modeling approach employed by UA SWE is limited in that it only relies on daily average temperature data and there is a significant deviation of true cumulative ablation from the cumulative ablation predicted by the quadratic regression. In this chapter, we proposed a novel intuitive, straightforward, and versatile approach for predicting SWE, based on predicting snow ablation or SWE on a daily scale. Our methodology can be easily applied to available point- and gridded data to predict SWE at point locations as well as over gridded fields.

Through thorough experimentation, we found that incorporating covariates beyond precipitation and daily mean air temperature, namely terrain attributes, historical weather conditions, and additional granular air temperature data, improves model predictive performance at out-of-sample site-dates. We also found that predicting SWE directly and predicting SWE via improved ablation modeling exhibit comparable performance, particularly when limiting the number of constraints placed on model training and prediction. Amongst the models considered, the machine learning models tended to outperform standard regression models, likely due to the ability of the ML models to capture more intricate non-linear relationships between input variables and their effect on snow ablation.

When the pre-trained models are extended to the grid level and applied to gridded input data, including 4 km PRISM data for meteorological covariates, the results indicate that our approach may not perform as well as UA SWE. Additional analysis against true gridded snow data, such as satellite-based MODIS snow cover data, is needed to draw definitive conclusions on the performance of our gridded predictions. It would also be worthwhile to

explore alternative gridded datasets for the meteorological inputs, as they could enhance the grid level performance of our models.

Presently, our technique, like the majority of other publicly available gridded SWE products, produces deterministic spatiotemporal SWE estimates. One particular direction of interest lies in transforming our deterministic SWE estimates into probabilistic estimates, incorporating uncertainties to provide a more comprehensive understanding of the range of possible outcomes.

Moreover, including additional relevant physical covariates, such as solar radiation, humidity, and wind speed, has strong potential to elevate the predictive performance of our models. The feasibility of this, however, is contingent upon the availability of point and gridded data for these variables, which is currently limited. Nevertheless, as the data space rapidly evolves and when data on these additional variables becomes more widely available in the future, retraining our models with additional covariates should be a consideration.

Chapter 4

Consideration of Spatial Scales and Representativeness in Assessing Gridded Snowpack Estimates

4.1 Introduction

Due to data limitations in ground truth snow depth (SD) and snow water equivalent (SWE) data on regular spatial and temporal intervals, the accuracy and representativeness of gridded SWE data products are often assessed at the point-level with individual grid cells compared against in-situ measurements or at the grid-level with the full gridded dataset compared against point-level measurements interpolated to the grid level [47, 4, 82]. Considering the differences in scale and uncertainty between point and gridded SWE estimates, these evaluation approaches are inadequate in for assessing gridded SWE estimates [47, 82].

With the emergence of a new remote sensing LiDAR-based measurement method — the ASO dataset, a very high spatial resolution measurement-based SD and SWE dataset is now publicly available [18, 55]. While ASO flights are only operated on select days in select mountain regions, the rich spatial information contained within the flight-derived dataset provides an unprecedented opportunity to understand, empirically, the true heterogeneity in SD and SWE in mountainous regions.

In this chapter, we leverage ASO data to evaluate the accuracy of two gridded SWE products, namely UA SWE and our gridded SWE product proposed in Chapter 3 (henceforth termed "MY SWE"). Section 4.2 outlines the datasets used in the analyses presented in this chapter. In Section 4.3, the two gridded data products are evaluated against in-situ measurements at the site level, specifically from the SNOTEL network. We then discuss challenges and issues associated with making grid-to-point comparisons. Taking advantage of the fine spatial resolution of ASO observations within key hydrological basins, we evaluate UA SWE and MY SWE on the grid-level against aggregated (gridded) ASO SWE in Section 4.4. Finally, in Section 4.5, we conduct analyses to identify potential reasons for the mismatch

between empirically-based gridded products, such as UA SWE and MY SWE, and the true SWE surface.

4.2 Datasets

Datasets Used

A list of SWE datasets was outlined and described in Sections 1.3 and 1.5. This chapter leverages SD and SWE data from SNOTEL, ASO, and UA SWE, as well as SWE data from the MY SWE product developed in Chapter 3. The results related to MY SWE presented in this chapter are derived from applying the full daily ablation RF model.

We also consider a coarsened version of ASO SD and SWE data, in order to match the spatial resolution of publicly available PRISM data and UA SWE. We generate these datasets by aggregating daily ASO data to a 4 km $\times$ 4 km spatial resolution. Within each 4 km $\times$ 4 km UA SWE grid cell, we average all 3 m SD ASO pixels to produce "Gridded ASO SD" and average all 50 m SWE ASO pixels to produce "Gridded ASO SWE" datasets. These observation-based gridded datasets are considered a well-informed representation of ground truth at the 4 km resolution for grid-level evaluation.

In sum, this chapter uses observational data from SNOTEL and ASO and gridded data from UA SWE, MY SWE, and Gridded ASO.

To align with ASO data availability in Colorado, the study area is the Colorado ASO basins, which are located within or immediately adjacent to the UCRB. A list of these basins during the 2016-2022 study period is presented in Table 4.1.

Basin Name	Flight Dates
Taylor River	03-30-2018
	04-08-2019
	06-09-2019
	02-20-2020
	04-21-2022
	05-25-2022
East River	03-31-2018
	05-24-2018
	04-07-2019
	06-10-2019
	02-14-2020
	04-21-2022
	05-18-2022
Blue River	04-19-2019
	06-24-2019
	04-19-2022
Crested Butte	04-04-2016
	03-30-2018
Conejos River	04-03-2016
	04-15-2022
Rio Grande River	04-03-2016
Grande Mesa	04-07-2019
	06-10-2019
Windy Gap	04-18-2022
Dolores	04-15-2022
Animas	05-15-2021

Table 4.1: List of basins in or adjacent to the UCRB surveyed by ASO. The associated flight dates and approximate basin area are provided.

Maps of Datasets

Below, the various datasets outlined above are illustrated. 3 m SD and 50 m SWE maps encompassing a select basin (specifically, the Taylor River basin) on select days (specifically, two dates within the 2022 water year) from the data products are shown in Figures 4.1-4.4. The 3 m or 50 m fine resolution ASO data are displayed alongside 4 km resolution Gridded ASO SD/SWE, 4 km SD/SWE from the UA SWE product, and 4 km SWE from the MY SWE product. The location of SNOTEL sites are indicated on the maps.

SD

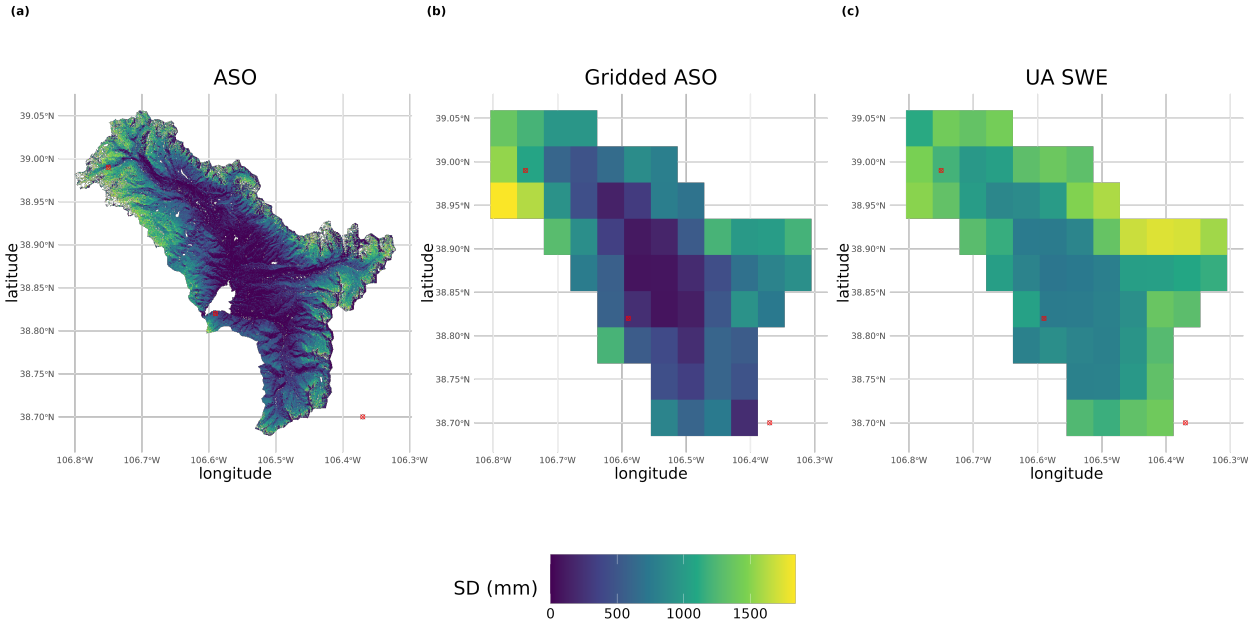


Figure 4.1: Taylor River basin on April 21, 2022. (a) 3 m ASO SD; (b) 4 km Gridded ASO SD; (c) SD from 4 km UA SWE.

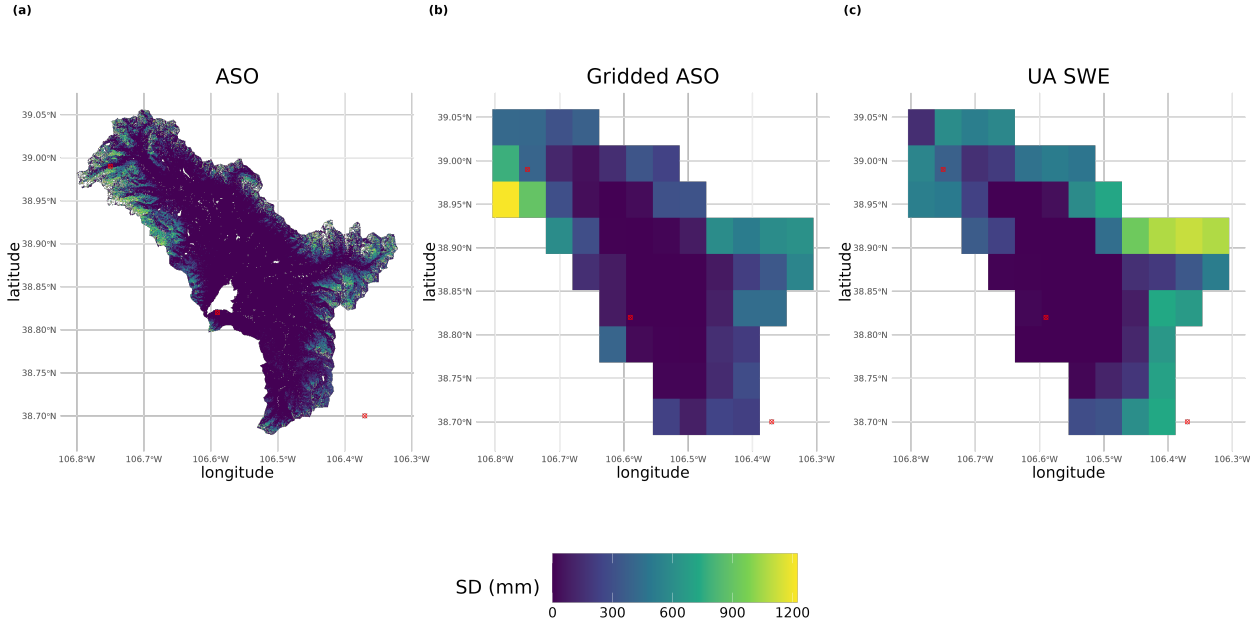


Figure 4.2: Taylor River basin on May 25, 2022. (a) 3 m ASO SD; (b) 4 km Gridded ASO SD; (c) SD from 4 km UA SWE.

SWE

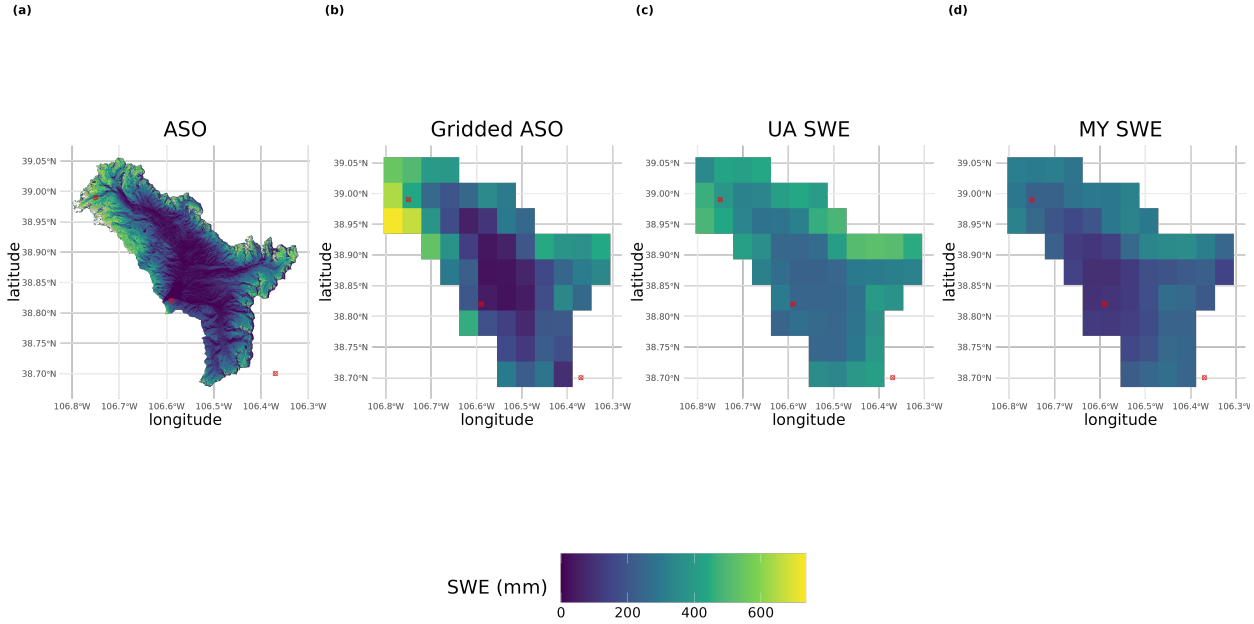


Figure 4.3: Taylor River basin on April 21, 2022. (a) 50 m ASO SWE; (b) 4 km Gridded ASO SWE; (c) SWE from 4 km UA SWE; (d) SWE from 4 km MY SWE.

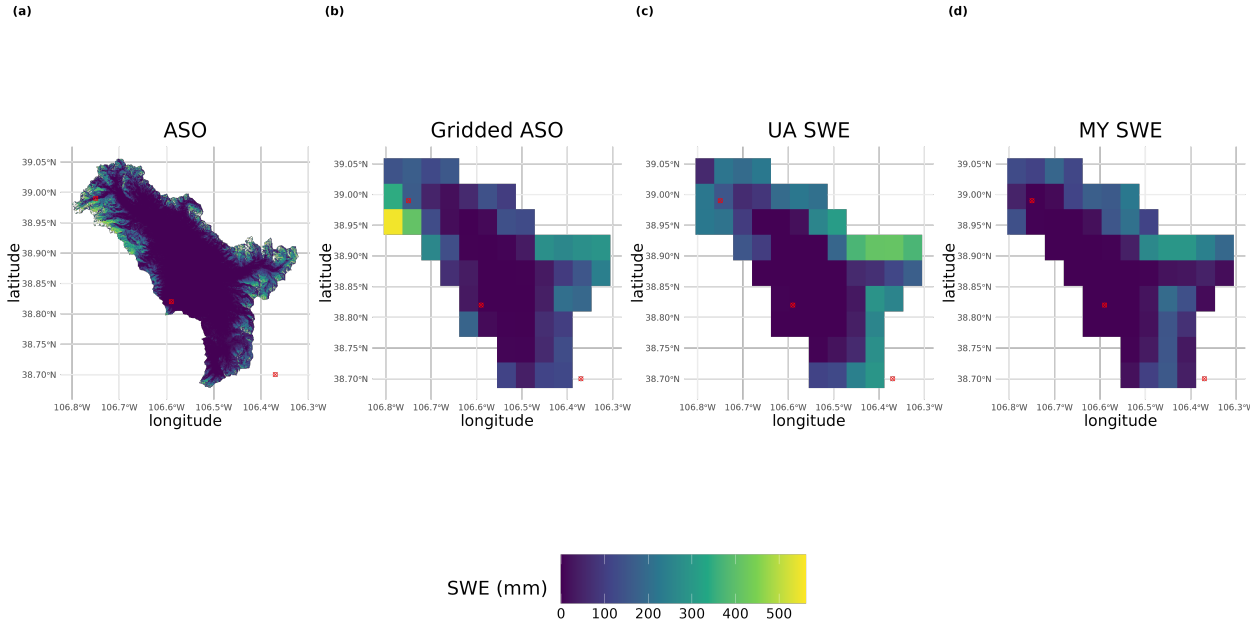
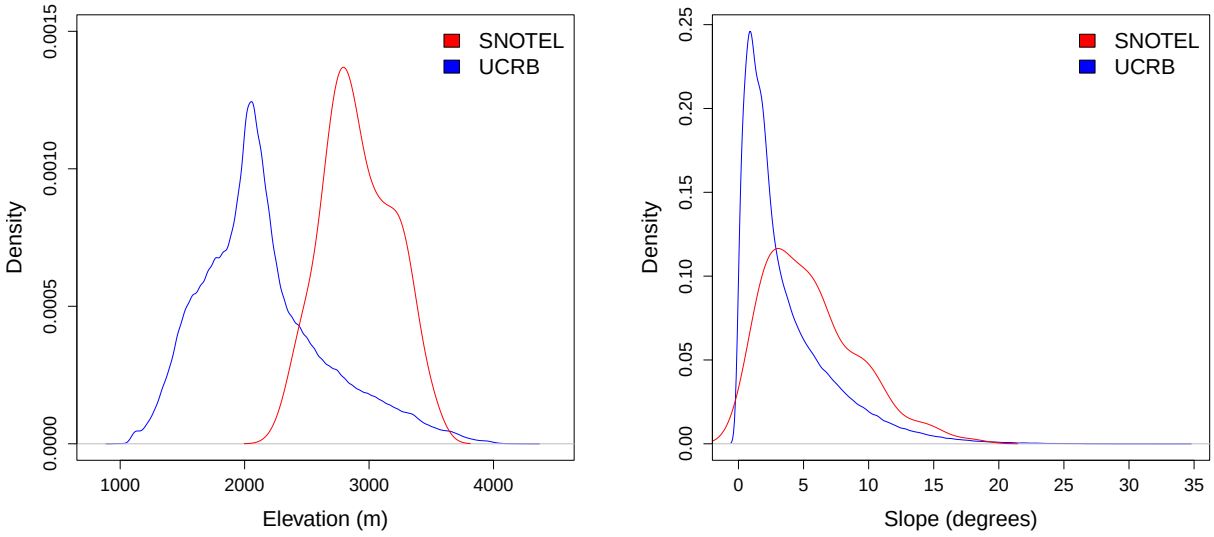


Figure 4.4: Taylor River basin on May 25, 2022. (a) 50 m ASO SWE; (b) 4 km Gridded ASO SWE; (c) SWE from 4 km UA SWE; (d) SWE from 4 km MY SWE.

4.3 Assessing Gridded SWE against Point Measurements

In this section, we conduct a point-level analysis of the various SD and SWE datasets. We start by comparing the distributions of SNOTEL elevations and slopes to the distributions of elevations and slopes in the UCRB and the ASO basins, identifying possible bias or skew in the representation of SNOTEL sites in the context of the expansive regions in which they reside. Then, we assess gridded SWE accuracy against observed SWE at SNOTEL sites located within ASO basins on available ASO flight dates. This analysis is carried out to prepare for a comprehensive grid-level evaluation in the subsequent section. We close out the section by outlining some challenges associated with comparing gridded estimates to point observations.

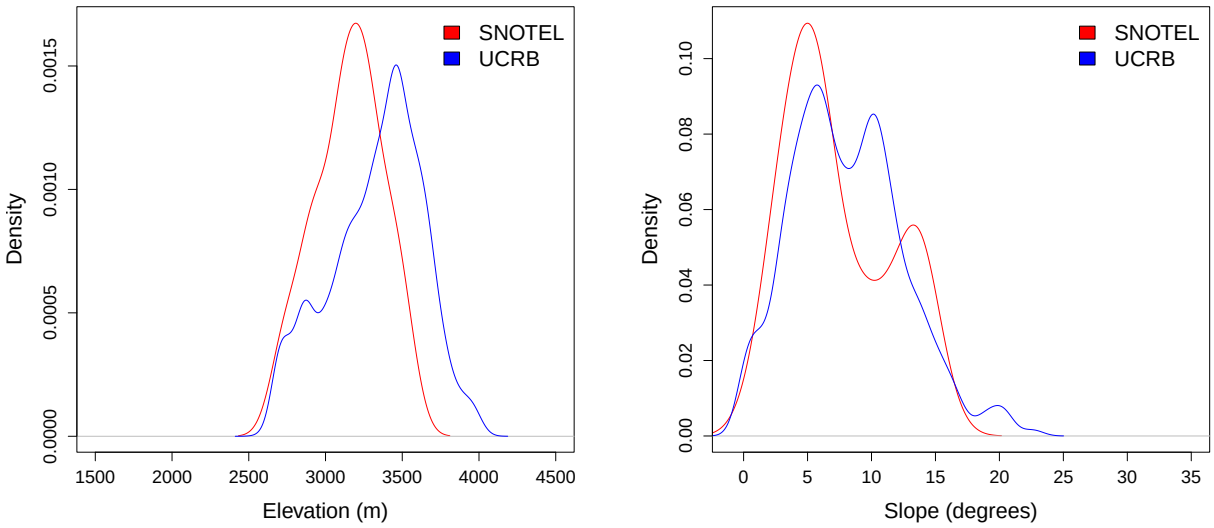
Distribution of SNOTEL site attributes in comparison with UCRB and ASO attributes



(a) Elevation

(b) Slope

Figure 4.5: Distribution of elevation and slope quantities of SNOTEL sites in the UCRB and all 30 m grid cells in the UCRB.



(a) Elevation

(b) Slope

Figure 4.6: Distribution of elevation and slope quantities of SNOTEL sites in the UCRB and all 4 km UCRB grid cells containing SNOTEL sites.

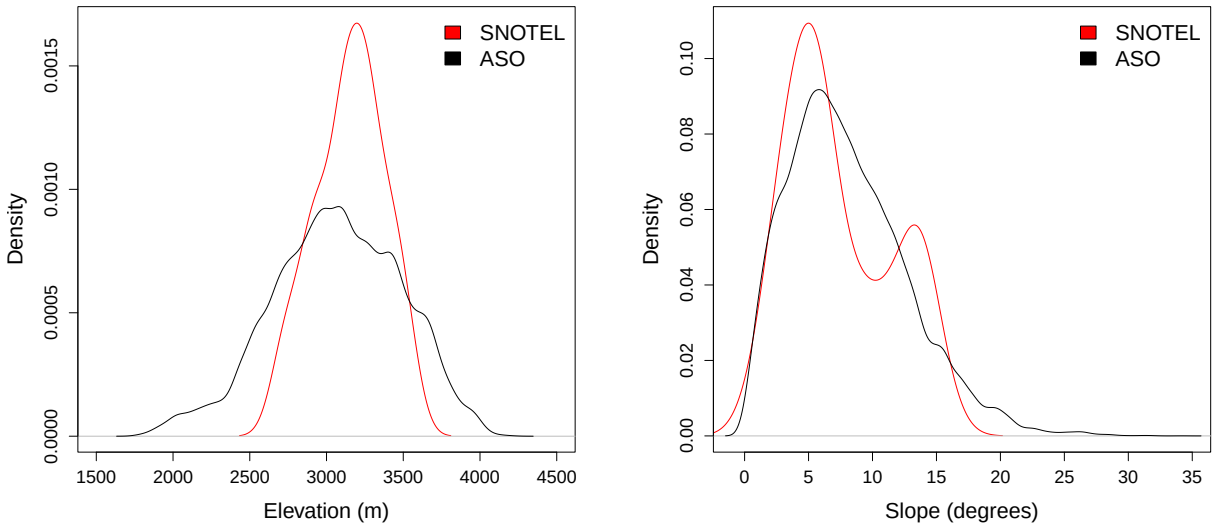
Figure 4.5 compares the distribution of (a) elevations and (b) slopes of SNOTEL sites in the UCRB against the total range of elevations and slopes represented by the UCRB (computed from the GTOPO30 DEM). SNOTEL sites represent higher elevation areas of the UCRB, as evidenced by the higher overall range of SNOTEL elevation values, as well as steeper areas, as evidenced by higher peak of SNOTEL slope values.

The reason the elevation distribution of SNOTEL is higher than the distribution of elevations in the UCRB is likely because in sampling, SNOTEL attempts to capture areas with more snow, which is most common at mid-high elevations. However, from a practical standpoint, it is difficult to implement and maintain equipment at the highest elevations.

The reason the slope distribution of SNOTEL congregates around a steeper slope value is likely because the UCRB covers a very large region, inevitably there are a lot of flat areas, which skews the distribution heavily toward zero slope, even though there are very steep areas.

Figure 4.6 compares the distribution of (a) elevations and (b) slopes of SNOTEL sites in the UCRB to that of 4 km $\times$ 4 km grid cells containing the respective SNOTEL sites. Comparing these distributions accentuates the topographic conditions of SNOTEL site locations within their more localized areas, providing insight on how representative SNOTEL sites are in snowy areas that may experience heterogeneity over short distances. We observe that within their respective local areas, SNOTEL sites are generally located at slightly lower elevations and on more gradual slopes, presumably due to practical difficulties in installing

and maintaining measurement instruments at very high elevations and on very steep terrain.



(a) Elevation

(b) Slope

Figure 4.7: Distribution of elevation and slope quantities of SNOTEL sites within Colorado ASO basins and all 30 m pixels in the Colorado ASO basins.

Figure 4.7 compares the distribution of (a) elevations and (b) slopes of SNOTEL sites within the ASO basins in relation to the distribution of elevations and slopes of all ASO grid cells across the Colorado ASO basins. The distribution of SNOTEL elevations in the ASO basins is more aligned with the distribution of ASO elevations than the distribution of SNOTEL elevations in the UCRB compared to the distribution of UCRB elevations for cells that contain SNOTEL sites (Figure 4.6), as exhibited by the two distributions peaking around the same elevation band of 3000 m to 3500 m. The elevations in the ASO basins as a whole span a wider range. Similar to in 4.6, the distribution of slopes of SNOTEL in the ASO basins is similar to that of slopes in the ASO basin as a whole, peaking around the same slope value, although the latter includes some steeper slopes. The longer tail of the ASO basins slope distribution suggests that there are steeper slope areas within the ASO basins at which SNOTEL sites are not located.

Point-scale comparison at SNOTEL sites within ASO basins

First, we evaluated the gridded SD and SWE predictions from the data products outlined in Section 4.2 against SD and SWE measured at SNOTEL sites within Colorado ASO basins on available flight dates. To carry out the grid-to-point comparison, for each SNOTEL site,

we extract data from the pixel (in the case of ASO) or grid cell (in the case of UA SWE, MY SWE, and Gridded ASO) containing the site. The evaluation exercise here differs from the site-level SWE evaluation in Section 3.4 in that only a small number of SWE values on noncontinuous dates at a partially overlapping subset of SNOTEL sites are being compared here, whereas the evaluation in Chapter 2 focused more on a climatological comparison of all sites across the UCRB. The study period is also different, in that here we focus on the years 2016 to 2022 containing ASO data, while Chapter 2 studied a 30-year period between the 1986 and 2016 water years.

For SD, we compare SNOTEL measured SD to 3 m ASO observed SD, 4 km Gridded ASO SD, and 4 km UA SWE predicted SD. For SWE, we compare SNOTEL measured SWE to 50 m ASO modeled SWE, 4 km gridded ASO SWE, 4 km UA SWE predicted SWE, 4 km MY SWE predicted SWE (with PRISM and GTOPO30 inputs), and MY SWE point-predicted SWE (with SNOTEL and GTOPO30 inputs).

In the evaluation results that follow, we stratify SNOTEL site-dates by time of year. In total, 30 out of the 43 total SNOTEL site-dates fall in the peak SWE period (defined as the earliest flight date in late February to the end of April) and the remaining 13 site-dates fall in the ablation period (defined as May 1 to the last flight date in late June).

SD

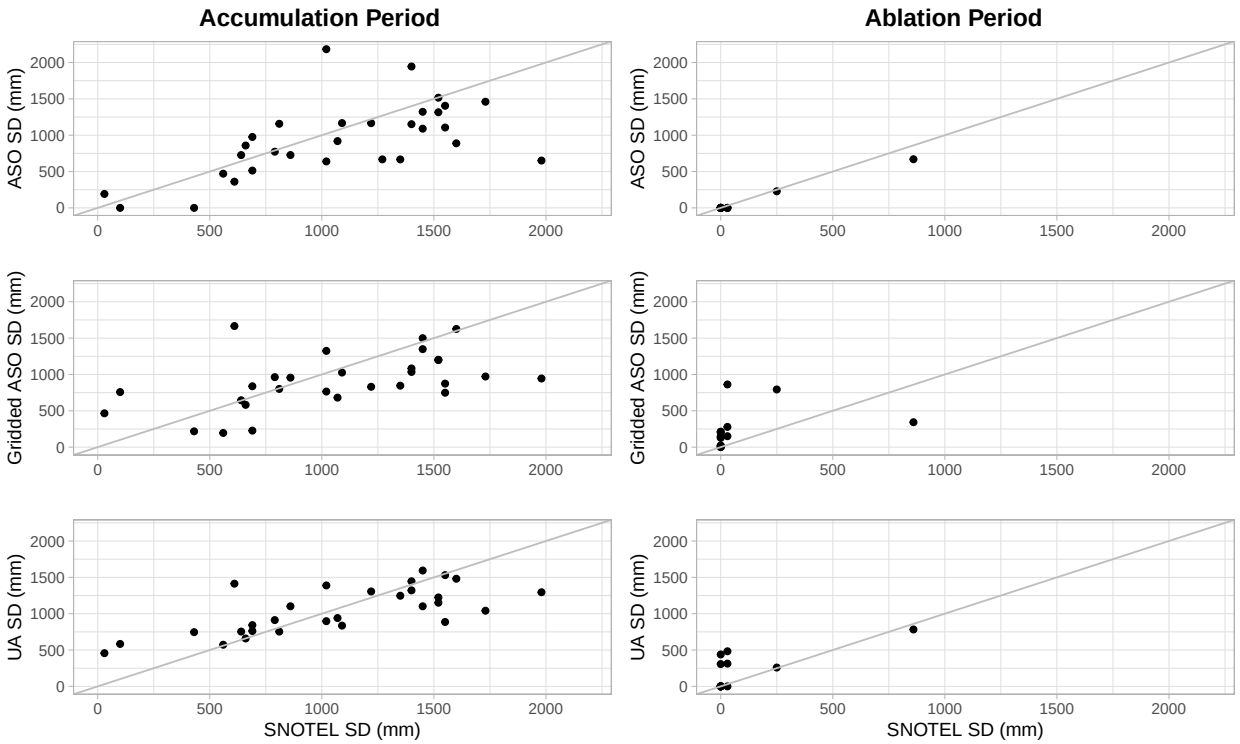
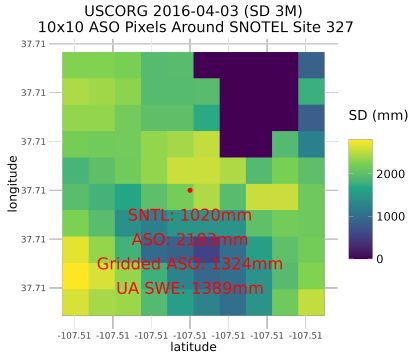
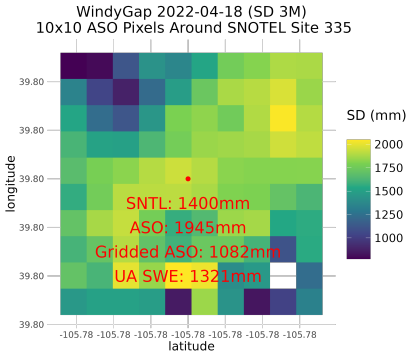


Figure 4.8: Scatterplots of SD from SNOTEL compared to ASO (top row), Gridded ASO (middle row), and UA SWE (bottom row) during the peak SWE period (left column) and the ablation period (right column).

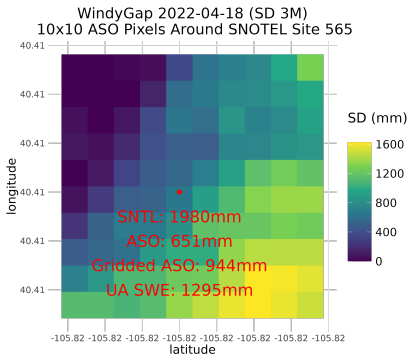
Comparing ASO measured SD to SNOTEL measured SD at SNOTEL sites (top panel of Figure 4.8), there is some agreement during the peak SWE period and very strong alignment during the ablation period. During the peak SWE period, there is generally a positive correlation where higher SNOTEL SD values correspond to higher ASO SD values; however, with larger SD values, we see greater deviation between the SNOTEL and ASO measurements. ASO SD measurements tend to be lower than SNOTEL SD measurements, especially as SD increases. During the ablation period, SNOTEL and ASO SD measurements are almost perfectly aligned, with both observing zero or near-zero SD on the same dates and positive SD values of similar magnitudes on dates with non-zero SD.



(a) The 100 (10 × 10) 3 m ASO SD pixels surrounding the pixel containing SNOTEL site 327 on April 3, 2016, a day in the peak SWE period.



(b) The 100 (10 × 10) 3 m ASO SD pixels surrounding the pixel containing SNOTEL site 335 on April 18, 2022, a day in the peak SWE period.



(c) The 100 (10 × 10) 3 m ASO SD pixels surrounding the pixel containing SNOTEL site 565 on April 18, 2022, a day in the peak SWE period.

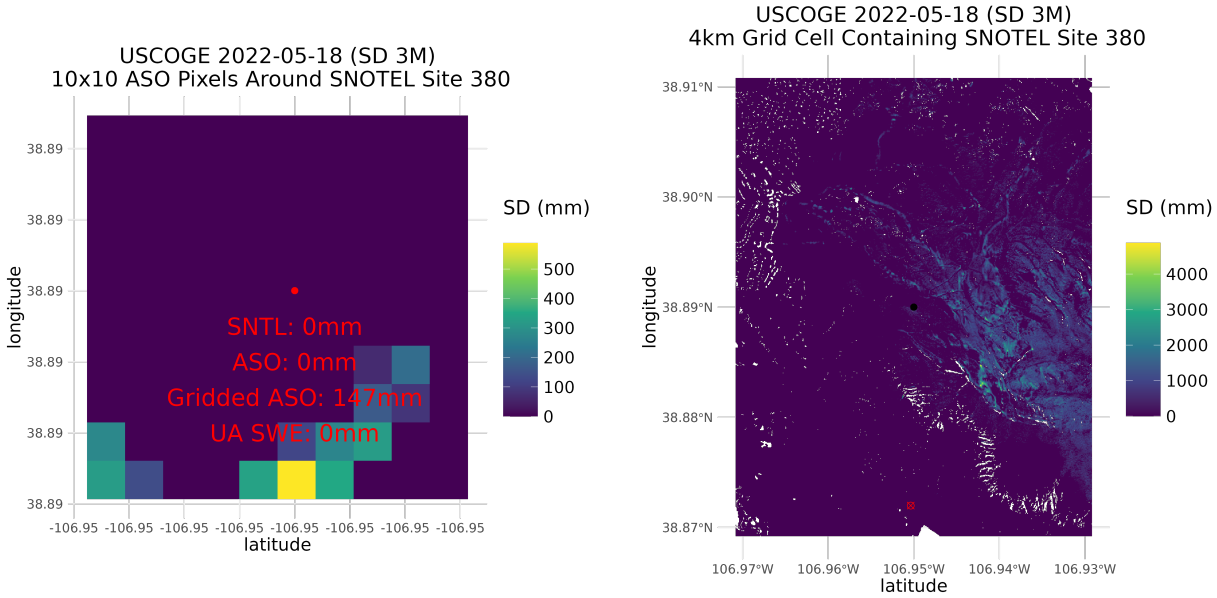
Figure 4.9: Example of regions of ASO basins overlaid by a selected SNOTEL site on dates in the peak SWE period, when ASO and SNOTEL at the point-level differ significantly.

There are several possible reasons for discrepancies observed between SNOTEL SD and ASO SD. First, ASO observes SD on a gridded surface via remote sensing technology, whereas SNOTEL SD is measured at single points by ultrasonic depth sensors. Figure 4.9 shows the distribution of observed SD, as measured by ASO, across ASO pixels surrounding SNOTEL sites where SNOTEL and ASO SD measurements diverge drastically. Seeing that there is heterogeneity across ASO pixels, factors such as wind redistribution, vegetation, and

topography can also contribute additional heterogeneity within each 3 m ASO pixel. A single SNOTEL point in a pixel may not correspond with the full distribution of SD quantities within the $9m^2$ area.

Furthermore, the data collection and subsequent processing and calibration procedures of ASO and SNOTEL differ completely, and may create bias, error, or differences in the reported SD quantities. The SNOTEL depth sensors measure SD by emitting ultrasonic waves toward the ground covered with snow, and using the time it takes for the waves to reflect back to calculate the distance to the snow surface (i.e., SD) [26]. While reported SD values are directly taken from the depth sensors with little to no processing, the sensors do face limitations in their data collection. The most common causes for erroneous measurements include sensor misalignment with the snow surface, poor reflection due to terrain roughness or unevenness, sound absorption by low-density ($< 5\%$) new snow, obstruction of the transducer by ice or debris, and presence of large or numerous snowflakes. Automated data quality checks are applied to the SNOTEL SD data to filter out values deemed implausible; however, errors falling within the range of possible SD values but are incorrect are unlikely to be detected and flagged [63]. ASO, on the other hand, measures SD by subtracting LiDAR data of gridded surface elevation on snow-off from that on snow-on days, followed by data processing and calibration steps that address missing data, vegetation, and alignment of gridded surfaces [18, 55]. These steps may at times introduce error into the snow depth observations. Additionally, temporal variability might play a role. SNOTEL SD is measured at various points throughout a day, but the daily reported quantity is the measurement taken at midnight [63]. Conversely, ASO aircrafts are flown during a seven hour interval. While this may not be a prominent driver for differences in SNOTEL and ASO SD on a number of days, this temporal difference could be a contributor, especially if a sizeable amount of accumulates or melts on a certain day over the course of several hours.

When SNOTEL measured SD is compared to Gridded ASO SD (middle panel in Figure 4.8), there continues to be non-negligible deviation between the two during the peak SWE period, but the variance in the predictions is smaller than when SNOTEL SD is compared to ASO SD. During the ablation period, however, Gridded ASO SD performs worse than ASO SD in matching SNOTEL SD. In particular, Gridded ASO SD overestimates zero or near-zero SNOTEL SD. This may be attributed to heterogeneity in SD in the pixels surrounding the SNOTEL sites, as illustrated by the pixels surrounding an example SNOTEL site in Figures 4.9 and 4.10. This heterogeneity is reflected in Gridded ASO because pixels surrounding the site are averaged, resulting in SD values higher than zero or near-zero SNOTEL SD.



(a) The 100 (10 × 10) 3 m ASO SD pixels surrounding the pixel containing SNOTEL site 380 on May 18, 2022, a day in the ablation period.

(b) 3 m ASO pixels in the 4 km × 4 km grid cell containing SNOTEL site 380 on May 18, 2022, a day in the ablation period.

Figure 4.10: Example of a region of an ASO basin (East River Basin) overlaid by a selected SNOTEL site on a date in the ablation period. (a) The 10 × 10 ASO pixels surrounding the SNOTEL site, and (b) the 4 km × 4 km grid cell containing the SNOTEL site. Generally, there is less heterogeneity in ASO pixels surrounding SNOTEL sites, especially during the ablation season.

Comparing SD predicted by the UA SWE algorithm to SNOTEL measured SD (bottom panel in Figure 4.8), during the peak SWE period, the UA SWE-modeled SD tends to overestimate SNOTEL SD at lower SD values and underestimate at higher SD values. At SNOTEL sites that experience no snow (i.e., zero SD) during the ablation period (when all snow has melted), SD predicted by UA SWE tends to overestimate. This overestimation in the ablation period, in conjunction with SD alignment in the peak SWE period, suggests that the ablation rate predicted by UA SWE may be too slow.

SWE

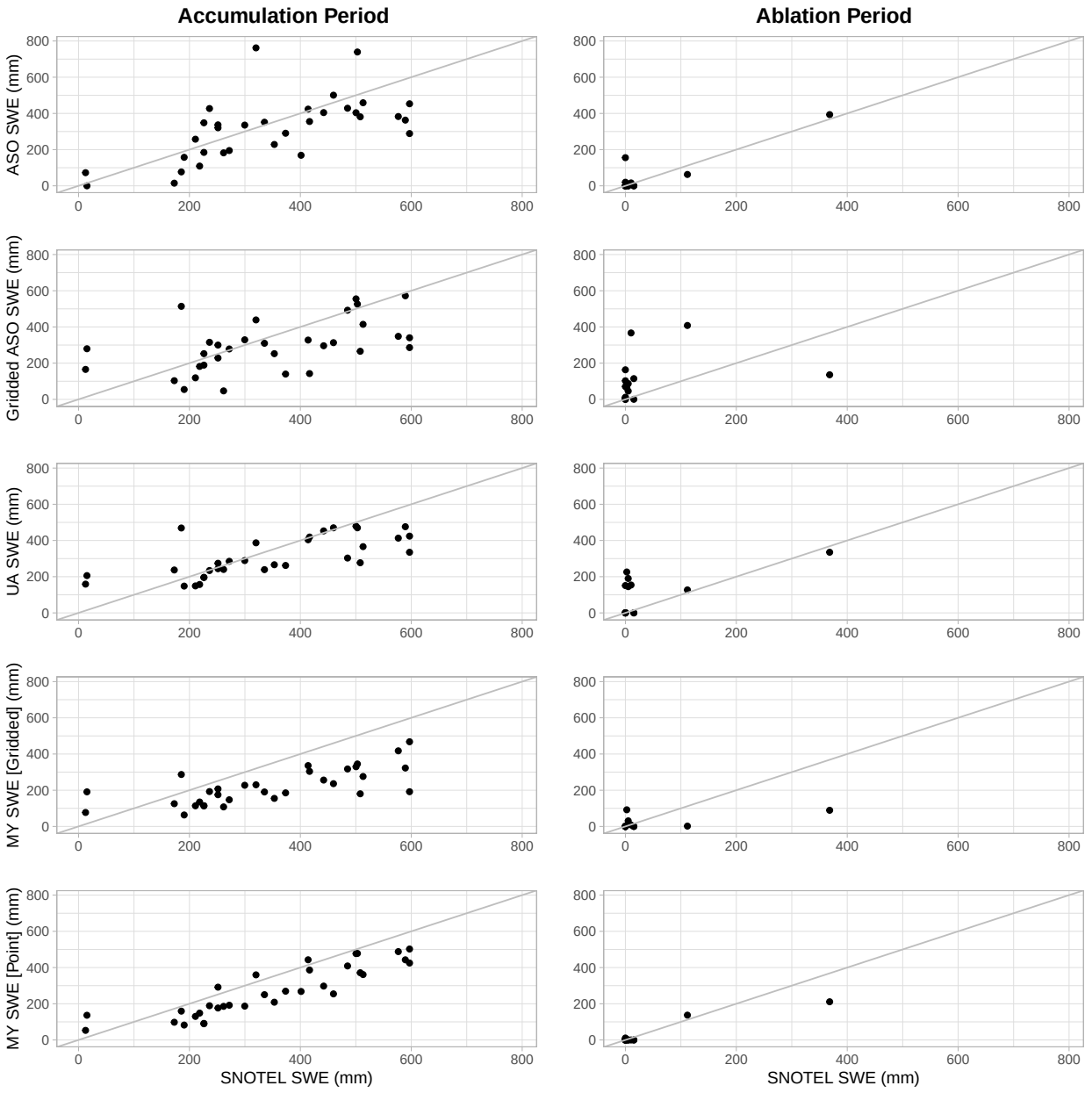


Figure 4.11: Scatterplots of SWE from SNOTEL compared to ASO (top row), Gridded ASO (second row), UA SWE (third row), MY SWE with gridded PRISM meteorological inputs (fourth row), and MY SWE with point-level SNOTEL meteorological inputs (bottom row) during the peak SWE period (left column) and the ablation period (right column).

At SNOTEL sites, ASO-estimated SWE and SNOTEL-measured SWE (top row panels of Figure 4.11) are positively correlated during both the peak SWE and ablation periods. For flight dates in the peak SWE period, the positive agreement does not exhibit a heteroskedas-

tic pattern, unlike in the SD case. Aside from two noticeable outliers where ASO SWE is significantly larger than SNOTEL SWE, the difference between ASO SWE and SNOTEL SWE is generally greater when ASO SWE is lower than SNOTEL SWE than when ASO SWE is higher than SNOTEL SWE. Moreover, there are more site-dates on which ASO SWE is lower than SNOTEL SWE. For flight dates in the ablation period, ASO SWE and SNOTEL SWE are generally in agreement, both in terms of zero or near-zero SWE and positive SWE.

When SNOTEL SWE is compared to Gridded ASO SWE (second row panels in Figure 4.11), the estimates during the peak SWE period exhibit a similar pattern to that of SNOTEL SWE compared to ASO SWE. However, the outliers that were observed in ASO SWE are no longer present and there is a slightly more noticeable pattern of general underestimation. This may be explained by the averaging procedure — that averaging means incorporating more lower SWE values, due potentially to including more lower elevation pixels. Gridded ASO SWE in the ablation period is less well aligned to SNOTEL SWE, with Gridded ASO estimating positive SWE on many days where SNOTEL observes zero or near-zero SWE. This is also due to the averaging procedure — when considering an entire grid cell, it becomes likely that at least a few pixels within the grid cell will experience persisting snow even if most of the snow has melted in the grid cell, thus increasing the average SWE above zero.

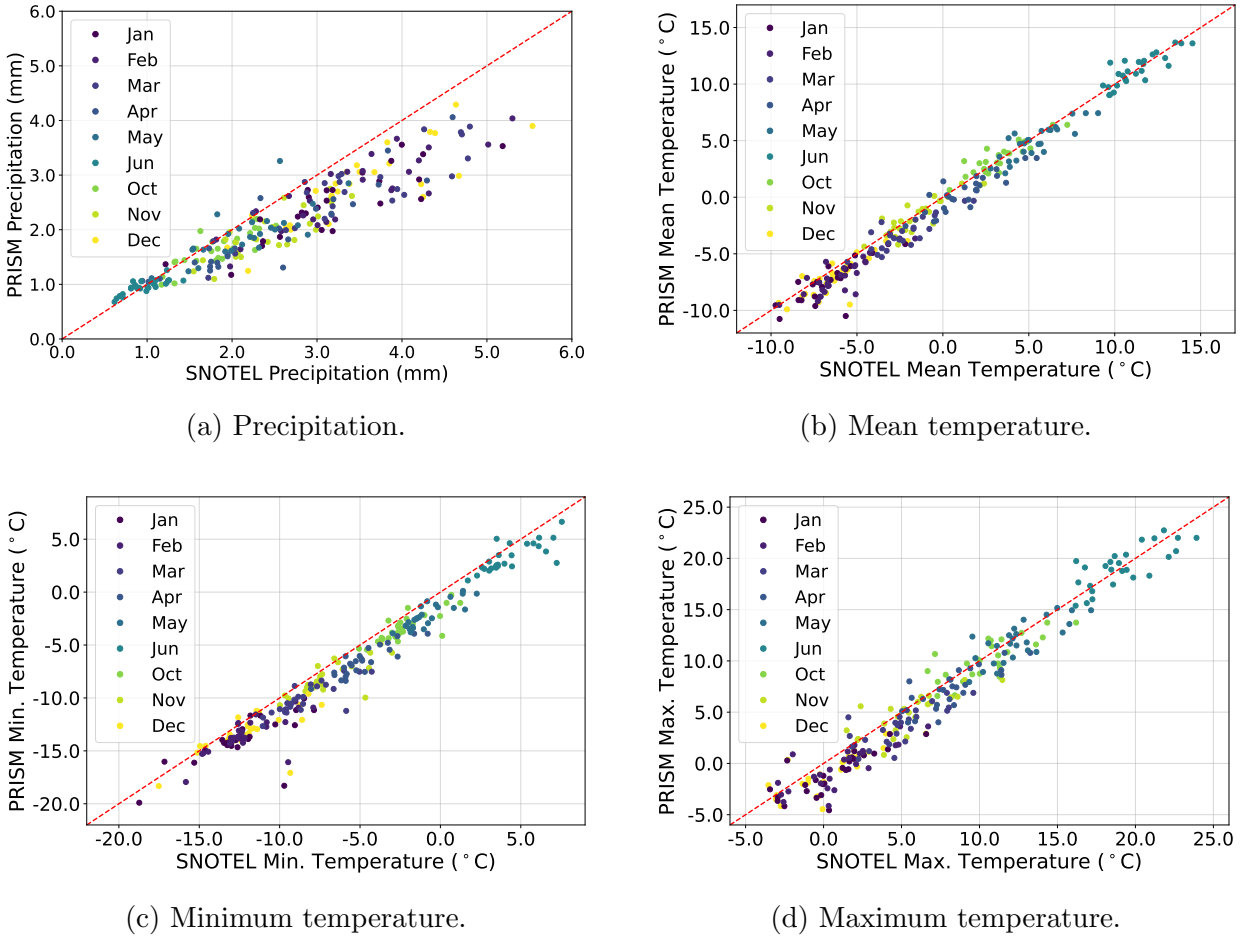
The deviation between UA SWE predicted SWE and SNOTEL measured SWE (third row panels in Figure 4.11) is less pronounced than ASO SWE compared to SNOTEL SWE during the peak SWE period. In general, UA SWE underestimates SNOTEL SWE. During the ablation period, UA SWE predicted SWE has a tendency to overestimate zero or near-zero SNOTEL SWE.

MY SWE with gridded PRISM meteorological inputs also generally underestimates SNOTEL SWE during the peak SWE period (fourth row panels in Figure 4.11). The magnitude of underestimation is larger than that exhibited by UA SWE. During the ablation season, MY SWE with gridded PRISM meteorological inputs is generally aligned with zero and near-zero SNOTEL SWE with the exception of a couple site-dates, while significantly underestimating positive SWE.

When MY SWE is generated with point-level SNOTEL meteorological input data rather than gridded PRISM meteorological input data (bottom row panels contrasted with fourth row panels in Figure 4.11), the positive correlation with SNOTEL SWE becomes stronger in both the peak SWE and ablation periods. During the peak SWE period, the underestimation of MY SWE persists even when site-level meteorological input data is used; however, the magnitude of underestimation is lower than when MY SWE is generated with gridded PRISM meteorological input data. Notably, MY SWE with SNOTEL meteorological inputs shows a more consistent pattern of underestimation relative to SNOTEL SWE than UA SWE does, with a more uniform bias. This pattern is similar to that seen in Figures 3.14 and 3.15, which illustrate that for the daily ablation RF model, the climatological mean of MY SWE underestimates SNOTEL SWE to a greater extent than UA SWE during the peak SWE period, but the climatological standard deviation of MY SWE aligns more closely with that of SNOTEL SWE. Therefore, MY SWE with SNOTEL meteorological inputs shows a stronger

systematic low bias compared to UA SWE, but the spread of the MY SWE predictions better reflects the variability in SNOTEL SWE. During the ablation period, MY SWE generated using site-level input meteorological input data more accurately captures zero or near-zero SNOTEL SWE values than MY SWE generated using gridded input meteorological input data as well as UA SWE.

The improvement in alignment of MY SWE derived from point SNOTEL meteorological input data over MY SWE derived from gridded PRISM meteorological input data with SNOTEL SWE may be attributed to several possible reasons. One, PRISM meteorological data generally underestimates SNOTEL meteorological data at the site-level. This is evidenced in Figure 4.12 (as well as in contrasting Figure 3.18 against Figure 3.14), which shows that for precipitation and the temperature variables, PRISM tends to underestimate during most months at most SNOTEL sites in the ASO basins. This may be attributed to the average precipitation and/or temperature within a larger grid area being lower than measurements at a single in-situ site, influenced by the site's specific topographic placement. Second, the interpolation and averaging procedures involved in producing gridded meteorological datasets, such as PRISM, may lead to biased predictions. Third, the discrepancy may be due to a fundamental scale-related difference underlying point- and grid-level predictions.



(a) Precipitation.

(b) Mean temperature.

(c) Minimum temperature.

(d) Maximum temperature.

Figure 4.12: Comparison of SNOTEL and PRISM meteorology at SNOTEL sites within ASO basins. Each dot represents a monthly average over the study period for a SNOTEL site.

Challenges and issues associated with grid-to-point comparisons

Evaluating gridded data against point-level measurements presents several challenges primarily related to differences in spatial representation, scale, and variability of the two types of data sources.

By definition, gridded data are spatially continuous, representing values across grid cells, while point-level measurements provide localized information at specific locations. Environmental processes also naturally exhibit spatial heterogeneity, especially in regions with complex terrain or land cover variability. The spatial averaging inherent in gridded data is likely to smooth out fine-scale variations and heterogeneity that point-level measurements are better able to capture at select locations, leading to discrepancies between the two types of data.

Moreover, datasets of different resolutions, including gridded datasets and point-level measurements, often represent underlying physical processes at different scales, rendering direct comparison unsuitable.

Point-level measurements are further subject to uncertainties and variability inherent in the measurement process, such as instrumentation errors, sampling biases, and environmental variability. Gridded datasets may also contain uncertainties arising from data assimilation, interpolation techniques, and model assumptions. Directly comparing point-level measurements to gridded data without accounting for these uncertainties can lead to misleading conclusions about the accuracy and reliability of the gridded dataset.

In light of these issues that arise when comparing gridded estimates, such as the UA SWE gridded product and MY SWE generated with gridded input data, against point measurements, such as SNOTEL, we seek a more appropriate reference for evaluating gridded estimates. Specifically, we define the gridded reference as Gridded ASO SWE, which, as described in Section 4.2, is the integral (average) over ASO pixels in each corresponding 4 km $\times$ 4 km PRISM grid. Recall that ASO measures SD at a 3 m spatial resolution and applies a model to estimate SWE at a 50 m spatial resolution, so Gridded ASO SD and SWE are derived by averaging all 3 m pixels and 50 m pixels, respectively, within each 4 km grid cell. This allows us to better conduct one-to-one comparisons of UA SWE or MY SWE against an observation-based gridded dataset. Note that the point-level mismatch between SNOTEL and ASO does raise some concern about using Gridded ASO SD and SWE as ground truth for gridded evaluation.

4.4 Assessing Gridded SWE against Gridded Observations

In this section, we evaluate UA SWE and MY SWE against Gridded ASO at each 4 km $\times$ 4 km grid cell within ASO basins on available flight dates. We assess model predictions based on SWE within each grid cell, total snow water volume in each ASO basin, and spatial variability across ASO basins. Since MY SWE is only available for SWE and not SD, we will only show results for SWE in this section.

SWE

First, we evaluate model-predicted SWE within each grid cell against Gridded ASO SWE. Figure 4.13 stratifies grid cells based on elevation and Figure 4.14 stratifies grid cells by month.

According to Figure 4.13, UA SWE and MY SWE both tend to overestimate Gridded ASO SWE at the lowest elevation band (2000 m to 2500 m). At the second lowest elevation band (2500 m to 3000 m), UA SWE tends to overestimate Gridded ASO SWE, but at grid cells where UA SWE underestimates, the magnitude of the underestimation is more drastic. On the other hand, the distribution of under- and overestimation of MY SWE

is more balanced. In the two highest elevation bands (3000 m to 3500 m and 3500 m to 4000 m), there is a notable number of grid cells at which UA SWE and MY SWE both underestimate Gridded ASO SWE, with the latter underestimating Gridded ASO SWE by a greater magnitude.

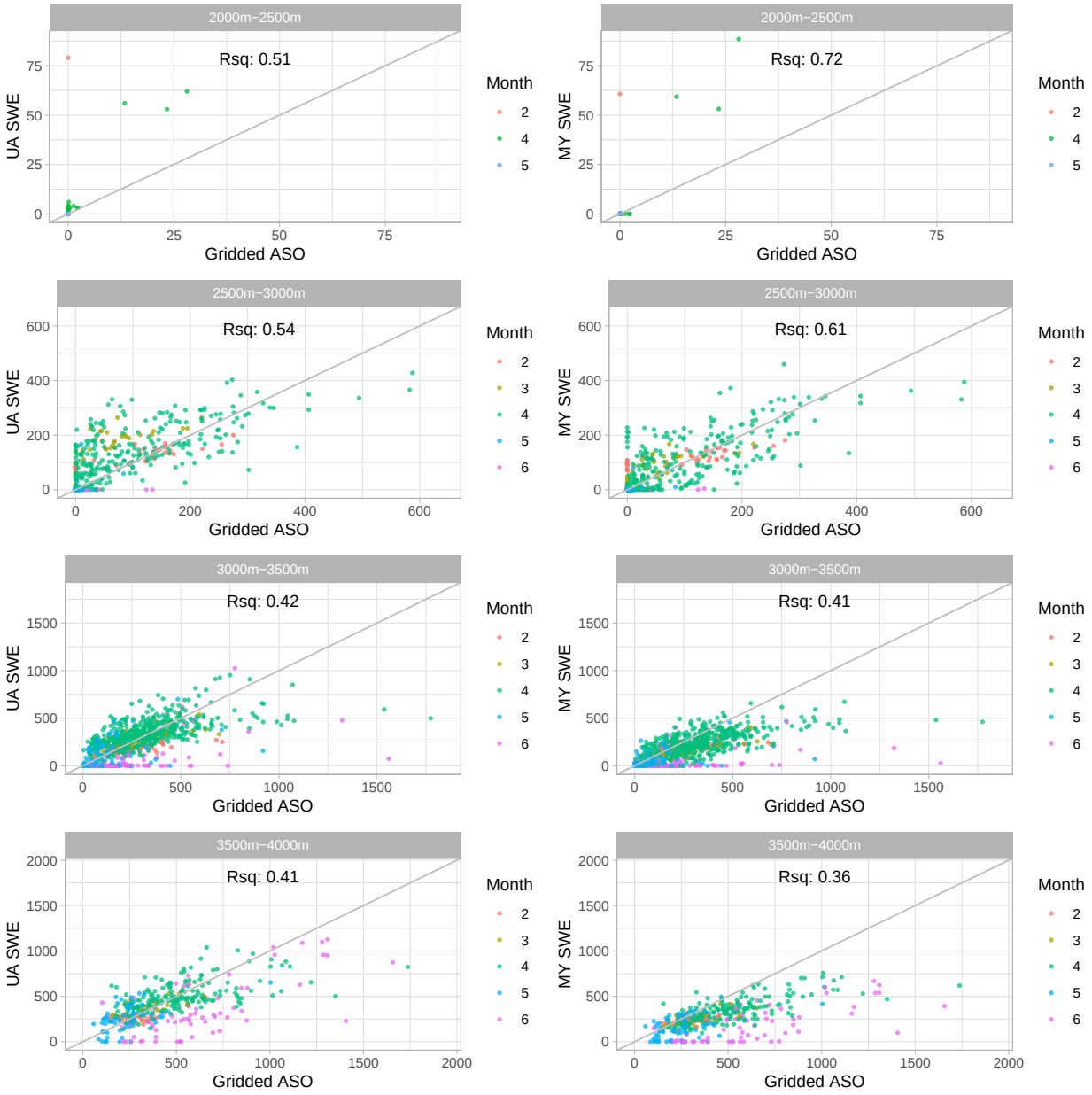


Figure 4.13: Scatterplots of Gridded ASO SWE compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, arranged in ascending order from top to bottom rows. Each dot in each panel corresponds to a grid cell in an ASO basin on a flight date.

When stratifying solely by month (Figure 4.14), we observe that UA SWE and MY SWE on the February flight dates tend to underestimate Gridded ASO SWE. UA SWE in March tends to overestimate Gridded ASO SWE, while MY SWE tends to underestimate Gridded

ASO SWE, especially at grid cells exhibiting larger amounts of SWE according to Gridded ASO. In April, the under- and overestimation of UA SWE to Gridded ASO SWE is more evenly distributed, even though the magnitude of overestimating values tends to be greater than the magnitude of underestimating values at lower Gridded ASO SWE values and the magnitude of overestimating values is lower than the magnitude of underestimating values at larger Gridded ASO SWE values. The spread in MY SWE predictions is more narrow, with more MY SWE values underestimating Gridded ASO SWE, especially at larger Gridded ASO SWE values. On May flight dates, the spread in UA SWE increases as Gridded ASO SWE increase, while the spread in MY SWE grid cell values is smaller, as are the predicted avlues in general when compared to Gridded ASO SWE. In June, UA SWE tends to underestimate Gridded ASO SWE, as does MY SWE, with the latter exhibiting more underestimation, in terms of quantity and magnitude, especially when Gridded ASO SWE is larger.

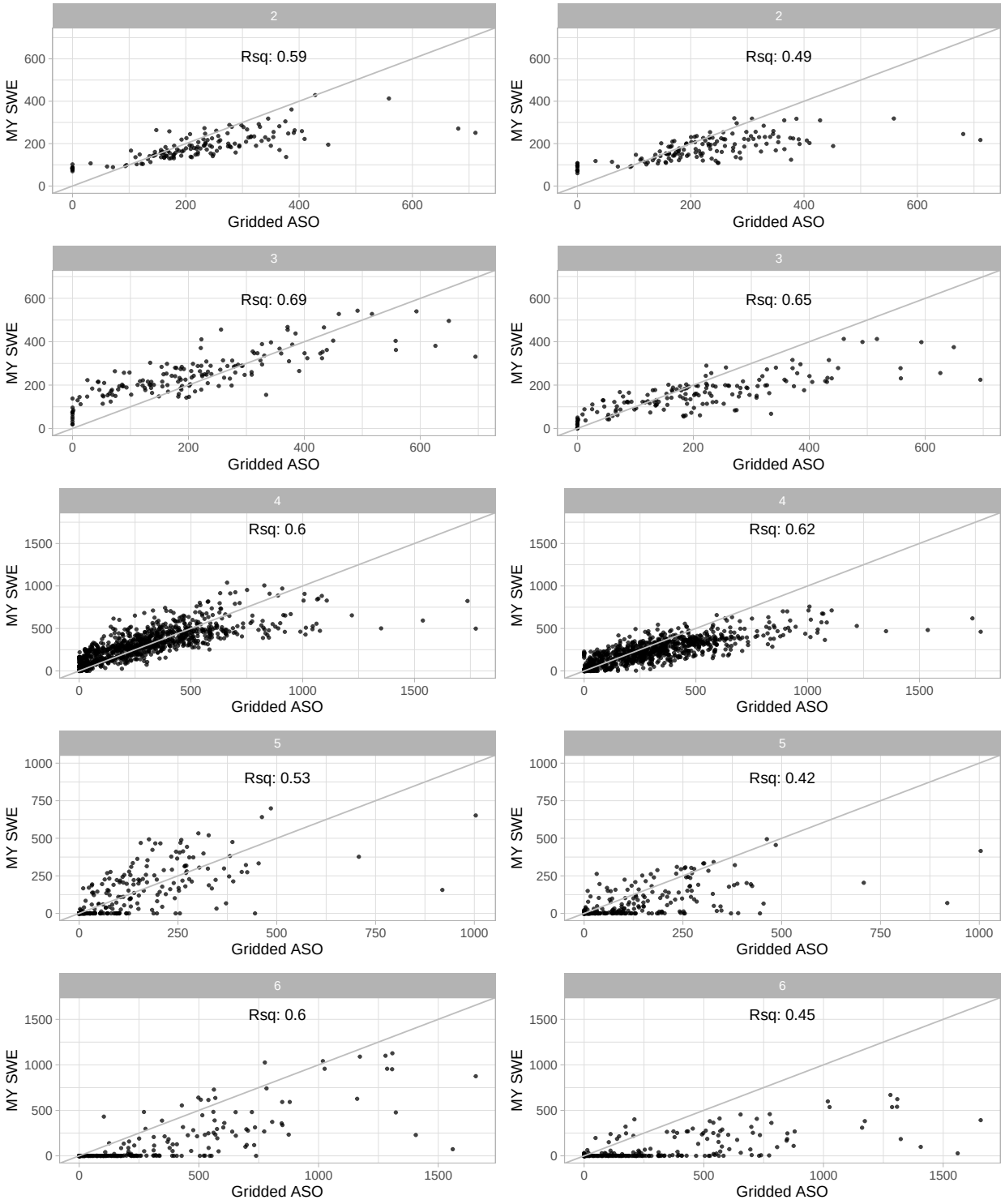


Figure 4.14: Scatterplots of Gridded ASO SWE compared to UA SWE (left column) and MY SWE (right column) stratified by month. Each dot in each panel corresponds to a grid cell in an ASO basin on a flight date.

A notable pattern in both Figures 4.13 and 4.14 is that there are clusters of dots along the x = 0 and y = 0 lines, meaning that there is a significant number of grid cells at which Gridded ASO has zero SWE while UA SWE and/or MY SWE predict non-zero SWE, and vice versa.

To try to understand this behavior, we first observe, from Figure 4.13, that at lower elevations (2000 m to 3000 m), there is a noticeable number of grid cells in which Gridded ASO SWE predicts full snowmelt while UA SWE predicts persisting snowpack (ASO = 0, UA SWE > 0). This same pattern is observed with MY SWE in the 2500 m to 3000 m elevation band, but the opposite pattern (i.e., ASO > 0, MY SWE = 0) is observed at the lowest 2000 m to 2500 m elevation band (although not many grid cells fall within this elevation band). At mid-high elevations (2500 m to 4000 m), there are noticeable occurrences of Gridded ASO predicting persisting snowpack while UA SWE and MY SWE both predict full snowmelt (ASO > 0, UA SWE = 0, MY SWE = 0).

According to Figure 4.14, during the peak SWE season (February-April), there is a noticeable number of grid cells in which Gridded ASO predicts full snowmelt while UA SWE and MY SWE predict persisting snowpack (ASO = 0, UA SWE > 0, MY SWE > 0). During the ablation season (May-June), there is a noticeable number of grid cells in which Gridded ASO predicts continued snowpack while UA SWE and MY SWE predict full snowmelt (ASO > 0, UA SWE = 0, MY SWE = 0).

Combining these two observations, we conclude that at lower elevations earlier in the season (during peak SWE season), Gridded ASO predicts zero SWE in some grid cells while UA SWE and MY SWE predict non-zero SWE. At mid to high elevations, especially later in the season (during the ablation period), Gridded ASO predicts that some grid cells still contain snow while UA SWE and MY SWE predict that all snow has melted.

This makes sense intuitively because at lower elevations, we would expect less snow in total, even during the peak SWE period. It suggests that UA SWE and MY SWE predict too much accumulation or too little ablation in these locations. As the overestimating UA SWE and MY SWE grid cells approach the ablation season, both catch up and begin to approach zero (which is why there is no longer a trend of UA SWE > 0, MY SWE > 0, ASO = 0 in the ablation period — UA SWE becomes zero). During the ablation season, however, other grid cells that previously experienced non-zero UA SWE/MY SWE and non-zero Gridded ASO SWE begin to experience a difference in timing of approaching zero SWE — with UA SWE and MY SWE reaching zero SWE before Gridded ASO does. This trend is especially true at higher elevations.

Overall, the patterns along the x- and y-axes are a reflection of difference in timing of reaching full ablation. At lower elevations, Gridded ASO predicts full ablation earlier than UA SWE and MY SWE. Whereas, at mid to higher elevations, UA SWE and MY SWE predict full ablation earlier than does Gridded ASO.

On the other hand, there are also some non-zero outliers in the UA SWE and MY SWE predictions when compared against Gridded ASO SWE. In particular, throughout all time periods (months), UA SWE and MY SWE noticeably underestimate SWE at grid cells that experience higher Gridded ASO SWE, with the underestimation from MY SWE being more

pronounced (Figure 4.14). At higher elevations (3000 m to 4000 m), there is also a noticeable number of grid cells at which UA SWE and MY SWE both underestimate Gridded ASO SWE, with MY SWE once again underestimating at a greater magnitude (Figure 4.13). Putting these two observations together, we conclude that at higher elevations when Gridded ASO exhibits higher SWE values, UA SWE and MY SWE both tend to underestimate. A probable explanation rests in the constrained representation of SNOTEL sites: since SNOTEL sites are sparse at higher elevations, gridded SWE predictions derived from SNOTEL (i.e., UA SWE and MY SWE) experience greater difficulty in accurately predicting higher SWE values at higher elevations (since presumably SWE is generally even higher at higher elevations). The limited coverage of station data at higher elevations may also impact the accuracy of point and gridded precipitation and temperature data. Gridded products such as PRISM rely on both station and remote sensing data, and sparse representation of station data in higher elevation areas can pose challenges for accurately interpolating these data. Errors in precipitation and temperature data at higher elevations are likely to propagate into UA SWE and MY SWE predictions, as these variables are key inputs to both models.

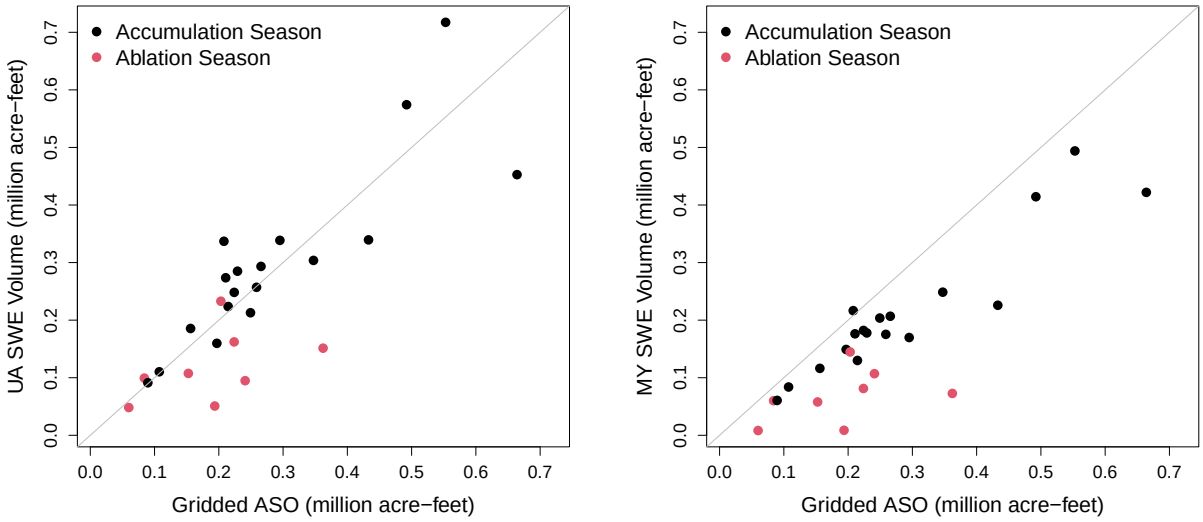
Note that due to the lack of spatial and temporal continuity in ASO basin locations and flight dates (as well as years), the interpretation of 4.13 and 4.14 discussed above is subject to limitations.

Total snow water volume

To provide a more comprehensive assessment of SWE model predictions, we also evaluate, for each basin, the total water volume estimated by UA SWE and MY SWE against Gridded ASO SWE. To compute total water volume (in million acre-feet) in a basin from gridded SWE observations or predictions (in mm), we follow these steps:

1. For each grid cell, convert SWE from mm to m.
2. Compute the area of each grid cell, in m^2.
3. For each grid cell, multiply total SWE in m (Step 1) by the grid cell area in m^2 (Step 2). This results in water volume of each grid cell in cubic meters (m^3).
4. Sum water volumes across all grid cells to obtain total basin water volume in cubic meters.
5. Convert total basin water volume from cubic meters to million acre-feet, by multiplying the total basin water volume in cubic meters by 0.000810714 and then by 1,000,000.

Figure 4.15 summarizes the total water volume per basin derived from Gridded ASO SWE compared to (a) UA SWE and (b) MY SWE.



(a) Gridded ASO compared to UA SWE

(b) Gridded ASO compared to MY SWE

Figure 4.15: Total water volume of each basin-date according to Gridded ASO SWE versus UA SWE and MY SWE. Each dot in the scatterplots correspond to a basin-date at which an ASO flight took place. The black dots correspond to basin-dates in the peak SWE season and the red dots correspond to basin-dates falling in the ablation season.

With regard to water volume derived from UA SWE in comparison to water volume derived from Gridded ASO SWE, there is decent overall alignment, especially in basin-dates experiencing lower total water volume. Basin-dates where UA SWE underestimates Gridded ASO SWE have a larger magnitude of difference compared to basin-dates where UA SWE overestimates Gridded ASO SWE. The underestimation also tends to correspond more to basin-dates in the ablation period (May-June) than in the peak SWE period (February-April).

On the other hand, with the exception of one basin-date, total water volume derived from MY SWE categorically underestimates total water volume derived from Gridded ASO SWE. Similar to in the UA SWE case, the magnitude of underestimation tends to be more pronounced in basin-dates that fall within the ablation period (May-June) than those that fall in the peak SWE period (February-April). Basin-dates that observe higher total water volume, according to Gridded ASO SWE, are also underestimated to a greater degree by MY SWE.

Spatial variability

To assess the spatial variability underlying the Gridded ASO SWE, UA SWE, and MY SWE fields, we employ the use of variograms. Variograms measure the degree of spatial depen-

dence between pairs of data points as a function of their separation distance. The x-axis of a variogram signifies the lag distance, indicating the spatial distance (i.e., separation distance) between data points. The y-axis represents semivariance, a measure that quantifies the variability between data points that are separated by certain distances. A higher semivariance indicates greater variability between data points separated by the given distance. In Figure 4.16, distances between pairs of grid cells within each ASO basin are aggregated into a single diagram and then an exponential theoretical variogram is fit for each gridded SWE dataset.

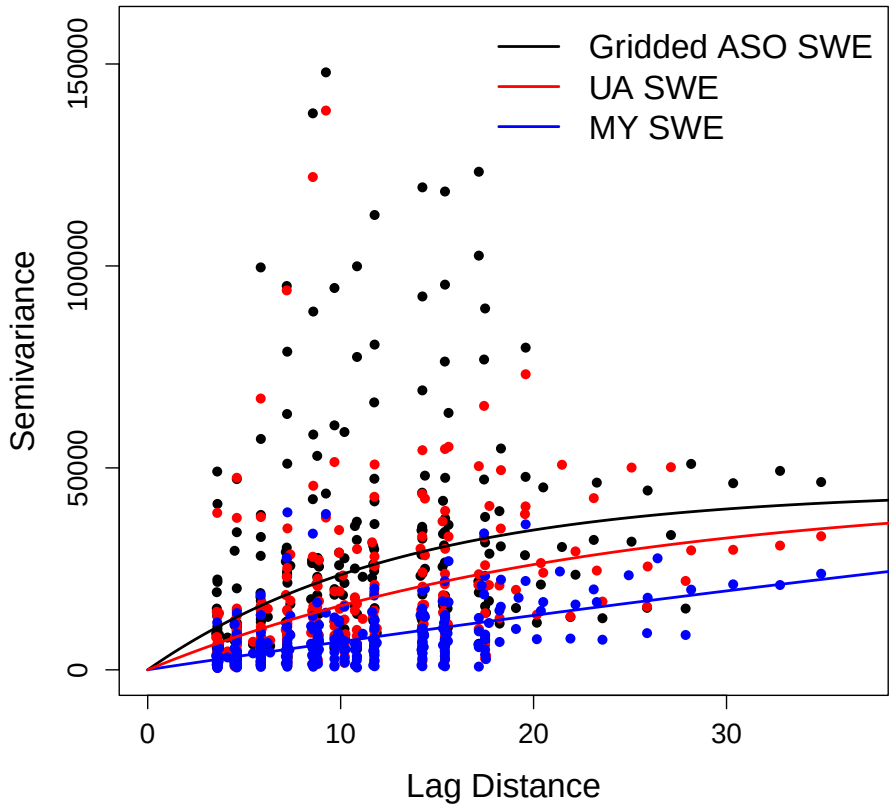


Figure 4.16: Comparison of variograms of Gridded ASO SWE, UA SWE, and MY SWE, computed from grid cells aggregated across all basins.

Both MY SWE and UA SWE consistently demonstrate lower variability compared to Gridded ASO SWE across all non-zero lag distances in the variograms shown in Figure 4.16. The variograms suggest that UA SWE and MY SWE show more spatially smooth surfaces.

4.5 Discrepancy Between Gridded SWE and True SWE

As outlined and discussed in previous sections, there exist discrepancies between SWE predictions from algorithms such as UA SWE and MY SWE and observed SWE, such as SNOTEL in-situ SWE measurements and ASO LiDAR-derived SWE. These disparities underscore limitations of empirically-derived gridded datasets in fully capturing complexities of the true SWE surface. In this section, we explore potential factors contributing to why gridded MY SWE may not be fully representative of true SWE. When possible, and for additional context, UA SWE is also included in discussions.

There are several characteristics in the training and gridding procedures of the MY SWE methodology that may play a part in affecting its accuracy in representing true SWE. One potential driver is the difference in scale of some of the key training and prediction data. To determine the optimal model parameters, MY SWE leverages SNOTEL data for its meteorological inputs (including precipitation and temperature, which are key variables driving SWE prediction). When applying the trained model in the gridding procedure, however, in-situ data are unavailable at the grid-level. As a result, gridded products such as PRISM are substituted for the meteorological inputs. Since SNOTEL and PRISM are at different scales and resolutions, issues around information loss and mismatched feature representation may arise. Another possible driver for discrepancies in MY SWE compared to the true SWE field is in the limited representativeness of the SNOTEL training data. SNOTEL sites are generally located in mid- and high-altitude areas with little to no tree cover [30] and therefore do not represent the full range of topographic features inherent in the true surface. This may introduce bias in the model parameter optimization procedure during MY SWE model training, leading to more accurate predictions at grid locations that match the range of SNOTEL site attributes and less accurate predictions at grid locations outside these represented ranges.

In the following subsections, we carry out several analyses to explore these two potential drivers in detail.

Scale differences in model training and prediction

To assess the potential impact of scale mismatch between input data used for model training and prediction on the representativeness and accuracy of MY SWE, we carry out two analyses. First, we investigate the effect of applying the pre-trained MY SWE model to gridded covariate data of a different spatial resolution. Specifically, we compare the effect of 800 m PRISM meteorological input data compared to 4 km PRISM meteorological data on MY SWE prediction. Second, we examine the impact of a more smoothed surface. We do this by filtering out ASO pixels whose elevation falls outside a narrow band (e.g., 100 meters) around the 4 km cell-wise average elevation prior to gridding, or averaging, the ASO pixels to produce Gridded ASO SWE.

800 m vs 4 km PRISM input data for prediction The first analysis assesses the impact on gridded MY SWE of altering the resolution of gridded meteorological input variables from 4 km to 800 m. Due to data availability limitations in the proprietary 800 m PRISM data, our analysis (including Figures 4.17 and 4.18) only encompasses basin-dates covering 2016 to 2019. This corresponds to 15 out of the 26 basin-dates for which ASO data are available (Table 3.1).

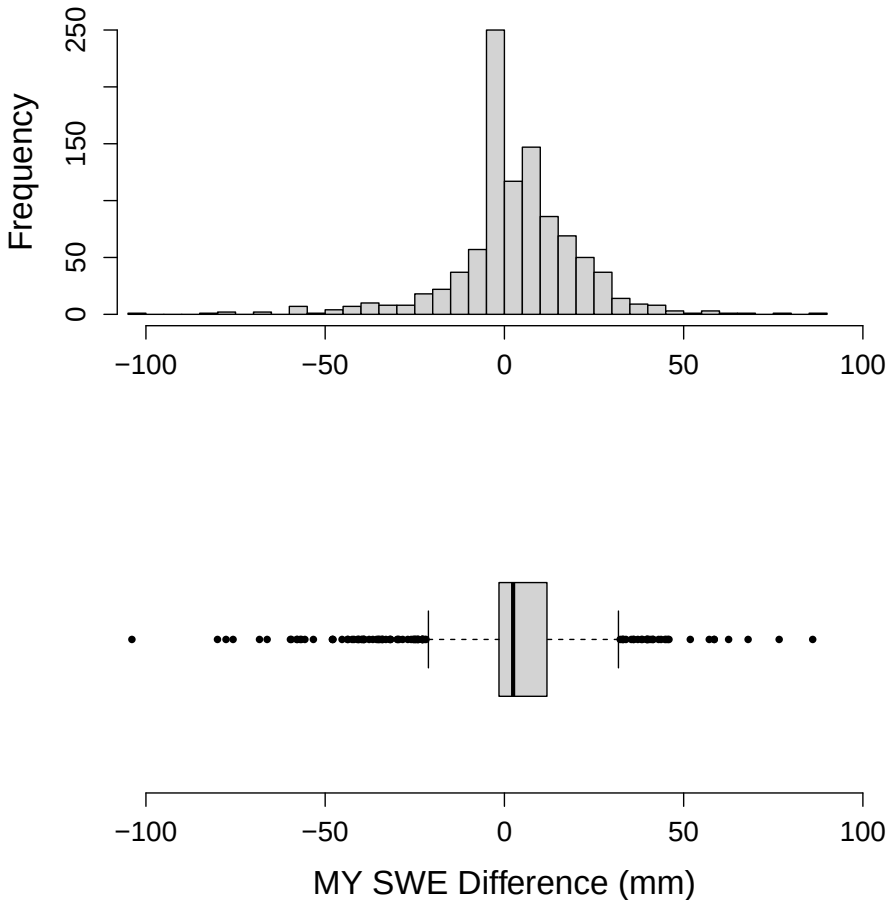


Figure 4.17: Histogram and boxplot of differences between predictions from MY SWE with 4 km meteorological input data and MY SWE with 800 m meteorological input data.

Figure 4.17 shows the distribution of differences between MY SWE gridded using 4 km meteorological input data and MY SWE gridded using 800 m meteorological input data across all 4 km grid cells in the 15 ASO basin-dates. The predicted differences range from approximately -100 mm to 100 mm of SWE, with most values concentrated around 0 mm SWE. Given that SWE in a grid cell located in snowy regions during the snow season

can typically run into the hundreds and low thousands of millimeters, the SWE differences observed in Figure 4.17 are not dramatic.

This observation is reinforced in Figure 4.18, where MY SWE gridded using 4 km meteorological input data against Gridded ASO SWE (middle panel) exhibits little difference in overall pattern compared to MY SWE gridded using 800 m meteorological input against Gridded ASO SWE (bottom panel). In both cases, MY SWE tends to underestimate Gridded ASO SWE, particularly as Gridded ASO SWE increases. Compared to UA SWE (top panel), MY SWE with meteorological input data of both resolutions generally underestimates Gridded ASO SWE to a greater degree.

The similarities in MY SWE predictions given input data of varying resolutions suggests that difference in scale of training and prediction input data may not be a major concern driving discrepancies in MY SWE from true SWE. Nevertheless, previous results show that 4 km PRISM meteorological inputs in place of SNOTEL meteorological inputs decreases SWE estimates considerably. That result, in conjunction with the spatial and temporal constraints of the 800 m meteorological data available for this analysis, warrants further investigation to reach a more definitive conclusion.

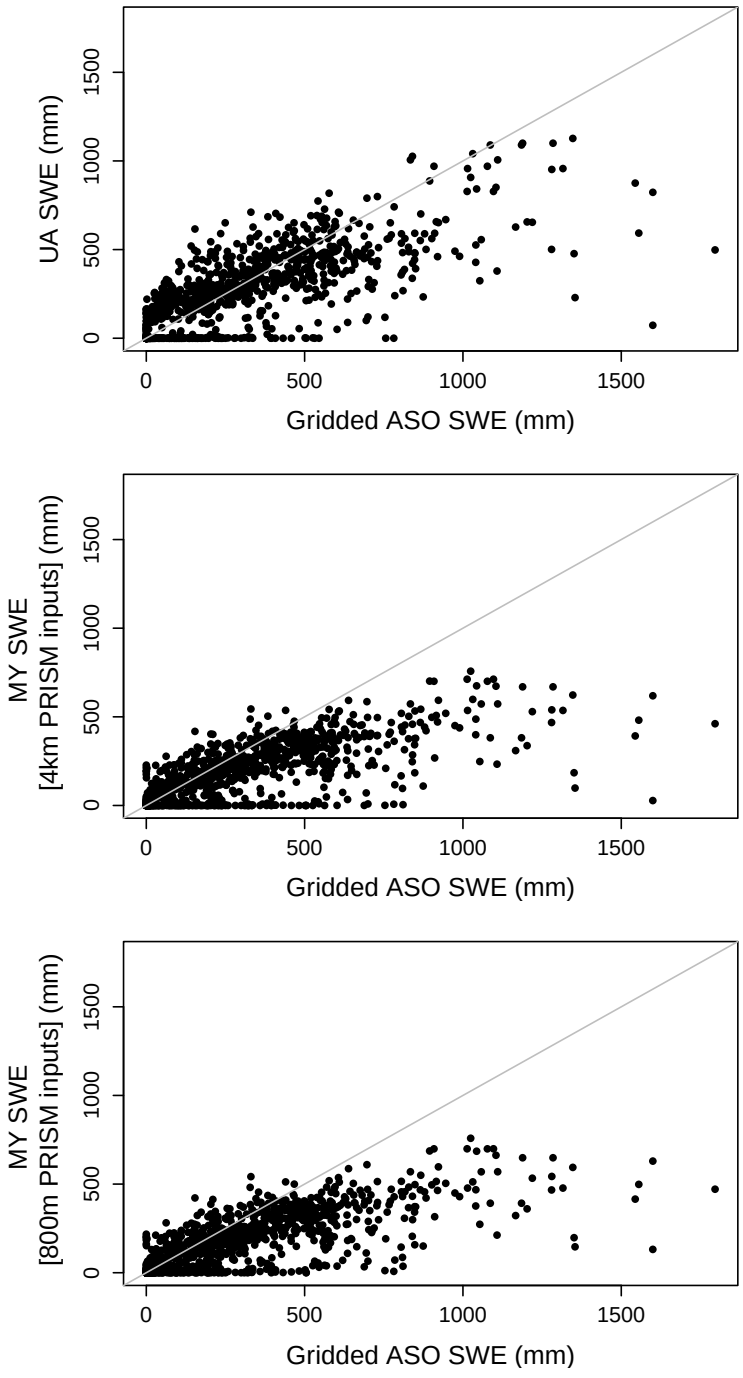


Figure 4.18: Scatterplots comparing Gridded ASO SWE to UA SWE (top row), MY SWE with 4 km PRISM meteorological input data (middle row), and MY SWE with 800 m PRISM meteorological input data (bottom row). Each dot in the scatterplots represents a single grid cell in a basin-date between 2016 and 2019 when proprietary 800 m PRISM data are available.

Filtering by average elevation The second analysis looks at the effect of gridding ASO SWE after filtering out pixels whose elevations differ from their respective cell-wise average elevations. Since MY SWE predictions are generated based on more homogeneous inputs, comparing MY SWE to data from more homogeneous input pixels can provide insight into whether heterogeneity of input data contributes to mismatch of MY SWE to true SWE. Figure 4.19 shows the distribution of differences between Gridded ASO SWE and Gridded ASO SWE generated by averaging 50 m pixels within each 4 km grid cell after filtering out pixels that fall outside of a 100 m elevation band around the cell-wise average elevation (hereafter "Average Elevation Filtered Gridded ASO SWE"). Despite a wide spread, the differences in SWE values are generally small, most of which are concentrated around zero and within 100 mm in both the negative and positive direction.

Contrasting Figure 4.20 against Figure 4.13 provides a more granular view of the Gridded ASO SWE difference in relation to UA SWE and MY SWE (first and second column, respectively), stratified by elevation and month. While there is no strong systematic difference in the pattern of Gridded ASO SWE and Average Elevation Filtered Gridded ASO SWE compared to UA SWE and MY SWE, the R^2 values of UA SWE and MY SWE decrease marginally across all elevation bands, when compared against Average Elevation Filtered Gridded ASO SWE (as opposed to Gridded ASO SWE). This suggests that heterogeneity in ASO does not explain the mismatch between MY SWE and the true SWE surface.

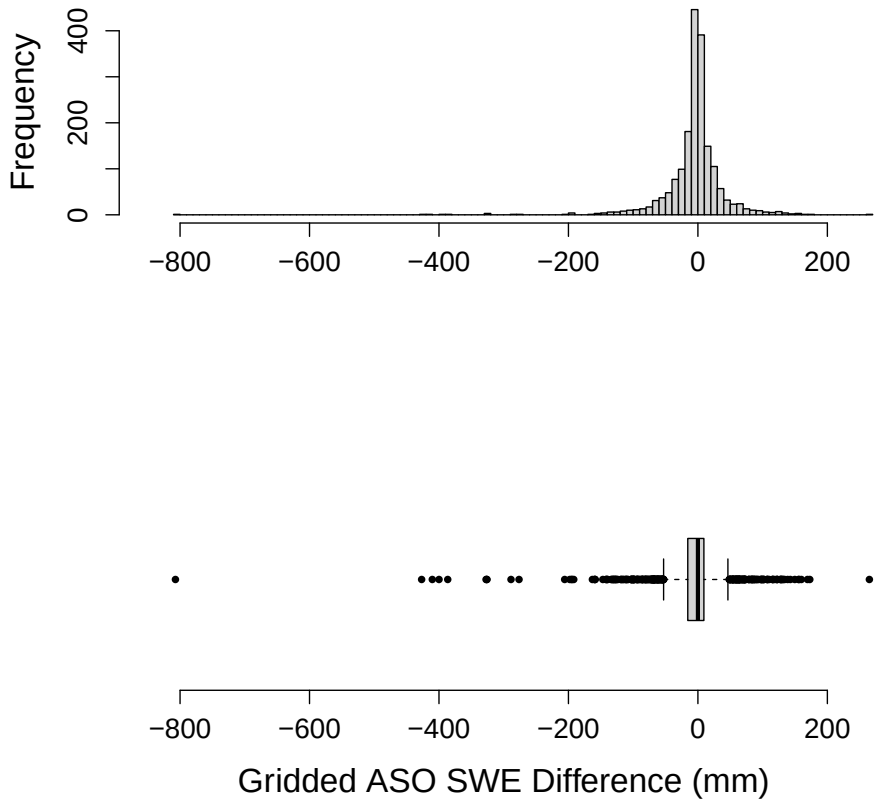


Figure 4.19: Histogram and boxplot of the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside bands around the 4 km cell-wise average elevation (Average Elevation Filtered Gridded ASO SWE).

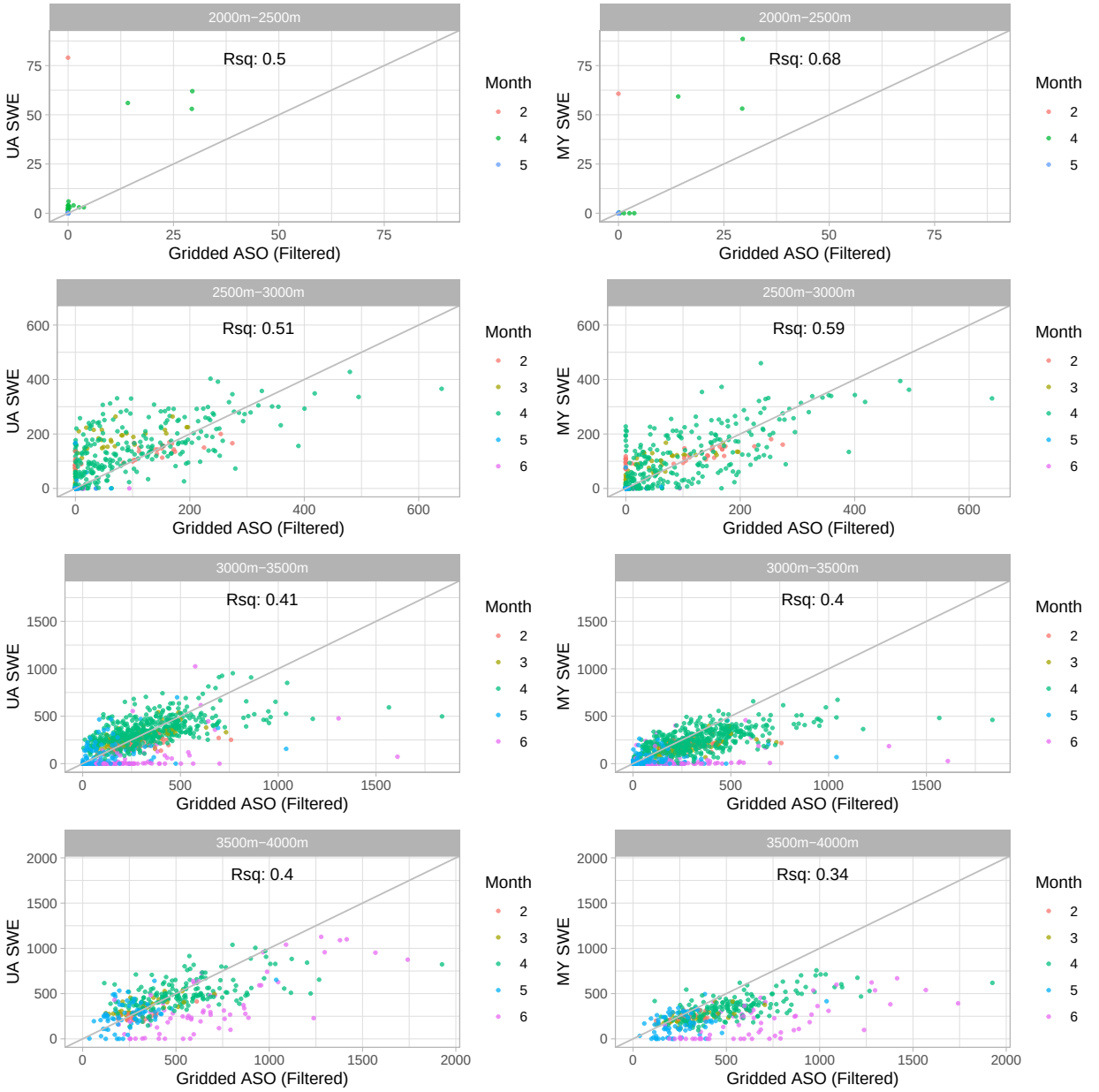


Figure 4.20: Scatterplots showing the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside 4 km cell-wise average elevation bands (Average Elevation Filtered Gridded ASO SWE) compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, stratified by month.

Representativeness of training data: Filtering by static attributes

To assess how limitations in training data representation affects gridded SWE predictions, we evaluate the impact of SNOTEL non-representativeness on MY SWE. This is done by filtering out ASO pixels that do not match topographic attributes (specifically, elevation and slope) of SNOTEL stations in the UCRB prior to the gridding of pixels to produce Gridded ASO SWE.

Figure 4.21 shows the distribution of differences between Gridded ASO SWE and Gridded ASO SWE generated by averaging 50 m pixels within each 4 km grid cell after filtering out pixels that fall outside of the range of SNOTEL elevations and slopes observed at SNOTEL sites across the UCRB (hereafter "Attribute Filtered Gridded ASO SWE"). The spread of SWE differences is wide, but most of the differences are small, concentrated around zero, with a slight positive skew.

Contrasting Figure 4.22 against Figure 4.13 provides a more granular view of the Gridded ASO SWE difference in relation to UA SWE and MY SWE (first and second column, respectively), stratified by elevation and month. Apart from the highest elevation band (3500 m to 4000 m) in May and June, there is no strong systematic difference in Gridded ASO SWE and Attribute Filtered Gridded ASO SWE compared to UA SWE and MY SWE, as exhibited by similar patterns and R^2 values. In May and June, Attribute Filtered Gridded ASO SWE values tend to be lower than Gridded ASO SWE values for grid cells in the highest elevation band. This is ascribed to two main factors. First, SWE can be very high at higher elevations. During the ablation season (May and June), many pixels undergo complete or substantial snowmelt, resulting in lower SWE values. However, pixels located at extremely high elevations retain higher SWE levels. Thus, there is notable variation in SWE values between higher and lower elevations during the ablation period. Second, out of the total ASO pixels across basins and dates, 17.53% exceed the maximum elevation monitored by SNOTEL, 3.14% exhibit slopes steeper than those observed at SNOTEL sites, and 0.59% are located at elevations lower than the minimum recorded by SNOTEL. This means that high elevation is the predominant topographic feature eliminating ASO pixels in the generation of Attribute Filtered Gridded ASO SWE. This finding indicates that higher elevation pixels are less represented by SNOTEL and therefore tend to be excluded during the filtering process. If the modeling procedure does not account for topographic features absent in the training data (SNOTEL in the case of MY SWE), the resulting SWE predictions from the trained model are prone to inaccuracies or misrepresentations at locations where these features are present.

Furthermore, UA SWE and MY SWE align better with Attribute Filtered Gridded ASO SWE than Gridded ASO SWE at high elevations. The stronger agreement between UA SWE/MY SWE and ASO gridded from sites more similar to SNOTEL sites provides some evidence that mismatch between ASO and UA SWE/MY SWE may be related to UA SWE/MY SWE being trained on data from non-representative SNOTEL sites. This indicates that non-representativeness of gridded SWE data could in part be attributed to non-representativeness of training inputs.

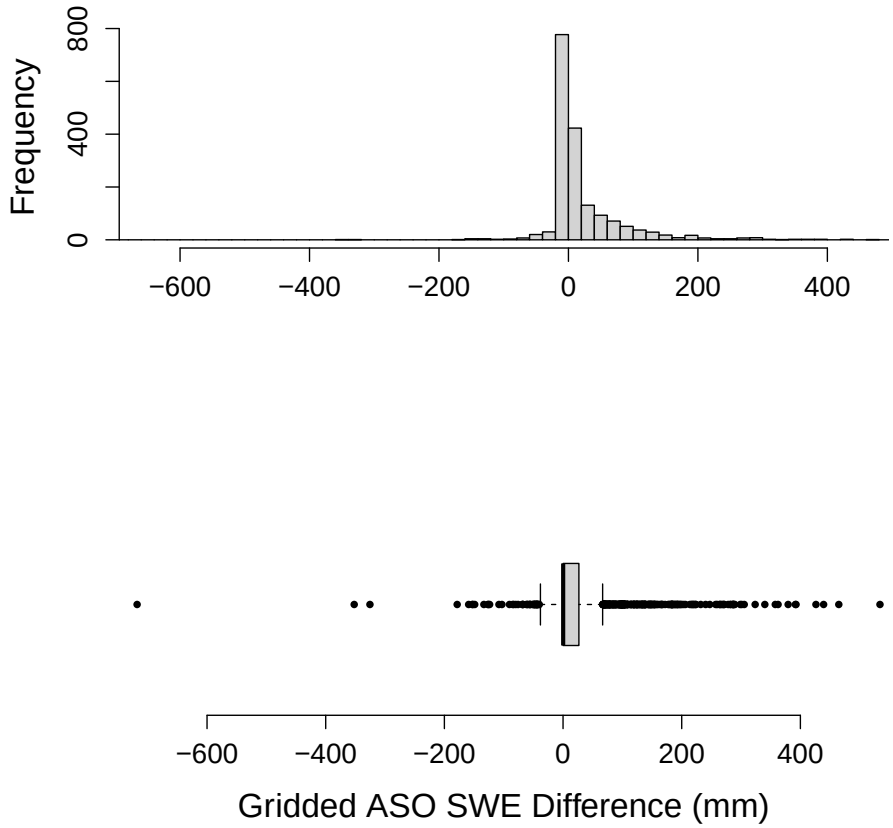


Figure 4.21: Histogram and boxplot of the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside the range of elevation and slope values represented by SNOTEL sites in the UCRB (Attribute Filtered Gridded ASO SWE).

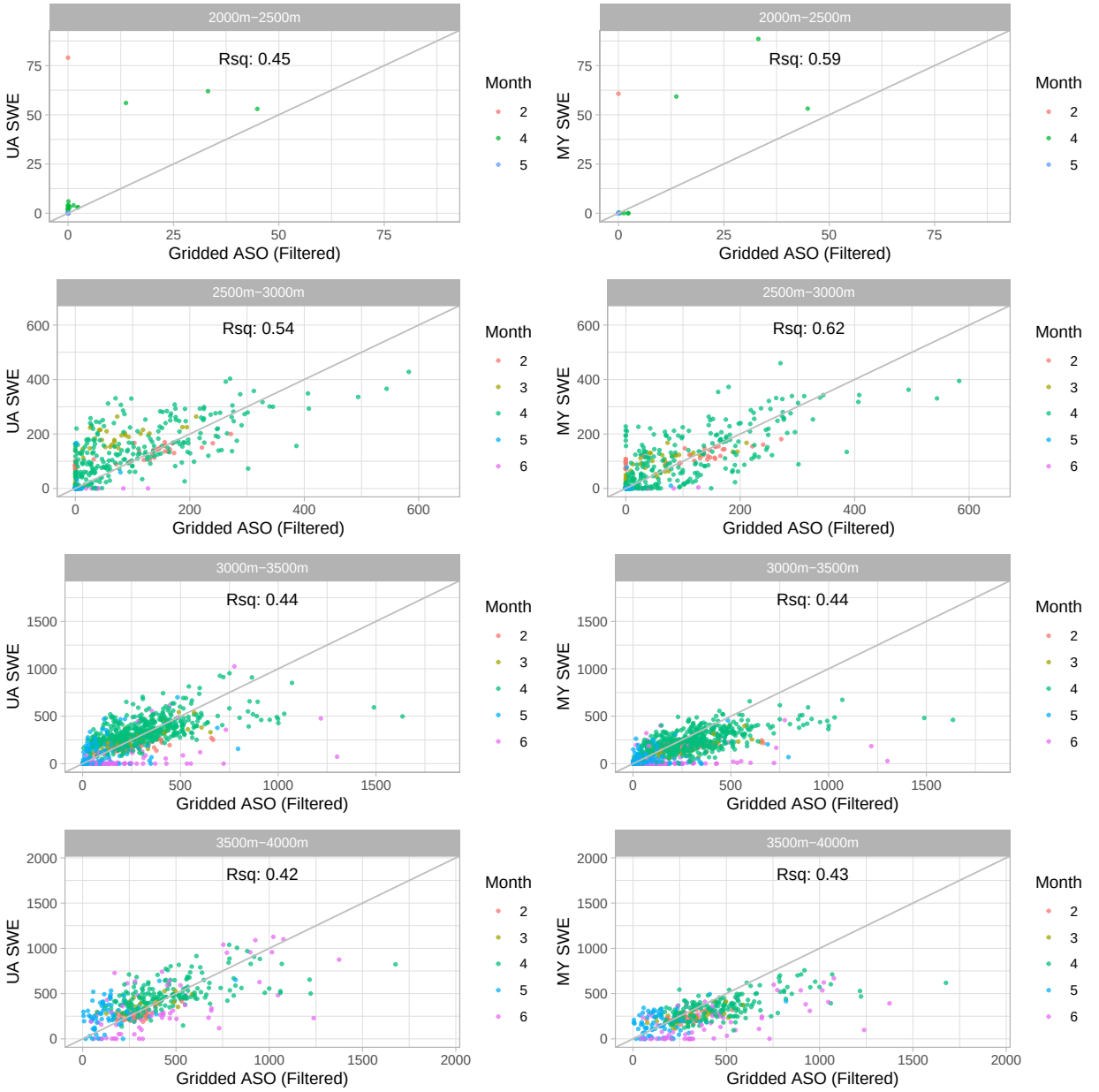


Figure 4.22: Scatterplots showing the difference between Gridded ASO SWE and Gridded ASO SWE derived after filtering out pixels that fall outside the range of elevation and slope values represented by SNOTEL sites in the UCRB (Attribute Filtered Gridded ASO SWE) compared to UA SWE (left column) and MY SWE (right column) at four elevation bands, stratified by month.

Spatial variability

Figure 4.23 shows the variograms of Average Elevation Filtered Gridded ASO SWE and Attribute Filtered Gridded ASO SWE. Variograms of Gridded ASO SWE, UA SWE, and MY SWE from Figure 4.16 are included as reference. Across lag distances, Attribute Filtered Gridded ASO SWE exhibits comparable spatial variability as UA SWE, which is lower than that of Gridded ASO SWE. Since pixels at very high and very low elevations are excluded, the remainder of the pixel SWE values are more homogeneous, leading to lower spatial variability. This suggests that the geographic location of observation sites has some effect on the mismatch between MY SWE and the true SWE surface.

Meanwhile, across lag distances, Average Elevation Filtered Gridded ASO SWE exhibits similar spatial variability (and in fact, slightly greater) compared to Gridded ASO SWE. This similarity might be attributed to the filtering of pixels by cell-wise averages (as in Average Elevation Filtered Gridded ASO SWE), which tends to result in more disparate grid-cell averages, particularly over larger spatial distances. This disparity may be influenced by the substantial heterogeneity across pixels within various areas and on various days, which persists even within individual grid cells (for instance, as seen in examples shown in Figures 4.9 and 4.10). In contrast, when no pixel filtering is applied (as in Gridded ASO SWE), grid-cell averages show less disparity across distances. Overall, the agreement in spatial variability implies that within-cell heterogeneity may not explain the discrepancy between MY SWE and the true SWE surface.

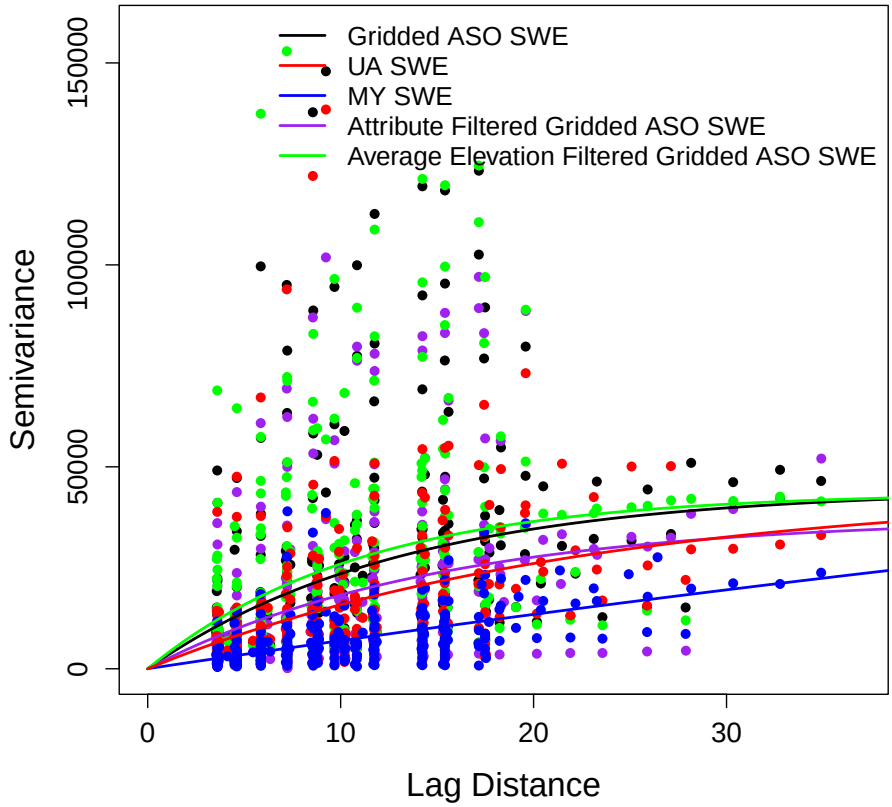


Figure 4.23: Comparison of variograms of Gridded ASO SWE, UA SWE, MY SWE, Average Filtered Gridded ASO SWE, and Attribute Filtered Gridded ASO SWE computed from grid cells aggregated across all basins.

4.6 Conclusion

In this chapter, we evaluated gridded SWE estimates derived from statistical models against in-situ and remote sensing SWE data, in particular leveraging state-of-the-art ASO data that reflects SD and SWE data at very fine spatial resolutions. Following a direct point-level evaluation of UA SWE and MY SWE model predictions against measurements at SNOTEL sites and a discussion on issues related to making point-to-grid comparisons, we conducted a direct scale-matched evaluation of UA SWE and MY SWE against a gridded dataset derived directly from ASO. Finally, an in-depth investigation on potential drivers for inconsistencies between gridded SWE products, such as MY SWE, and the true SWE surface, which is unknown, is presented.

Our analyses suggest that when compared against SNOTEL and ASO, UA SWE and MY SWE exhibit a general trend of underestimation. During the peak SWE period between February and May, UA SWE and MY SWE have a tendency to overestimate SWE at grid cells located in lower elevation regions. Meanwhile, during the ablation season in May and June, UA SWE and MY tend to underestimate SWE in higher elevation areas when ASO indicates persisting snow on the ground. Overall, gridded SWE products, such as MY SWE, are not fully representative of the true SWE surface. We explored several potential factors contributing to this lack of representation, namely input data resolution, small-scale heterogeneity of the input data, and spatial distribution of input training data relative to the full range of true terrain attributes. Due to spatial and more importantly temporal limitations in ASO, in terms of both continuity and coverage, careful consideration is needed before generalizing these findings.

Bibliography

[1] J. T. Abatzoglou. "Development of gridded surface meteorological data for ecological applications and modelling". In: *International Journal of Climatology* 33.1 (2013), pp. 121–131.

[2] R. P. Allan et al. "Advances in understanding large-scale responses of the water cycle to climate change". In: *Annals of the New York Academy of Sciences* 1472.1 (2020), pp. 49–75.

[3] B. T. Anderson et al. "Insights into the physical processes controlling correlations between snow distribution and terrain properties". In: *Water Resources Research* 50.6 (2014), pp. 4545–4563.

[4] A. Behrangi, K. J. Bormann, and T. H. Painter. "Using the Airborne Snow Observatory to Assess Remotely Sensed Snowfall Products in the California Sierra Nevada". In: *Water Resources Research* 54.10 (2018), pp. 7331–7346.

[5] J. Boehner and O. Antonic. "Land-surface parameters specific to topo-climatology". In: *Developments in Soil Science* 33 (2009), pp. 195–226.

[6] G. de Boer and others. "Supporting Advancement in Weather and Water Prediction in the Upper Colorado River Basin: The SPLASH Campaign". In: *Bulletin of the American Meteorological Society* 104.10 (2023), E1853–E1874.

[7] L. Breiman. "Random Forests". In: *Machine Learning* 45 (2001), pp. 5–32.

[8] S. Brönnimann et al. "Historical weather data for climate risk assessment". In: *Annals of the New York Academy of Sciences* 1436.1 (2018), pp. 121–137.

[9] R. D. Brown, B. Brasnett, and D. Robinson. "Gridded North American monthly snow depth and snow water equivalent for GCM evaluation". In: *Atmosphere-Ocean* 41.1 (2003), pp. 1–4.

[10] P. D. Broxton, N. Dawson, and X. Zeng. "Linking snowfall and snow accumulation to generate spatial maps of SWE and snow depth". In: *Earth and Space Science* 3 (2016), pp. 246–256.

[11] P. D. Broxton, X. Zeng, and N. Dawson. "Why Do Global Reanalyses and Land Data Assimilation Products Underestimate Snow Water Equivalent?" In: *Journal of Hydrometeorology* 17.11 (2016), pp. 2743–2761.

[12] T. Carroll et al. “NOAA’S NATIONAL SNOW ANALYSES”. In: *NOAA’S NATIONAL SNOW ANALYSES*. 2006.

[13] E. Cho and J. M. Jacobs. “Extreme Value Snow Water Equivalent and Snowmelt for Infrastructure Design Over the Contiguous United States”. In: *Water Resources Research* 56.10 (2020), pp. 1–20.

[14] N. Cressie. *Statistics for Spatial Data*. John Wiley Sons, 1993.

[15] N. Cressie. “The origins of kriging”. In: *Mathematical Geology* 22.3 (1990), pp. 239–252.

[16] N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley Sons, Inc., 2015.

[17] C. Daly et al. “Challenges in Observation-Based Mapping of Daily Precipitation across the Conterminous United States”. In: *Journal of Atmospheric and Oceanic Technology* 38.11 (2021), pp. 1979–1992.

[18] J. S. Deems, T. H. Painter, and D. C. Finnegan. “Lidar measurement of snow depth: a review”. In: *Journal of Glaciology* 59.215 (2013), pp. 467–479.

[19] J. Dozier, E. H. Bair, and R. E. Davis. “Estimating the spatial distribution of snow water equivalent in the world’s mountains”. In: *WIREs Water* 3 (2016), pp. 461–474.

[20] S. Duan et al. “Using Temporal Deep Learning Models to Estimate Daily Snow Water Equivalent Over the Rocky Mountains”. In: *Water Resources Research* 60.4 (2024), e2023WR035009.

[21] L. A. Ensor and S. M. Robeson. “Challenges in Observation-Based Mapping of Daily Precipitation across the Conterminous United States”. In: *Journal of Applied Meteorology and Climatology* 47.9 (2008), pp. 2468–2476.

[22] J. Fan, F. Han, and H. Liu. “National Science Review”. In: *IEEE* 1.2 (2014), pp. 293–314.

[23] Y. Fang, Y. Liu, and S. A. Margulis. “A western United States snow reanalysis dataset over the Landsat era from water years 1985 to 2021”. In: *Scientific Data* 9.677 (2022), pp. 1–17.

[24] D. R. Feldman et al. “The Surface Atmosphere Integrated Field Laboratory (SAIL) Campaign”. In: *Bulletin of the American Meteorological Society* 104.12 (2023), E2192–E2222.

[25] P. Y. Groisman et al. “Reducing biases in estimates of precipitation over the United States: Phase 3 adjustments”. In: *Journal of Geophysical Research: Atmospheres* 101.D3 (1996), pp. 7185–7195.

[26] A. Heggli et al. “Toward snowpack runoff decision support”. In: *iScience* 25.5 (2022), pp. 1–31.

[27] R. A. Houze Jr. "Orographic effects on precipitating clouds". In: *Reviews of Geophysics* 50.1 (2012), pp. 1–47.

[28] E. H. Isaaks and R. M. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, 1989.

[29] K. S. Jennings et al. "Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere". In: *Nature Communications* 9.1148 (2018), pp. 1–9.

[30] S. K. Kampf et al. "Increasing wildfire impacts on snowpack in the western U.S." In: *PNAS* 119.39 (2022), e2200333119.

[31] D. K. Karger et al. "CHELSA-W5E5: daily 1km meteorological forcing data for climate impact studies". In: *Earth System Science Data* 15.6 (2023), pp. 2445–2464.

[32] A. Karpatne et al. "Machine Learning for the Geosciences: Challenges and Opportunities". In: *IEEE* 31.8 (2018), pp. 1544–1554.

[33] S. W. Kienzle. "A new temperature based method to separate rain and snow". In: *Hydrological Processes* 22.26 (2024), pp. 5067–5085.

[34] N. J. Kinar and J. W. Pomeroy. "Measurement of the physical properties of the snowpack". In: *Review of Geophysics* 53.2 (2015), pp. 481–544.

[35] Z. W. Kundzewicz. "Climate change impacts on the hydrological cycle". In: *Ecohydrology Hydrobiology* 8.2-4 (2008), pp. 195–203.

[36] K. E. Lapo et al. "Impact of errors in the downwelling irradiances on simulations of snow water equivalent, snow surface temperature, and the snow energy balance". In: *Water Resources Research* 51.3 (2015), pp. 1649–1670.

[37] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[38] D. Li et al. "How much runoff originates as snow in the western United States, and how will that change in the future?" In: *Geophysical Research Letters* 44.12 (2017), pp. 6163–6172.

[39] D. Li et al. "The Role of Rain-on-Snow in Flooding Over the Conterminous United States". In: *Water Resources Research* 55.11 (2019), pp. 8492–8513.

[40] B. Livneh et al. "A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions". In: *Journal of Climate* 26.23 (2013), pp. 9384–9392.

[41] B. Livneh et al. "A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013". In: *Scientific Data* 2.150042 (2015), pp. 1–12.

[42] C. Ma, S. R. Fassnacht, and S. K. Kampf. "How Temperature Sensor Change Affects Warming Trends and Modeling: An Evaluation Across the State of Colorado". In: *Water Resources Research* 55.11 (2019), pp. 9748–9764.

[43] Y. Ma et al. "Remote sensing big data computing: Challenges and opportunities". In: *Future Generation Computer Systems* 51 (2015), pp. 47–60.

[44] S. A. Margulis et al. "A landsat-era Sierra Nevada snow reanalysis (1985-2015)". In: *Journal of Hydrometeorology* 17.4 (2016), pp. 1203–1221.

[45] G. Mazzotti et al. "Canopy structure, topography, and weather are equally important drivers of small-scale snow cover dynamics in sub-alpine forests". In: *Hydrology and Earth System Sciences* 27.11 (2023), pp. 2099–2121.

[46] R. R. Mccrary, S. McGinnis, and L. O. Mearns. "Evaluation of snow water equivalent in NARCCAP simulations, including measures of observational uncertainty". In: *Journal of Hydrometeorology* 18.9 (2017), pp. 2425–2452.

[47] L. Meromy et al. "Subgrid variability of snow water equivalent at operational snow stations in the western USA". In: *Hydrological Processes* 27 (2013), pp. 2383–2400.

[48] N. Mizukami and S. Perica. "Spatiotemporal Characteristics of Snowpack Density in the Mountainous Regions of the Western United States". In: *Journal of Hydrometeorology* 9 (2008), pp. 1416–1426.

[49] N. P. Molotch and R. C. Bales. "Scaling snow observations from the point to the grid element: Implications for observation network design". In: *Water Resources Research* 41.11 (2005), pp. 1–16.

[50] C. Mortimer et al. "Gridded North American monthly snow depth and snow water equivalent for GCM evaluation". In: *The Cryosphere* 14 (2020), pp. 1579–1594.

[51] R. Mott et al. "Orographic effects on snow deposition patterns in mountainous terrain". In: *Journal of Geophysical Research: Atmospheres* 119.3 (2014), pp. 1419–1439.

[52] National Water Natural Resource Conservation Service and Climate Center. *Air and Water Database*. https://nwcc-apps.sc.egov.usda.gov/. Accessed 6 March 2024.

[53] P. J. Neiman et al. "The Statistical Relationship between Upslope Flow and Rainfall in California's Coastal Mountains: Observations during CALJET". In: *Monthly Weather Review* 130.6 (2002), pp. 1468–1492.

[54] J. W. Oyler et al. "Artificial amplification of warming trends across the mountains of the western United States". In: *Geophysical Research Letters* 42.11 (2015), pp. 153–161.

[55] T. H. Painter et al. "The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo". In: *Remote Sensing of Environment* 184 (2016), pp. 139–152.

[56] R. Petersky and A. Harpold. "Now you see it, now you don't: a case study of ephemeral snowpacks and soil moisture response in the Great Basin, USA". In: *Hydrology and Earth System Sciences* 22.9 (2018), pp. 4891–4906.

[57] R. Petersky and A. Harpold. "Seasonal and Ephemeral Snowpacks of the Conterminous United States". In: *Hydrology* 8.32 (2021), pp. 1–19.

[58] J. Pomeroy and E. Brun. "Physical properties of snow". In: Snow Ecology: An Interdisciplinary Examination of Snow-Covered Ecosystem, 2001, pp. 45–16.

[59] Oregon State University PRISM Climate Group. *Parameter elevation Regression on Independent Slopes Model*. https://prism.oregonstate.edu. Created 4 Feb 2014. Accessed 24 Mar 2024.

[60] A. M. Rhoades, A. D. Jones, and P. A. Ullrich. "Assessing Mountains as Natural Reservoirs With a Multimetric Framework". In: *Earth's Future* 6.9 (2018), pp. 1221–1241.

[61] O. Schabenberger and C. A. Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC, 2005.

[62] M. C. Serreze et al. "Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data". In: *Water Resources Research* 35.7 (1999), pp. 2145–2160.

[63] Natural Resources Conservation Service. *Snow Depth FAQ - Natural Resources Conservation Service*. https://www.nrcs.usda.gov/wps/portal/wcc/home/quicklinks/states/idaho/snowdepth. Accessed 24 June 2024.

[64] E. R. Siirila-Woodburn et al. "A low-to-no snow future and its impacts on water resources in the western United States". In: *Nature Reviews Earth Environment* 2.11 (2021), pp. 800–819.

[65] N. L. Silverman et al. "Dynamically downscaled winter precipitation over complex terrain of the Central Rockies of Western Montana, USA". In: *Water Resources Research* 49.1 (2013), pp. 458–470.

[66] R. A. Slatyer, K. D. L. Umbers, and P. A. Arnold. "Ecological responses to variation in seasonal snow cover". In: *Conservation Biology* 36.1 (2021), e13727.

[67] M. Staudinger, K. Stahl, and J. Seibert. "A drought index accounting for snow". In: *Water Resources Research* 50.10 (2014), pp. 7861–7872.

[68] S. Stellars et al. "Computational Earth Science: Big Data Transformed Into Insight". In: *EOS* 94.32 (2013), pp. 277–278.

[69] I. T. Stewart. "Changes in snowpack and snowmelt runoff for key mountain regions". In: *Hydrological Processes* 23.1 (2008), pp. 78–94.

[70] M. Sturm, M. A. Goldstein, and C. Parr. "Water and life from snow: A trillion dollar science question". In: *Water Resources Research* 53.5 (2017), pp. 3534–3544.

[71] Z. J. Suriano. "North American rain-on-snow ablation climatology". In: *Climate Research* 87 (2022), :133–145.

[72] K. E. Taylor. "Summarizing multiple aspects of model performance in a single diagram". In: *Journal of Geophysical Research* 106.D7 (2001), pp. 7183–7192.

[73] C. W. Thackeray et al. “Snow and Climate: Feedbacks, Drivers, and Indices of Change”. In: *Climate Change and Snow/Sea Ice* 5 (2019), pp. 322–333.

[74] E. Trujillo and N. P. Molotch. “Snowpack regimes of the Western United States”. In: *Water Resources Research* 50.7 (2014), pp. 5611–5623.

[75] L. Tsang et al. “Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing”. In: *The Cryosphere* 16 (2022), pp. 3531–3573.

[76] A. Umirbekov, R. Essery, and D. Müller. “GEMS v1.0: Generalizable Empirical Model of Snow Accumulation and Melt, based on daily snow mass changes in response to climate and topographic drivers”. In: *Geoscientific Model Development* 17 (2024), pp. 911–929.

[77] G. Urban. “Snow water equivalent in the Western Sudetes and its changes in the light of a changing climate”. In: *Journal of Hydrology: Regional Studies* 54 (2024), pp. 1–26.

[78] K. G. Wheeler et al. “What will it take to stabilize the Colorado River?” In: *Science* 377.6604 (2022), pp. 373–375.

[79] M. L. Wrzesien et al. “Characterizing Biases in Mountain Snow Accumulation From Global Data Sets”. In: *Water Resources Research* 55.11 (2019), pp. 9873–9891.

[80] M. Xiao and D. P. Lettenmaier. “Atmospheric Rivers and Snow Accumulation in the Upper Colorado River Basin”. In: *Geophysics Research Letters* 48.16 (2021), pp. 1–9.

[81] Y. Xu, A. Jones, and A. Rhoades. “A quantitative method to decompose SWE differences between regional climate models and reanalysis datasets”. In: *Scientific Reports* 9.16520 (2019), pp. 1–11.

[82] K. Yang et al. “Intercomparison of snow water equivalent products in the Sierra Nevada California using airborne snow observatory data and ground observations”. In: *Frontiers in Earth Science* 11.1106621 (2023), pp. 1–22.

[83] X. Zeng, P. Broxton, and N. Dawson. “Snowpack Change From 1982 to 2016 Over Conterminous United States”. In: *Geophysical Research Letters* 45.23 (2018), pp. 12, 940–12, 947.

[84] S. Zhong et al. “Machine Learning: New Ideas and Tools in Environmental Science and Engineering”. In: *Environmental Science Technology* 55.19 (2021), pp. 12741–12754.

Appendix A

Supplementary Material for Chapter 2

A.1 Sensitivity Analysis

Mathematical Formulation of Total Sum of Squares (SST) Decomposition

The full mathematical formulation of the sum of squares (SST) decomposition introduced in Equation 2.1 is provided below. This can be easily extended to Equation 2.2, which includes an additional factor, ϕ.

Let

- P be precipitation, with $a = 3$ levels.
- T be temperature, with $b = 3$ levels.
- θ be the rain-snow partitioning threshold, which has $c = 2$ levels.
- L be location, which has d levels, corresponding to the number of sites or grid cells considered.
- W be year, which has e levels, corresponding to the number of water years considered.
- n be the total number of data points, or location-years.

As in Equation 2.1, the total sum of squares, SST, can be decomposed into the sum of squares of the factors, their interactions, and the remaining error:

$$\begin{aligned} SST &= SS_P + SS_T + SS_\theta + SS_L + SS_W \quad + \quad SS_I \quad + \quad SS_{resid} \\ &= \text{main effects} \quad + \quad \text{interactions} \quad + \quad \text{residuals} \end{aligned} \tag{A.1}$$

where

$$SS_I = SS_{P:T} + SS_{T:\theta} + SS_{P:\theta} + SS_{P:T:\theta}$$

.
Further defining:

- $y_{i,j,k,l,m,q}$ as the observed value (SWE) q for factor $P = i$, factor $T = j$, factor $\theta = k$, factor $L = l$, and factor $W = m$
- $\bar{y}_{.....}$ be the overall average across all observations
- $\bar{y}_{i.....}$ be the average when factor P has value i, $\bar{y}_{.j....}$ be the average when factor T has value j, and so on for factors θ, L, and W.
- $\bar{y}_{ij....}$ be the average when factor P has value i and factor T has value j, $\bar{y}_{i.k..}$ be the average when factor P has value i and factor θ has value k, and $\bar{y}_{.jk..}$ be the average when factor T has value j and factor θ has value k.
- $\bar{y}_{ijk.}$ be the average when factor P has value i, factor T has value j, and factor θ has value k.
- $\bar{y}_{ijklm.}$ be the average when factor P has value i, factor T has value j, factor θ has value k, factor L has value l, and W has value m.

The total sum of squares, SST, is defined mathematically as:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{d}\sum_{m=1}^{e}\sum_{q=1}^{n}(y_{ijklmq} - \bar{y}_{.....})^2$$

The sum of squares of each of the main effects is:

$$SS_P = n \cdot b \cdot c \cdot d \cdot e \sum_{i=1}^{a}(\bar{y}_{i.....} - \bar{y}_{.....})^2$$

$$SS_T = n \cdot a \cdot c \cdot d \cdot e \sum_{j=1}^{b}(\bar{y}_{.j....} - \bar{y}_{.....})^2$$

$$SS_\theta = n \cdot a \cdot b \cdot d \cdot e \sum_{k=1}^{c}(\bar{y}_{..k...} - \bar{y}_{.....})^2$$

$$SS_L = n \cdot a \cdot b \cdot c \cdot e \sum_{l=1}^{d}(\bar{y}_{...l..} - \bar{y}_{.....})^2$$

$$SS_W = n \cdot a \cdot b \cdot c \cdot d \sum_{m=1}^{e}(\bar{y}_{....m.} - \bar{y}_{.....})^2$$

The interaction effect, SS_I, in Equation 2.1 and A.1 encompasses the interactions between P, T, and θ (SS_P, SS_T, and SS_θ). The sum of squares of the interactions between the P, T, and θ main effects is:

$$SS_{P:T} = n \cdot c \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij\cdots} - \bar{y}_{i\cdots} - \bar{y}_{\cdot j\cdots} + \bar{y}_{\cdots})^2$$

$$SS_{P:\theta} = n \cdot b \sum_{i=1}^{a} \sum_{k=1}^{c} (\bar{y}_{i\cdot k\cdots} - \bar{y}_{i\cdots} - \bar{y}_{\cdot j\cdots} + \bar{y}_{\cdots})^2$$

$$SS_{T:\theta} = n \cdot a \sum_{j=1}^{b} \sum_{k=1}^{c} (\bar{y}_{\cdot jk\cdots} - \bar{y}_{\cdot j\cdots} - \bar{y}_{\cdot\cdot k\cdots} + \bar{y}_{\cdots})^2$$

$$SS_{P:T:\theta} = n \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (\bar{y}_{ijk\cdots} - \bar{y}_{ij\cdots} - \bar{y}_{i\cdot k\cdots} - \bar{y}_{\cdot jk\cdots} + \bar{y}_{i\cdots} + \bar{y}_{\cdot j\cdots} + \bar{y}_{\cdot\cdot k\cdots} + \bar{y}_{\cdots})^2$$

The residual effect in Equations 2.1 and A.1 encapsulates un-modeled variation, including the interactions of the location factor and the year factor with one another as well as with the other factors (P, T, and/or θ). The sum of squares of the residuals is:

$$SS_{resid} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{d} \sum_{m=1}^{e} \sum_{q=1}^{n} (y_{ijklmq} - \bar{y}_{ijklm\cdot})^2$$

Appendix B

Supplementary Material for Chapter 3

B.1 Conditions for Daily Ablation and Daily SWE Models

Physical Conditions	UA SWE-like dAbl Model	Physical dAbl Model	Full dAbl Model	dSWE Model
Warm-wet days, no SWE_t				✓
Warm-dry days, no SWE_t				
Warm-wet days, with SWE_t	✓	✓	✓	✓
Warm-dry days, with SWE_t	✓	✓	✓	✓
Mid-wet days, no SWE_t			✓	✓
Mid-dry days, no SWE_t				
Mid-wet days, with SWE_t	✓		✓	✓
Mid-dry days, with SWE_t	✓	✓	✓	✓
Cold-wet days, no SWE_t			✓	✓
Cold-dry days, no SWE_t				
Cold-wet days, with SWE_t			✓	✓
Cold-dry days, with SWE_t		✓	✓	✓

Table B.1: Physical conditions of days used for training and prediction for each daily ablation (dAbl) and dSWE model. This is the tabular representation of the methodology discussed in Section 3.3.

Warm days are days on which average temperature exceeds 2°C, mid days are days on which average temperature falls between 0°C and 2°C, and cold days are days on which average temperature falls below 0°C.

Wet days are days that experience non-zero precipitation and dry days are days that experience no precipitation.

"no SWE_t" refers to days where there is no snow on the ground and "with SWE_t" refers to days where there is snow on the ground.

B.2 Gridded SWE

dSWE NN Gridded Field

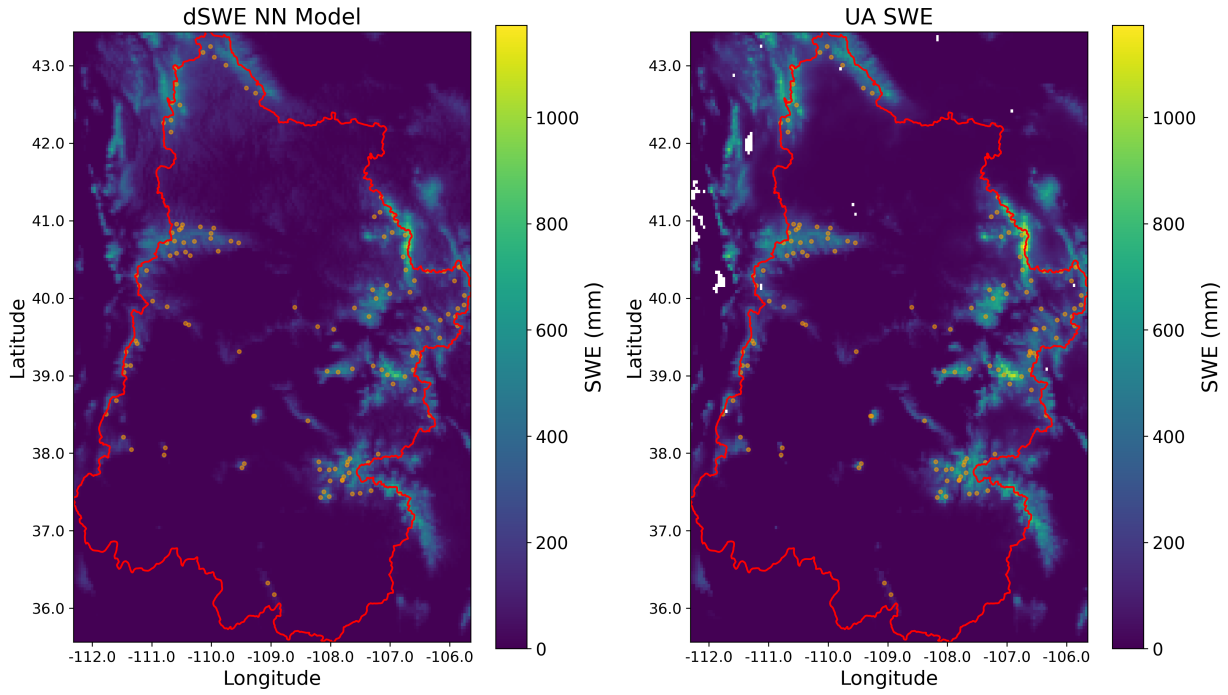
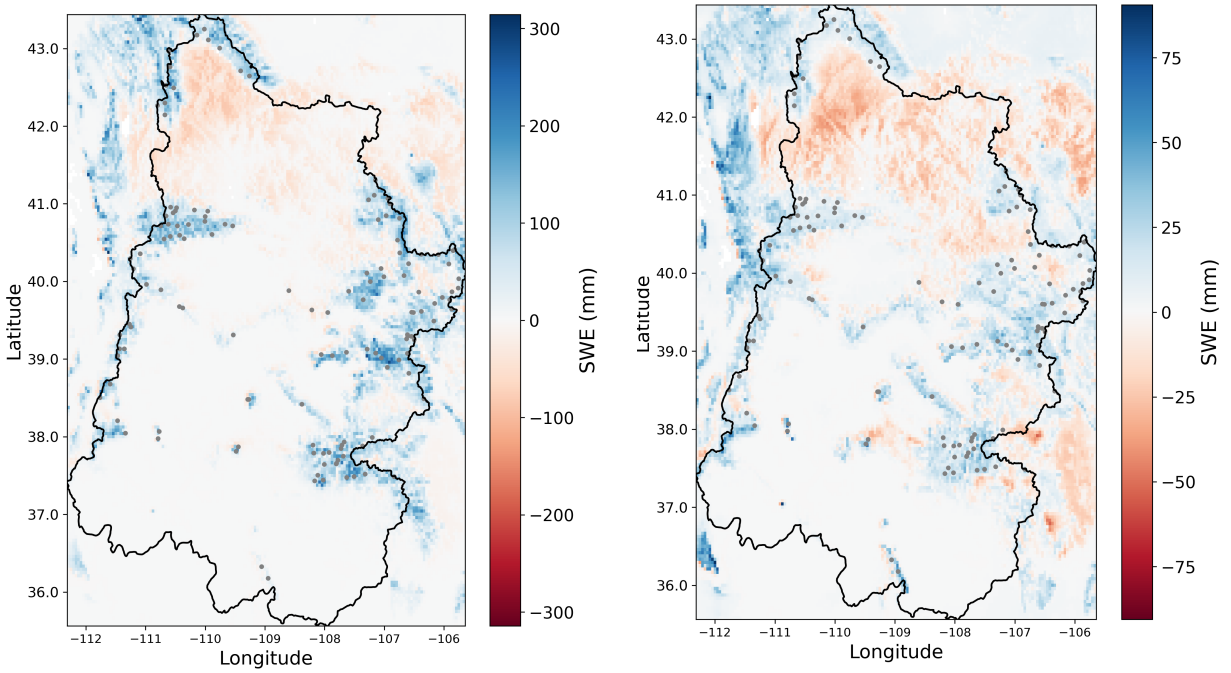


Figure B.1: Gridded SWE prediction from the dSWE NN model compared to UA SWE on April 1.



(a) Differences in mean on April 1.

(b) Differences in standard deviation on April 1.

Figure B.2: Differences in mean and standard deviation of our dSWE NN model versus UA SWE by 4km x 4km grid.

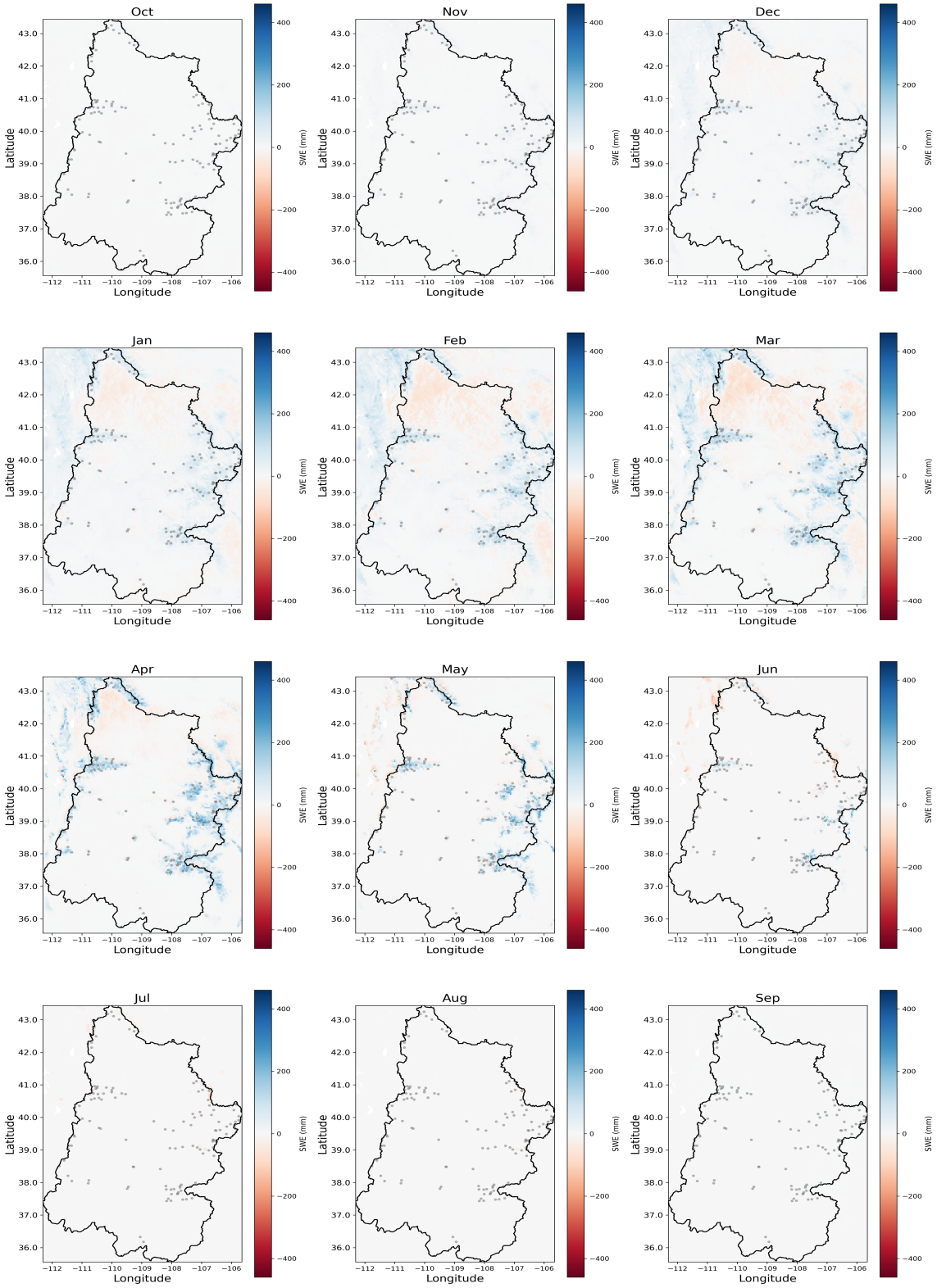
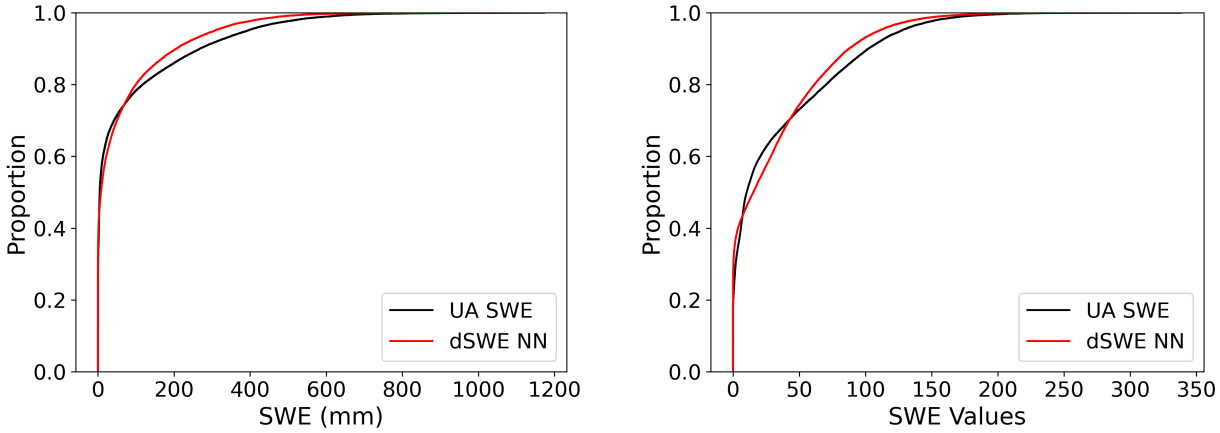


Figure B.3: Difference in mean of our dSWE NN model versus UA SWE by 4km x 4km grid for each month.



(a) ECDF of means.

(b) ECDF of standard deviations.

Figure B.4: Comparison of ECDFs of climatological April 1 means and standard deviations of SWE predicted by UA SWE compared to the dSWE NN model at the grid level over the UCRB.