Understanding the Distribution of Snow Using Remotely Sensed Snow Covered Area

Dominik Schneider

University of Colorado at Boulder, dominik.schneider@colorado.edu
Understanding the Distribution of Snow Using Remotely Sensed
Snow Covered Area

by

Dominik Schneider

B.S., University of Delaware, 2007
M.S., Villanova University, 2011

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Geography
2017
This thesis entitled:
Understanding the Distribution of Snow Using Remotely Sensed Snow Covered Area

written by Dominik Schneider
has been approved for the Department of Geography

____________________________________
Prof. Noah Molotch

____________________________________
Prof. Mark Williams

____________________________________
Prof. Peter Blanken

____________________________________
Prof. Ben Livneh

____________________________________
Prof. Rajagopalan Balaji

____________________________________
Dr. Jeff Deems

Date______________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
Schneider, Dominik (Ph.D., Hydrology)

Understanding the Distribution of Snow Using Remotely Sensed Snow Covered Area

Thesis directed by Prof. Noah P. Molotch

Snowmelt makes up a large portion of the streamflow in the mountainous western United States. The spatial distribution of snow water equivalent (SWE) can affect the magnitude and timing of the spring and summer runoff represented in the hydrograph. Hence, efforts to improve our understanding of the spatial distribution of SWE are vital for good management of our ecological and water resources. SWE is traditionally monitored at measuring stations spread across the western United States, but these stations have been shown to poorly represent the unsampled areas. Remote sensing from satellites has existed since the 1960s but is still unable to measure SWE at scales relevant for water resources. This research utilizes spatio-temporal datasets to promote the use of historical observations of fractional snow covered area (fSCA) to improve estimates of SWE. First, I show that retrospective models of historical SWE distributions from observed fSCA depletion patterns augment existing ground observations of SWE to improve real-time estimates of SWE in unsampled locations. Second, I show that remotely sensed observations of fSCA improve the temporal transferability of the relationship between topography and SWE. Third, a high resolution spatio-temporal dataset is used to observe depletion curves for the first time and evaluate the topographic controls on the relationship between fSCA and snow depth inherent in these depletion curves. Each of these chapters leverages fSCA as an important component and together imply that fSCA has historically been an underutilized observation. Observations of fSCA are available globally for about three decades but necessitate spatially explicit observations of snow depth or SWE to make the most of this long record. Emerging technologies, such as Light Detection and Ranging (LiDAR) and Ground Penetrating Radar (GPR), that provide high resolution spatio-temporal observations of the snowpack and other environmental variables should continue to be exploited to provide insights regarding the physical processes controlling snow dynamics and more generally
our water resources. Future adoptions to climate change rely on improving our understanding of the controlling processes and our ability to monitor them at the relevant scales.
Acknowledgements

There are many people to thank for this accomplishment:

• My parents who believe there is no such thing as too much education.

• Boulder for the opportunity to meet the perfect woman.

• Jess for her love, support, and steadfast example of work hard, play harder.

• Friends and roommates who provided respite from the grueling adventure that is higher education.

• Labmates who spent hours in the field freezing their hands off with me and provided intellectual stimulation with their varied interests.

• My advisor who devoted hours discussing the implications of the thousands of figures I generated. And the opportunity to get involved in a diverse set of projects.

• My committee for their continued interest in my work, encouragement, and stimulating discussions.

• Members of the scientific community in Boulder and abroad. I was very lucky to have interactions with such a diversity of scientifically inclined folks.

• Members of the athletic community in Boulder who continue to push me to be stronger and faster. And encourage a steady stream of endorphins.

• Mother Nature for providing the beauty and serenity of the backcountry. And providing some sweet POW POW once in a while.

• The online community of R and statistical gurus and FOSS developers who inspired me beyond research.
## Contents

### Chapter

1 Introduction

1.1 Motivation ....................................... 1

1.2 Background ...................................... 2

1.3 Chapter 2: Real-time estimation of snow water equivalent in the Upper Colorado River Basin using MODIS-based SWE reconstructions and SNOTEL data .... 7

1.4 Chapter 3: Estimating relationships between snow water equivalent, snow covered area, and topography to extend the Airborne Snow Observatory dataset ........ 8

1.5 Chapter 4: Topographic Controls on Depletion Curves ......................... 8

1.6 References ....................................... 9

2 Real-Time Estimation of Snow Water Equivalent in the Upper Colorado River Basin using MODIS-based SWE Reconstructions and SNOTEL data 14

2.1 Introduction ....................................... 14

2.2 Study Area ....................................... 17

2.3 Methods ......................................... 17

2.3.1 Data Sources ................................... 19

2.3.2 SWE Interpolation Methodologies ..................... 21

2.3.3 Model Evaluation ................................ 25

2.4 Results ......................................... 28
2.4.1 SNOTEL Cross-validation Model Evaluation ................ 28
2.4.2 Distributed Snow Survey Measurement Model Evaluation .......... 29
2.4.3 Regional Differences in Model Performance ................ 31
2.4.4 Spatial SWE Differences ........................... 34
2.5 Discussion ....................................... 36
2.6 Conclusion ...................................... 46
2.7 Acknowledgements ................................. 47
2.8 Appendix A ..................................... 47
2.9 References ....................................... 48
3 Estimating relationships between snow water equivalent, snow covered area, and topography to extend the Airborne Snow Observatory dataset 56
3.1 Introduction ...................................... 56
3.2 Site Description .................................... 60
3.3 Data Sources ..................................... 60
3.4 Methods ........................................ 62
   3.4.1 SWE Models: Discrete Time ......................... 63
   3.4.2 SWE Models: Transferred in Time ...................... 65
   3.4.3 Statistical Model ............................... 66
3.5 Results ......................................... 68
   3.5.1 SWE Models: Discrete Time ......................... 68
   3.5.2 SWE Models: Transferred in Time ...................... 69
   3.5.3 Evaluating the Selected Model Performance ............... 75
3.6 Discussion ........................................ 79
   3.6.1 SWE Distribution Modeling ............................ 79
   3.6.2 Considerations for Extending the ASO Record ............. 84
3.7 Conclusion ........................................ 86
### 4 Topographic Controls on Depletion Curves Observed from Airborne LiDAR Snow Depth Data

#### 4.1 Introduction

96

#### 4.2 Methods

98

- **4.2.1 Site Description**
  
98

- **4.2.2 Data Sources**
  
98

- **4.2.3 Depletion Curves**
  
100

- **4.2.4 Topographic Controls on Depletion Curves**
  
101

#### 4.3 Results and Discussion

104

- **4.3.1 Depletion Curves**
  
104

- **4.3.2 Topographic Controls on Depletion Slopes**
  
108

- **4.3.3 Implications**
  
111

#### 4.4 Conclusions

113

#### 4.5 References

113

### 5 References

119
Tables

Table

2.1 Summary of Snow Surveys. .................................................. 27
2.2 Cross-validated statistics from model estimates compared with SNOTEL observa-
tions. .................................................................................... 29
3.1 List of physiographic variables ............................................. 64
3.2 Mean prediction performance for FSCA, PHV, and PHV-FSCA from each obser-
vation date using a split-sampling approach. .............................. 69
3.3 Mean model performance comparison with PHV-FSCA for the split sample model
and the best transferred model for each simulation date. .............. 75
3.4 Mean model performance comparison with PHV-FSCA for the best transferred
model and selected model for each simulation date. .................... 76
4.1 Variables used in a recursive partitioning regression to explain depletion curve slopes.103
4.2 Single variable Kendall correlations between each of the topographic variables and
the depletion slope. ................................................................. 108
Figures

Figure

2.1 Map of the Upper Colorado River Basin. .......................... 18
2.2 Range of daily squared Pearson Correlation coefficients. .......... 22
2.3 Timeseries of cross-validated skill scores. .......................... 30
2.4 Boxplots of SWE comparing PHV-baseline and PHV-RCN against independent
      snow survey data. .................................................. 32
2.5 Histograms of differences in percent bias. .......................... 33
2.6 Map of percent of years at each SNOTEL station where PHV-RCN exhibited a
      lower percent bias. .................................................. 35
2.7 SWE maps of predictions. ............................................. 37
2.8 Elevation bands of average SWE depth on April 1 from 2001-2012 for each HUC4
      basin by elevation band within the Upper Colorado River Basin. ....... 38
2.9 Elevation distribution of snow accumulating area ...................... 41

3.1 Landcover map of the Tuolumne Basin ............................. 61
3.2 Example SWE distribution ............................................ 70
3.3 The range of $r^2$ for each simulation date from the transferred models. .... 72
3.4 The range of %MAE for each simulation date from the transferred models. .... 73
3.5 The range of %Bias for each simulation date from the transferred models. .... 74
3.6 The increase in prediction error for %MAE and %Bias between the best model and
      the selected model. ................................................. 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>The transferred prediction errors for PHV-FSCA for each date.</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Landcover and location map of Tuolumne basin, California, USA.</td>
<td>99</td>
</tr>
<tr>
<td>4.2</td>
<td>Example depletion curves from the Tuolumne.</td>
<td>106</td>
</tr>
<tr>
<td>4.3</td>
<td>Maps of depletion slope and temporally averaged standard deviations of snow depth</td>
<td>107</td>
</tr>
<tr>
<td>4.4</td>
<td>The regression tree for the controls on the depletion slope.</td>
<td>110</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

Regional water managers and various other agencies rely upon accurate assessments of water availability in order to make critical management decisions, allocating available water among varied stakeholder demands. Estimates of snow water equivalent (SWE) are critical for understanding ecologic impacts and a key component of forecasting water supply, as SWE distribution controls the magnitude of spring and summer runoff in mountainous regions globally. Changes in water and energy partitioning due to climate change have resulted in hydrographs with an earlier rising limb and peak. Current forecasting commonly uses a historical regression approach between SWE measured at index locations and water availability, but these relationships are decreasing in utility as streamflow variability increases. As such, there is a need to improve our capacity to estimate the spatial distribution of SWE to enable more accurate hydrologic models and decrease uncertainty in water allocation and flood mitigation.

Automated hydrologic measuring stations, such as the NRCS SNOw TELemetry (SNOTEL) network, provide high temporal resolution (hourly) data over a large area, but too large of spacing to capture important snowpack dynamics as SNOTEL sites are often not representative of surrounding terrain. Existing satellite remote sensing cannot yet retrieve direct estimates of SWE relevant to water resource management in mountainous snow-dominated regions. In this regard, emerging technologies such as Light Detection and Ranging (LiDAR) and Structure from Motion photogrammetry (SfM) that can measure SWE over large spatial extents at high resolution need to
be leveraged to improve water resource management and our understanding of the physical processes that control our water resources.

These opportunities motivate the development of innovative techniques that can leverage both decades of research and new technologies. This dissertation focuses on developing new insights by combining observations from multiple scales with physically based and statistical modeling. The first chapter merges complementary information from operational snow measuring stations with historical SWE distributions to general real-time maps of SWE distribution. The historical SWE distributions are based on 12 years of remotely sensed fSCA and provide important information regarding the distribution of SWE in unsampled locations. The second chapter utilizes the Airborne Snow Observatory (ASO) dataset, a novel airborne LiDAR mission that provides a high resolution spatio-temporal dataset of SWE. This chapter builds upon previous work in the literature in that statistical relationships between SWE, physiography, and fSCA are transferred to other points in time based on similarities in SWE distribution. Lastly, the spatio-temporal ASO dataset is used to examine the topographic controls on the relationship between snow depth and fSCA in order to improve our understanding of depletion dynamics. These three chapters are unified under the premise that historical records of fSCA have been under-utilized for estimating and understanding the processes that control SWE distribution. Emerging spatio-temporal datasets provide key insights for adding new value to the historical fSCA record. What follows is a brief description of the research objectives for each chapter.

1.2 Background

The distribution of SWE is a result of the interaction between meteorology and topography. Typically about 70% of precipitation in the Colorado Rockies consists of snow which falls in the autumn, winter and spring [Williams et al., 1999]. Snow accumulation patterns result from canopy interception, redistribution by wind and gravity, and orographic influences [Liston, 2004]. Typically, wind redistribution in forested regions is less than in the alpine because wind speeds are reduced. Consequently, SWE accumulation may be lower under canopy but also more homoge-
nous. The redistribution component is strongly affected by the orientation of the terrain relative to
the dominant wind direction. At a broad scale, the dominant wind direction is east due to atmo-
spheric circulation patterns. However, regional orography will dictate a specific wind direction for
each basin that generates a unique SWE pattern. In this regard, it is difficult to generalize a SWE
pattern for all regions.

Intensive field campaigns to capture SWE distributions with manual point measurements
are primarily only performed at the headwater catchment scale due to limited time and human
resources. Interpolation of these points based on their relationship with physiography remains a
common approach for estimating spatially continuous patterns of SWE. Binary regression trees
have been shown to explain up to 65% of variability in SWE when fit with data of a high point
density and modeled at relatively small resolution for small areas, e.g. 30 m for 3 km² [Elder et al.,
1998; Balk and Elder, 2000; Molotch and Bales, 2005]. The advantage of binary regression trees is
that it is able to model complex nonlinear interactions, however they can be less robust compared
to other statistical methods when used for prediction [López-Moreno and Nogués-Bravo, 2006].

In the context of water resources, SWE is largely unsampled, limited mostly to the SNOTEL
point network in the USA with roughly 1 station per 1000 km². Attempts have been made to
interpolate SWE at scales more relevant for water supply forecasting, e.g. at resolutions between
100m and 1000m. Similar to the headwater studies, these studies interpolated point measurements
with global regression techniques using local physiographic variables. Some also included regional
variables to account for differences in storm patterns across basins >1000 km² [Fassnacht et al.,
2003; López-Moreno and Nogués-Bravo, 2006; Harshburger et al., 2010; Jörg-Hess et al., 2014].

Remote sensing provides a means for observing otherwise unsampled locations, however di-
rect retrievals of SWE are not robust in mountainous environments. SWE monitoring methodolo-
gies include the NASA passive microwave AMSR-E SWE products [Tedesco et al., 2004] and air-
borne gamma flights managed by the National Oceanic and Atmospheric Administration (NOAA)
National Operational Hydrologic Remote Sensing Center (NOHRSC) [Carroll et al., 2001]. How-
ever, AMSR-E SWE does not have the spatial resolution required for accurate estimates of SWE in
complex terrain, and suffers from large signal attenuation in forested terrain or when liquid water is present in the snowpack, as is the case in spring when these estimates are most crucial for water managers [Mizukami and Perica, 2012]. Airborne gamma measurements require a low flight altitude and have high errors in complex topography with high variability in SWE or when there is large variability in the underlying soil moisture [Carroll and Carroll, 1989; Carroll et al., 2001]. Active microwave systems designed for measuring SWE do not currently exist operationally and also perform poorly with heterogeneous snowpacks [Davis et al., 2009]. Light Detection and Ranging (LiDaR) techniques (e.g. [Deems et al., 2013]) have enabled accurate remote detection of snow depth, and, combined with modeled density, unprecedented SWE retrievals [Painter et al., 2016]. Currently this technique is limited in extent to spatial scales on the order of $10^3$ km$^2$.

Snow covered area (SCA) can be reliably obtained at various spatial and temporal scales from remote sensing due to the distinctive appearance of snow compared to bare earth and vegetation [Bloeschl et al., 1991; Kirnbauer and Bloeschl, 1994; Rosenthal and Dozier, 1996; König and Sturm, 1998; Painter et al., 2009], but in itself does not provide the information needed to make water management decisions. Nonetheless, several studies have blended interpolated SWE with remotely sensed SCA, primarily by using observations of snow free areas to mask interpolated surfaces. Another innovative use of observed SCA is the SWE reconstruction framework whereby the distribution of peak SWE is estimated by coupling the satellite observed depletion of SCA with a snow melt model [Cline et al., 1998; Molotch, 2009]. This method, however, is limited to retrospective retrievals, forcing uncertainty, and potentially high computational demand depending on the type of snowmelt model used (temperature index vs energy balance).

Complementary to the SWE reconstruction technique, precipitation-driven models partition precipitation into snow and rain, accumulate SWE, and model the energy balance to simulate ablation processes forward in time [Raleigh and Lundquist, 2012]. These models also suffer from forcing uncertainty and potentially high computational demand depending on the scales at which processes are being resolved. An additional uncertainty in these models is the sub-grid distribution of snow. This is important because patchy snow cover can have significant impacts on the en-
ergy balance as sensible heat is advected from bare ground to snow cover [Shook and Gray, 1997; Kuchment and Gelfan, 2001; Flanner and Zender, 2006; Swenson and Lawrence, 2012]. Remote sensing images can be used to constrain the pixel-scale melt [Martinec, 1975], and the assimilation of snow covered area into large scale models have similarly improved streamflow simulations [Andreadis and Lettenmaier, 2006; Clark et al., 2006]. However, more often snow distribution is parameterized with an empirical areal depletion curve [Anderson, 1973] or as a statistical frequency distribution [Liston, 2004], which may not adequately scale with the physiographic gradients inherent within a pixel.

Snow depth and SWE distributions that display elements of repetition each year typically align with persistent SCA patterns. These have been repeatedly confirmed using manual sampling [König and Sturm, 1998; Sturm and Wagner, 2010], imaging [Bloschl et al., 1991; Kirnbauer and Bloschl, 1994], and Light Detection and Ranging (LiDaR) [Deems et al., 2008; Schirmer et al., 2011]. These result from spatially consistent interactions between precipitation and topography, e.g. wind and gravitational redistribution and mid-winter ablation. Snow covered area (SCA) patterns, in particular, have been observed for a long time as an indication of expected streamflow volume [Potts, 1944; Good and Martinec, 1987].

Only a few past studies directly incorporate fSCA to estimate the spatial variability in SWE. Sturm [1998] showed in situ measurements of SWE could be linked to relationships with topography and snow cover patterns from aerial photographs taken during the arctic melt season to estimate SWE distribution in unsampled locations. Sturm et al. [2010] utilized a climatological snow distribution pattern in the Alaskan tundra over nine years and successfully mapped snow depth distribution with only a few index measurements of snow depth. Additionally, they showed that including the climatological snow distribution pattern in a physically-based model further improved results by forcing the simulated distribution to match past observed patterns. In the same vein, Livneh et al. [2014] successfully used a peak SWE climatology that is based on the remotely sensed depletion pattern of fSCA to force the effective winter precipitation (i.e. precipitation after wind redistribution) pattern of a physically-based model in the Rockies of Colorado. Each of these studies supports
the idea that repetition in the patterns of snow cover and SWE provide valuable information with which to inform estimates of SWE distribution in between sparse point measurements.

Patterns of fSCA depletion are largely dictated by interactions between sub-pixel terrain variability, the mean snow depth, and the variability of snow depth. These relationships have been previously explored in the literature. In this context, Marchand and Killingtveit [2005] and Lopez et al. [2014] found significant correlations between sub-pixel terrain variability and the mean and standard deviation of pixel scale snow depth. In addition, [Lehning et al., 2011; Grünewald et al., 2013] showed that sub-pixel terrain variability and surface roughness yield controls on snow depth that exhibit elements of consistency from year to year. New and existing datasets of remotely sensed fSCA present an opportunity to estimate the distribution of SWE and snow depth directly since fSCA also decreases as SWE decreases and the sub-grid terrain variability is increasingly exposed throughout the melt season.

In summary, this dissertation uses a wide variety of data including observations of SWE from ground stations, physically based and statistical modeling, and remotely sensed observations of snow-covered area to improve our understanding of the physiographic controls on SWE and snow depth distribution. The novel contribution of this dissertation is the explicit use of remotely sensed snow covered area. A relationship between fSCA and SWE is well established in that depletion curves are commonly used to estimate areal snow cover based on SWE simulated from a physically based model and the yearly repetition of fSCA patterns has been observed and utilized in past studies. However, the long historical record of remotely sensed images of fSCA have been largely ignored as an opportunity to learn more about the distribution of SWE despite yearly repetition in fSCA patterns that are evidence of a persistent interaction between meteorology and topography. The various datasets mentioned above help to link these repetitive patterns of the areal coverage of snow to the distribution of the height of the snow. What follows are short summaries of the research projects that address this knowledge gap.
1.3 Chapter 2: Real-time estimation of snow water equivalent in the Upper Colorado River Basin using MODIS-based SWE reconstructions and SNOTEL data

Changes in climate necessitate improved snowpack information to better represent anomalous distributions of snow water equivalent (SWE) and improve water resource management. We estimate the spatial distribution of SWE for the Upper Colorado River basin weekly from January to June 2001-2012 in quasi-real-time by two regression techniques: a baseline regression of in situ operationally measured point SWE using only physiographic information and regression of these in situ points combining both physiographic information and historical SWE patterns from a remote sensing-based SWE reconstruction model. We compare the baseline regression approach to our new regression in the context of spatial snow surveys and operational snow measuring stations. When compared to independent distributed snow surveys, the new regression reduces the bias of SWE estimates from -5.5% to 0.8%, and RMSE of the SWE estimates by 8% from 0.25 m to 0.23 m. Notable improvements were observed in alpine terrain with bias declining from -38% to only 3.4%, and RMSE was reduced by 13%, from 0.47 m to 0.41 m. The mean increase in cross-validated $r^2$ for the new regression compared to the baseline regression is from 0.22 to 0.33. The largest increase in $r^2$ in any one year is 0.19, an 83% improvement. The new regression estimates, on average, 31% greater SWE depth than the baseline regression in areas above 3000 m elevation, which contributes up to 66% of annual SWE volume in the driest year. This indicates that the historical SWE patterns from the reconstruction adds information to the interpolation beyond the physiographic conditions represented by the SNOTEL network. Given that previous works using SWE reconstructions were limited to retrospective analyses by necessity, the work presented here represents an important contribution in that it extends SWE reconstructions to real-time applications and illustrates that doing so significantly improves the accuracy of SWE estimates.
1.4 Chapter 3: Estimating relationships between snow water equivalent, snow covered area, and topography to extend the Airborne Snow Observatory dataset

A new spatio-temporal dataset from the ongoing Airborne Snow Observatory (ASO) provides an unprecedented look at the spatial and temporal patterns of snow water equivalent (SWE) over multiple years in the Tuolumne Basin, California, USA. We found that fractional snow covered area (fSCA) significantly improves our ability to model the distribution of SWE based on relationships between SWE, fSCA, and topography. Further, the broad availability of satellite images of fSCA facilitates the transfer of these relationships to different years with minimal degradation in performance ($r^2=0.85$, %MAE=33%, %Bias=1%) compared with models fit on the same day, by considering variations in SWE depth as expressed by differences in fSCA between years. The crux of this proposition is in selecting the model to transfer. We offer a method with which to select a model from another year based on the similarity in SWE distribution at existing snow pillows in the area. Comparison of the best transferred predictions and the selected predictions results in a mild decrease in $r^2$ (0.02) and moderate increases in %MAE (15%) and %Bias (10%). The results motivate further refinement in the technique used to select the best model because if these dates can be identified then SWE can be modeled at accuracy levels equivalent to models generated from ASO data collected on the day of interest. Lastly, we found that models from ASO observations of anomalously low snowpacks in 2015 still transferred to other years, although the same cannot be said for the reverse. This research provides a first attempt at extending the value of ASO beyond the observations and we expect it will continue to provide insights for improving water resource management for years to come.

1.5 Chapter 4: Topographic Controls on Depletion Curves

The repetition of the spatial pattern of snow cover depletion from year to year is well established in the literature. For several decades, theoretical snow cover depletion curves have been used in hydrologic, land surface, and general circulation models to parameterize fractional snow
covered area (fSCA) based on modeled estimates of snow accumulation and snowmelt. Given the difficulty of directly measuring the spatio-temporal snow distribution, characterization of depletion curves and how they vary across mountainous landscapes has not been possible. Here we show, for the first time, that high resolution spatio-temporal snow depth information, from the NASA Airborne Snow Observatory (ASO), can be used to directly characterize depletion curve characteristics and how they vary across physiographic gradients. We utilize a bilinear regression model, relating snow depth and fSCA, to characterize depletion curve characteristics such as the slope of the curves and the breakpoint at which the depletion slope changes. We focus on the slope of the lower portion of the depletion curve and show it is related to the sub-pixel standard deviation of snow depth ($r^2=0.61$). Moreover, we use a regression tree analysis to show that vegetation height, terrain roughness, and northness represent the primary controls on the slope of the depletion curves. This study illustrates that repeat Light Detection and Ranging-based snow measurements can improve process-level understanding of snow distribution and improve the parameterization of snow cover depletion curves in hydrologic and land surface models.

1.6 References


Carroll, T., D. Cline, G. Fall, and A. Nilsson (2001), NOHRSC operations and the simulation of snow cover properties for the coterminal US.


Kimbauer, R., and G. Bloschl (1994), Wie ähnlich sind Ausaperungsmuster von Jahr zu Jahr (How similar are snow cover patterns from year to year)? *Deutsche Wässerkundliche Mitteilungen*, 37(5/6), 113–121.


Liston, G. E. (2004), Representing Subgrid Snow Cover Heterogeneities in Regional and Global


Painter, T. H., K. E. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier (2009), Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, *Remote Sensing*


Chapter 2

Real-Time Estimation of Snow Water Equivalent in the Upper Colorado River Basin using MODIS-based SWE Reconstructions and SNOTEL data

2.1 Introduction

Snowmelt from the Colorado River Basin provides 75% of the water supply for 40 million people in 7 states and 2 countries [Deems et al., 2013b]. Regional water managers rely upon accurate assessments of water availability in order to make critical management decisions, allocating available water among varied stakeholder demands. Estimates of snow water equivalent (SWE) are a key component of forecasting water supply, as SWE distribution largely controls the magnitude of spring and summer runoff.

Operational SWE measurements from the U.S. Natural Resource Conservation Service’s (NRCS) SNOw TELemetry (SNOTEL) network are available in real time and are considered to be accurate at the point scale, with measurement errors of < 1 mm [Beaumont, 1965]. Current water supply forecasting techniques commonly use a regression between historical runoff and SWE measured at SNOTEL stations and snow courses, but recent studies have noted increased variability in streamflow, thereby limiting the utility of historical regression forecasts based on index locations [Pagano and Garen, 2005]. Regression forecasts based on fully distributed spatial estimates of SWE have more skill than those from point locations [Rosenberg et al., 2011]. This motivates the development of improved estimates of spatial SWE distribution since spatially explicit measurements of SWE in complex terrain are currently not available.

Passive microwave remote sensing techniques lack the spatial resolution to provide useful
estimates of SWE in mountainous regions and are subject to issues of signal saturation when liquid water is present and when SWE magnitudes are greater than ~150 mm. Additionally, errors are particularly high in forested terrain because the canopy attenuates the signal [Mizukami and Perica, 2012]. Active microwave systems designed for measuring SWE do not currently exist operationally and also perform poorly with heterogeneous snowpacks [Davis et al., 2009]. Similarly, operational gamma flights from the National Oceanic and Atmospheric Administration (NOAA), National Operational Hydrologic Remote Sensing Center (NOHRSC) have high errors when estimating SWE in complex topography where there is large variability in the underlying soil moisture and SWE [Carroll and Carroll, 1989; Carroll et al., 2001]. Light Detection and Ranging (LiDaR) techniques (e.g. [Deems et al., 2013a]) have enabled accurate remote detection of snow depth and SWE, but such techniques are currently limited to relatively small spatial scales (i.e. ~ 10^3 km^2).

As it stands, researchers and practitioners rely on existing station networks such as the NRCS SNOTEL network to determine the magnitude of SWE on the ground in areas too large to be feasibly sampled manually. Several works have interpolated these data on the basis of topography in order to more directly estimate SWE volume, but predictions are inaccurate because relationships between physiography and SWE are poor at large scales [Carroll et al., 1999; Fassnacht et al., 2003; Harshburger et al., 2010; Jörg-Hess et al., 2014]. Fassnacht et al. [2003] examined a variety of models including inverse distance weighting, local hypsometric regression and multivariate linear regression for interpolating SWE within the Upper Colorado River Basin based on physiographic variables (e.g. elevation, slope, aspect, distance to ocean, etc.) that are also the basis for the Parameter-elevation Relationships on Independent Slopes Model (PRISM) used for estimating the distribution of precipitation [Daly et al., 1994; 2008]. Several other studies have interpolated point measurements in different locations with different techniques to estimate spatial SWE patterns with varying skill [Harshburger et al., 2010; Rice et al., 2011; Jörg-Hess et al., 2014]. Each of these studies suffers from a lack of representative observations beyond the extent of the pillow network and are forced to make gross assumptions about the relationship between physiography and SWE.
Physically based modeling also provides an opportunity to estimate the spatial distribution of SWE. These include precipitation-driven models, e.g. [Winstral and Marks, 2002; Clark et al., 2011; Kumar et al., 2013] and SWE reconstructions, e.g. [Cline et al., 1998; Molotch, 2009]. The latter type reconstructs the snowpack based on the energy balance (i.e. snowmelt) and remote sensing observations of snow covered area depletion [Cline et al., 1998; Liston, 1999; Molotch, 2009]. This method is unavailable in real-time because it is retrospective and similarly suffers from potentially large uncertainties in forcing data from which the potential melt flux is calculated [Slater et al., 2013]. That said, the approach has value in that it does not rely on precipitation estimates which represents a large source of uncertainty in forward models [Raleigh and Lundquist, 2012].

Uncertainty in SWE reconstructions is particularly affected by missing satellite images of fractional snow covered area (fSCA) [Girotto et al., 2014]. Notwithstanding, Slater et al. [2013] showed that missing satellite images of fSCA greater than five days in a row occur less than 5-10% of the time in the western U.S., but that uncertainties in meteorological data can also produce equivalent errors in the energy balance. In their respective studies, both Raleigh et al. [2012] and Girotto et al. [2014] recommended a combined approach whereby the best estimates of peak SWE were obtained by averaging the results from a forward and a backward model [Raleigh and Lundquist, 2012] and including precipitation estimates in a snowmelt model using a data assimilation scheme [Girotto et al., 2014].

Given that SWE patterns (i.e. SWE distribution) often exhibit some level of repetition from year-to-year, SWE reconstructions from previous years may be useful as an explanatory variable in a statistical regression to explain the spatial distribution of SWE in real time. In this regard, we offer an improvement on the multivariate linear regression of Fassnacht et al. [2003] by including past patterns of SWE from a reconstruction model as an additional explanatory variable, i.e. in addition to terrain variables. Past patterns of SWE provide complementary information to point-based interpolations because the spatially continuous reconstruction-based estimates are constrained by observations of the two-dimensional satellite-observed snow cover. In this regard, Guan et al. [2013] showed that spatial distributions of SWE can be more accurately estimated retrospectively.
by blending station observations of SWE with retrospective SWE reconstructions. Our objective is to evaluate whether past spatial patterns of SWE can be used to improve real-time estimates of SWE distribution from a statistical model. This study addresses two research questions: 1) Do MODIS-based SWE reconstructions improve regression-based real-time SWE estimates when combined with ground-based observations? And, 2) How does the spatial distribution of SWE differ when MODIS-based SWE reconstructions are incorporated into regression-based SWE estimates?

### 2.2 Study Area

The modeling domain is the Upper Colorado River Basin (UCRB) (bounding box 33°S, 43.75°N, 112.25°W, 104.125°W) and includes the southern Rocky Mountains from northern Arizona and New Mexico to Wyoming and the Front Range of Colorado to the Wasatch Mountains in Utah (Figure 2.1). The UCRB is a snow-dominated watershed from which snowmelt provides much of the streamflow to the Lower Colorado River Basin. It has a drainage area of 277,000 km², a mean elevation of 2150 m and an elevation range of 975-4260 m with about 60% of the basin above 2000 m [Fassnacht, 2004]. Approximately 65% of land is covered by grassland and shrub, and 25% is forested, with an average forest canopy density of 16%. Tree line is at about 3300 m, above which the land consists of alpine meadows or barren rock and ice, which make up 3% of the land area. Less than 2% of land is developed and the remaining 5% is open water or used for agriculture. The basin is subject to numerous inter-basin transfers, which divert water to population centers across the region.

### 2.3 Methods

To map the distribution of SWE across the UCRB, we fit two regression models approximately weekly from January to June for the years 2001-2012. The first regression model represents the baseline case in which the dependent variable is SNOTEL-observed SWE and the independent variables are the physiographic variables listed in Fassnacht et al. [2003] (described in Section...
Figure 2.1: The Upper Colorado River Basin and its Hydrologic Unit Code (HUC) 4 basins are outlined in black. SNOTEL stations (blue circles) and snow surveys (orange squares) are marked.
2.3.2.1). The second regression model is identical to the baseline model except past patterns of SWE from a MODIS-based SWE reconstruction are included as an independent variable in addition to the independent variables used in the baseline model (described in Section 2.3.2.2). For both models, we show cross-validation results at SNOTEL stations and a performance metric relative to 29 distributed snow surveys. Additionally, we discuss the results as they compare to the estimates of the operational National Weather Service Snow Data Assimilation System (SNODAS) SWE product.

2.3.1 Data Sources

2.3.1.1 Training Data - Observed Snow Water Equivalent

SWE observations used in this study are from 237 SNOTEL stations inside the aforementioned bounding box. Daily SWE data from these stations were acquired from October 1, 2000 to September 30, 2012 (available: http://www.wcc.nrcs.usda.gov/snow/). SNOTEL is an operational network managed by the NRCS primarily designed to measure SWE at point locations as indices for forecasting streamflow volume. The mean elevation of the stations is 2800 m and approximately 65% of stations are within one standard deviation of the mean; i.e. between 2440 m and 3160 m. The lowest station is at 1777 m and the highest station is at 3536 m. Stations are typically located in small clearings in evergreen forests that are relatively flat. The SWE record at each SNOTEL station was controlled for negative and other spurious values.

2.3.1.2 Predictor Data – Terrain Information and Reconstructed SWE

Sixteen independent physiographic variables were derived for the domain from a 500-m Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) DEM available: http://topotools.cr.usgs.gov/GMTED Viewer/. These variables are the same as described in Fassnacht et al. [2003] and include elevation, latitude, longitude, slope, eastness, northness, distance to ocean in three directions (i.e. NW, W, SW), barrier height in three directions, west footprint slope,
regional slope, regional eastness, and regional northness.

A 500 m MODIS-based SWE reconstruction product was derived for the Upper Colorado River Basin using the code and parameterizations described in Guan et al. [2013]. The SWE reconstruction approach is effectively a method of calorimetry whereby model-based estimates of snowmelt flux are integrated over the period of satellite observed snow cover to recover the initial mass of the snowpack at maximum accumulation [Cline et al., 1998]. Snow covered area is derived from MODIS at 500 m spatial resolution using a spectral unmixing algorithm [Painter et al., 2009]. Forcing data for the energy balance calculations are derived from the second generation of NASA’s North American Land Data Assimilation System (NLDAS-2). NLDAS variables used here have an hourly time step and 1/8th degree spatial resolution and include air temperature, specific humidity, surface air pressure, downward longwave and shortwave radiation and wind speed [Xia et al., 2012]. These variables are downscaled to 500-m spatial resolution using a combination of geometric terrain reflectance models [Dozier and Frew, 1990] and elevation lapse rates (see Guan et al. [2013]). Turbulent fluxes are calculated explicitly using the bulk aerodynamic roughness method [Tarboton and Luce, 1996]. All radiative and turbulent fluxes are adjusted for forest cover based on a yearly varying canopy product obtained from the University of Maryland Global Forest Cover Change product [Hansen et al., 2013]. Albedo is parameterized using a time decay curve [U.S. Army Corps of Eng., 1956]. More reconstruction algorithm specifics can be found in Guan et al. [2013].

From hourly energy exchange calculations, the sum of the daily potential melt flux is determined under the presumption of 100% fractional snow coverage in each pixel. This daily potential melt flux is then multiplied by the MODIS-observed fractional SCA for each pixel to determine the actual pixel-specific melt flux. When cloud cover or poor viewing geometry prevent MODIS SCA measurements, daily SCA estimates are obtained via a temporal interpolation scaled by the magnitude of the daily melt flux [Molotch, 2009]. For each pixel, depletion curves of seasonal cumulative actual melt flux versus MODIS-based fractional SCA are derived with the integral under each curve representing the peak SWE for a given pixel. For the purposes of this study, daily
estimates of SWE at 500-m resolution are produced for the snowmelt season for the years 2000 – 2012.

The impact of the errors in the reconstruction product are minimal in this application since we scale the daily estimates to have a mean of zero and standard deviation of one such that only the relative magnitudes are used in this study; see section 2.5 for further explanation. Historical reconstructed SWE explains between .02 and .62 of the variance of real-time SNOTEL SWE with a median of 0.36; this is more than twice the explained variance of any one of the physiographic variables and therefore we expect reconstructed SWE to provide significant information for the regression (Figure 2.2). It is important to note that single variable regressions contain, generally, little statistical power with respect to describing the spatial distribution of SWE.

The time period of this modeling study is currently limited by the availability of MODIS snow covered area images, which started on Feb 24, 2000, such that interpolated SWE could be masked for snow/no-snow areas. Snow covered area is derived from MODIS using the MODIS Snow-Covered Area and Grain size (MODSCAG) algorithm at 500 m spatial resolution [Painter et al., 2009] (available: snow.jpl.nasa.gov). The MODSCAG algorithm is a linear spectral unmixing method that uses Bands 1-7 of MODIS. These bands cover between 620 nm and 2155 nm and are used to distinguish between snow cover, vegetation, rock and clouds. The MODSCAG tiles were downloaded, mosaicked and cropped to the study domain. The tiles include a cloud mask based on empirical band thresholds developed by the JPL Snow Team. The images were manually screened for erroneous cloud masks. The average number of images retained per year was 65 with a range from 35 to 112 out of a possible 242 per year.

2.3.2 SWE Interpolation Methodologies

2.3.2.1 Baseline Physiographic Regression Framework

A generalized linear model (GLM) from the Gaussian family with an identity link function was used to model the effect of 16 independent physiographic variables on the dependent variable,
Figure 2.2: The range of daily squared Pearson Correlation coefficients for each independent variable used in the regression. Only significant correlations are shown (p<0.05). Variables not shown did not have any significant correlations (Regional Eastness, Footprint W). Each box represents the interquartile range with vertical bars for the 5th and 95th percentiles. The horizontal line is the median value.
SWE, as measured from 237 SNOTEL stations approximately weekly from January 1 – June 30, water years 2001-2012. Each day in the record is simulated with the GLM individually to optimize the parameter coefficients, resulting in at least 16 models per water year and 756 models in total for the 12-year study period. The GLM with a Gaussian distribution and identity link function is identical to using ordinary least squares but provides flexibility in the assumed distribution of the dependent variable. A log link function, log transformation of the dependent variable and a gamma distribution with an inverse link function were also tested, but the Gaussian distribution with the identity link function was chosen because the models converged consistently and the residuals were more homoscedastic.

We scaled the dependent variable, observed SWE from the SNOTEL stations, by remotely-sensed fractional snow covered area (adjusted for canopy density, [Molotch and Margulis, 2008]) to provide a better representation of the areal mean SWE for the grid cells encompassing SNOTEL sites. Past studies, e.g. Fassnacht et al. [2003], used the measured SNOTEL SWE directly in a regression, thereby assuming the observed point measurement of SWE is representative of the average SWE in the containing pixel. Making use of both the point observation and the remotely sensed observation limits this assumption such that the observed point measurement is only assumed representative of the area of the pixel that is observed to have snow cover. We assume that scaling SNOTEL SWE by the fractional snow covered area yields more representative average grid-cell SWE values than using the SNOTEL values independently. This assumption is further discussed in Section 5.

The independent variables used are those rated of moderate or high importance in Fassnacht et al. [2003] and are based on the variables used in Parameter-elevation Relationships on Independent Slopes Model (PRISM) [Daly et al., 1994]. Independent variables are scaled to a mean of 0 and standard deviation of 1. A full model with all possible physiographic variables is initially fit, but then is reduced using Bayesian Information Criteria as the objective function to only contain those variables that are statistically significant to avoid overfitting. This reduced model is henceforth referred to as the PHysiographic Variable regression (PHV-baseline) and represents our
baseline model; it is in effect the same as the multivariate regression from Fassnacht et al. [2003] except that we scaled SNOTEL SWE by the remotely sensed fractional snow covered area. The model residuals are examined for normality and variables for multicollinearity. Multicollinearity is when two or more predictor variables are strongly related to each other and is an issue in multiple linear regression because it can lead to unstable parameter estimates and inaccurate estimates of the extra sum of squares for the predictor variable. The Variable Inflation Factor was used to quantify multicollinearity with values greater than ten being considered problematic [Kutner et al., 2005]. Longitude was removed from the modeling framework due to strong multicollinearity with west distance to ocean; west distance to ocean was kept because it is representative of a physical quantity (proximity to moisture). Elevation and southwest barrier difference were also found to exhibit multicollinearity, but elevation was kept because of its known relationship with precipitation and SWE.

Simulations are limited before March 1 to those days for which a mostly cloud free MODSCAG fSCA image is available because fSCA images are used to mask snow-free areas in the simulation results. Days after March 1 utilize an interpolated fSCA product from the reconstruction model for masking and are therefore regularly simulated weekly on the 1st, 8th, 15th and 22nd of the months March, April, May, and June. Days after March 1 for which a mostly cloud-free MODSCAG image is available in between these dates are also simulated.

2.3.2.2 Physiographic Regression with Reconstruction

MODIS-based reconstructed SWE values from a previous year, corresponding to the pixels containing SNOTEL stations, are added to the PHV-baseline model as an additional independent variable to create a physiographic regression with reconstruction (hereafter referred to as PHV-RCN). Each day of simulation is modeled independently to ensure the most representative past spatial pattern is used to explain the real-time distribution of SNOTEL-observed SWE. In this regard, the SWE reconstruction is scaled to mean 0 and standard deviation 1, and is considered to represent the relative spatial distribution of SWE between pixels as observed from a climatologi-
cally similar year. The reconstructed SWE for a single date is chosen from a historical ensemble of reconstructed SWE consisting of the 1\textsuperscript{st} and 15\textsuperscript{th} of the months March to June, 2000-2012; the ensemble is limited from the daily time series for computational efficiency. Since real-time estimates of SWE are desired, only reconstruction dates from years prior to the simulation year are considered; the current and future years’ reconstruction would not be available yet as the method requires snow disappearance before retrospective SWE estimates can be obtained. The date of reconstructed SWE used in the interpolation was determined by optimizing a leave-one-out cross validation procedure that maximizes $r^2$ values at the SNOTEL stations. In this context, the day of year of reconstructed SWE was not limited to the same day of year as the simulation because each year can have considerable differences in the time series of SWE. The resulting model for PHV-RCN was tested using an F-test to confirm it was not statistically similar with PHV-baseline. SWE for the entire 12-year period was simulated weekly in real-time mode using the ensemble of previous years’ SWE reconstructions as an additional independent variable in the regression. This PHV-RCN approach represents the focus of our analyses given our goal of illustrating the utility of previous years SWE reconstruction in real-time applications.

2.3.3 Model Evaluation

2.3.3.1 In situ Model Evaluation

Validation of the PHV-baseline and PHV-RCN models is two-fold. First, a cross-validation scheme is used whereby each SNOTEL station is removed iteratively and the model prediction is compared to the measured value at the removed SNOTEL station. Relative root mean squared error (rRMSE) and $r^2$ are reported. The rRMSE is the RMSE divided by the spatial mean of MODSCAG-scaled SNOTEL-measured SWE. This facilitates comparison of model performance across days and years with different magnitudes of SWE. R-squared is defined as the square of the Pearson Correlation Coefficient, which is appropriate because we are using a linear model.

Second, an independent validation data set based on manually measured points has been col-
lated that includes SWE estimates from snow surveys throughout the domain (Table 2.1). Snow surveys with existing 30 m surface models (e.g. kriging model [Erickson et al., 2005] and binary regression trees [Molotch and Bales, 2005; Meromy et al., 2012]) were aggregated by averaging to the regression pixel size (500 m) and compared at the pixel scale. PHV-baseline and PHV-RCN exist on a grid with a different origin than the validation data. Therefore, comparisons were limited to regression pixels with at least 50% overlap with the observational dataset to avoid comparisons with small unrepresentative areas, but leaving some room for uncertainty in the MODIS georeferencing. With the exception of the Green Lakes Valley surveys, the extent of each of the surveys is centered on a SNOTEL pillow; for these surveys, PHV-baseline and PHV-RCN were simulated on the date of survey with removal of the co-located SNOTEL station such that no station was within the bounds of the survey that is used for validation. This ensures independent validation of the regression SWE estimates.

These two separate evaluation procedures (i.e. comparison against SNOTEL and snow surveys) provide complementary insight into the performance of the regression model. Cross-validation against SNOTEL provides high temporal information about model performance over the whole domain, but suffers from a mismatch in measurement support compared to the regression pixel and is more prone to data quality issues. The snow survey data are better suited for validating the model results because measurements are distributed over the model pixel. This provides a better representation of the true model pixel SWE, but surveys are only available at limited locations on select days. Other potential validation datasets from LiDAR, which measures snow depth, generally do not include corresponding estimates of density in order to calculate SWE accurately. The NASA Cold Land Processes Experiment (CLPX) is available in our modeling timeframe but only covers a 1km x 1km area and therefore would not add a convincing evaluation for a 500 m product.

### 2.3.3.2 Regional Variability in Model Performance

We calculate the percent bias (eqn 2.1) at each station to look at regional patterns in differences between PHV-baseline and PHV-RCN. The percent bias relative to the measured SWE at
Table 2.1: Summary of the snow surveys used for independent validation. Average interpolated survey SWE, and SNOTEL SWE (if applicable) are reported in meters. Notes: \textsuperscript{b} survey details can be found in Erickson et al. [2005], \textsuperscript{c} survey details can be found in Molotch and Bales [Molotch and Bales, 2005], \textsuperscript{d} survey details can be found in Meromy et al. [2012].

<table>
<thead>
<tr>
<th>Date Surveyed</th>
<th>Survey Avg SWE (m)</th>
<th>SNOTEL SWE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alpine Surveys</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green Lakes Valley\textsuperscript{b}</td>
<td>2001-05-09</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>2002-05-01</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>2003-05-14</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>2004-05-12</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>2005-05-10</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>2006-05-11</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>2007-05-10</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>Forested Surveys</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slumgullion (SNOTEL)\textsuperscript{c}</td>
<td>2001-04-22</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>2002-04-06</td>
<td>0.06</td>
</tr>
<tr>
<td>Upper San Juan (SNOTEL)\textsuperscript{c}</td>
<td>2001-04-23</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>2002-04-05</td>
<td>0.18</td>
</tr>
<tr>
<td>Wolf Creek Summit (SNOTEL)\textsuperscript{c}</td>
<td>2001-04-24</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>2002-04-04</td>
<td>0.18</td>
</tr>
<tr>
<td>Lily Pond (SNOTEL)\textsuperscript{c}</td>
<td>2001-04-27</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>2002-04-03</td>
<td>0.07</td>
</tr>
<tr>
<td>Niwot Ridge (SNOTEL)\textsuperscript{d}</td>
<td>2008-04-07</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>2008-05-05</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>2009-03-06</td>
<td>0.15</td>
</tr>
<tr>
<td>Joe Wright (SNOTEL)\textsuperscript{d}</td>
<td>2008-04-03</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>2008-05-01</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>2009-02-28</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>2009-05-02</td>
<td>0.66</td>
</tr>
<tr>
<td>Dry Lake (SNOTEL)\textsuperscript{d}</td>
<td>2008-04-04</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>2008-05-02</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>2009-02-28</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>2009-03-28</td>
<td>0.48</td>
</tr>
<tr>
<td>South Brush Creek (SNOTEL)\textsuperscript{d}</td>
<td>2008-04-04</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>2009-03-29</td>
<td>0.30</td>
</tr>
<tr>
<td>Lizard Head (SNOTEL)\textsuperscript{d}</td>
<td>2008-03-16</td>
<td>0.51</td>
</tr>
</tbody>
</table>
each SNOTEL station provides a normalized measure of the difference in bias at each station. The important outcome from this metric is a determination of which model estimates SWE closest to the observed value, independent of sign, and is used to test the utility of PHV-RCN versus PHV-baseline. The frequency when PHV-RCN has a lower percent bias compared to PHV-baseline is mapped to highlight potential geographic regions for which PHV-RCN is better/worse than PHV-baseline.

\[
\text{percent bias} = \frac{\text{resid}_{PHV\text{-}baseline} - \text{resid}_{PHV\text{-}RCN}}{SWE_{SNOTEL}}
\]  

(2.1)

where \(\text{resid}_{PHV\text{-}baseline}\) is the difference between SWE modeled with PHV-baseline and the SWE observation at a SNOTEL station, \(\text{resid}_{PHV\text{-}RCN}\) is the difference between SWE modeled with PHV-RCN and the SWE observation at a SNOTEL station, and \(SWE_{SNOTEL}\) is the observed SWE at the SNOTEL station; SWE is measured and modeled in meters. A measure of spatial autocorrelation, Global Moran’s I, is used to test for statistically significant clustering of the percent bias of these stations.

2.4 Results

2.4.1 SNOTEL Cross-validation Model Evaluation

The mean \(r^2\) of all simulations for PHV-baseline is 0.22 and for PHV-RCN is 0.33, an average improvement in \(r^2\) of 0.11 (Table 2.2). This can be seen in Figure 2.3a where the red dots, representing the \(r^2\) of PHV-RCN, are always higher than PHV-baseline (blue dots). The greatest increase in average \(r^2\) in any year is 0.19 in 2011 while the smallest increase was 0.02 in 2002. The timeseries of \(r^2\) (Figure 2.3a) show large differences in intra-annual variability in \(r^2\) from year to year. PHV-baseline and PHV-RCN mostly follow the same pattern within a given year, but the \(r^2\) for PHV-RCN is 36% less variable within a given year than for PHV-baseline with a mean standard deviation of 0.14 compared to 0.22, respectively. The minimum \(r^2\) generally occurs at the end of the snow season and the maximum \(r^2\) occurs on average on day of year 106. The standard devia-
Table 2.2: Cross-validated statistics from model estimates compared with SNOTEL observations.

<table>
<thead>
<tr>
<th></th>
<th>Cross-validated Scores</th>
<th>Alpine Snow Surveys</th>
<th>Forested Snow Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^2$ []</td>
<td>rRMSE [/]</td>
<td>RMSE [m]</td>
</tr>
<tr>
<td>PHV-baseline</td>
<td>0.22</td>
<td>1.14</td>
<td>0.47</td>
</tr>
<tr>
<td>PHV-RCN</td>
<td>0.33</td>
<td>1.07</td>
<td>0.41</td>
</tr>
<tr>
<td>PHV-fSCA</td>
<td>-</td>
<td>-</td>
<td>0.44</td>
</tr>
</tbody>
</table>

The day of year of maximum $r^2$ are 47 days for PHV-baseline and 40 days for PHV-RCN. The largest difference between PHV-baseline and PHV-RCN occurs in May, when $r^2$ improves by 121% from 0.14 to 0.31.

PHV-RCN decreases the rRMSE on all days compared to PHV-baseline and the maximum improvement in rRMSE is 0.28 (Figure 2.3b). The average rRMSE for PHV-baseline is 1.14 with a median of 0.72 while the average rRMSE of PHV-RCN is 1.07 with a median of 0.64 (Table 2.2). The average decrease in rRMSE when adding the reconstruction to the regression is 0.07, a 6% improvement; visually, the red points are lower than the blue points (Figure 2.3b). In the first half of the season, the average rRMSE for PHV-RCN is 0.5 before increasing drastically towards the end of the melt season; this increase is expected as the values are reported relative to mean observed SWE, which decreases late in the snow season. The yearly minimum rRMSE ranges from 0.37 to 0.49 and occurs on average on day of year 43, ranging between day of year 7 and day of year 66. On average, the maximum improvement occurs in April with monthly average rRMSE improving 14% from 0.7 to 0.6. The average yearly RMSE for PHV-RCN ranges from 0.07 m to 0.20 m with an average of 0.12 m. This represents an average annual reduction in RMSE between 0.03 m and 0.01 m, relative to PHV-baseline (data not shown).

### 2.4.2 Distributed Snow Survey Measurement Model Evaluation

Figure 2.4 and Table 2.2 compares the summary of SWE estimates from PHV-baseline and PHV-RCN against 29 different snow surveys (Table 2.1). PHV-RCN exhibited notable improvements when compared against snow surveys conducted in alpine environments (left subpanel).
Figure 2.3: Timeseries of cross-validated skill scores comparing estimated SWE and observed fSCA-scaled SNOTEL SWE. The top panel in each sub-figure shows the two regression models. The bottom panel shows the difference between the two regression models. 

a. The squared correlation coefficient ($r^2$) is shown for each day the models are fit. A positive difference indicates a higher $r^2$ for PHV-RCN.

b. The RMSE relative to the mean SNOTEL SWE (rRMSE) is shown for each day the models are fit. A negative difference indicates a lower rRMSE for PHV-RCN.
where the average RMSE values for the seven years surveyed in Green Lakes Valley, Colorado are 0.47 m and 0.41 m for PHV-baseline and PHV-RCN, respectively, a 13% improvement (Wilcoxon Test, \( p=0.05 \)). PHV-baseline yielded a higher RMSE five of seven years at Green Lakes Valley with percent bias of -38.0% compared to 3.4% for PHV-RCN. In contrast, the average RMSE values in the forested surveys (right subpanel) are 0.19 for PHV-baseline and 0.18 for PHV-RCN, respectively, not a significant difference (Wilcoxon Test, \( p=0.65 \)). The average RMSE for all the surveys is 0.25 m for PHV-baseline and 0.23 m for PHV-RCN. Of the 29 surveys, PHV-baseline exhibits the lowest RMSE 10 times and PHV-RCN 19 times. The bias of PHV-baseline (-5.5%) is reduced to 0.8% for PHV-RCN. Hence, addition of the SWE reconstruction significantly reduced PHV-baseline SWE underestimates in alpine terrain, which is poorly represented by the SNOTEL network; these estimates of model performance must, however, be viewed with caution as they are based on limited snow survey data.

### 2.4.3 Regional Differences in Model Performance

Figure 2.6 shows the frequency of the difference in percent bias at each station for all days binned into 5% increments; positive differences indicate PHV-RCN had a lower percent bias than PHV-baseline. The mean difference in percent bias between cross-validated PHV-baseline and PHV-RCN is 7.0% and the average and median differences each year are always greater than zero, i.e. PHV-RCN always has a lower percent bias. Each year the median percent bias difference is lower than the mean, suggesting that PHV-RCN yielded several large decreases in percent bias by magnitude but there were more small decreases in percent bias as well. In general, the distribution of the histogram is uni-modal with all years displaying a strong positive tail, i.e. there are many large decreases in percent bias in PHV-RCN compared to PHV-baseline. The mean skew is 1.5 (\( p<0.05 \)). The increased biases observed in PHV-baseline were similar for all elevation bands and months. The year 2012 exhibits particularly large mean (11.3) and median (10.4) differences in percent biases suggesting that the spatial pattern was greatly influenced by the reconstruction; note that this year had exceptionally low SWE with a high degree of spatial variability.
Figure 2.4: Boxplots of SWE comparing PHV-baseline and PHV-RCN against independent snow survey data. Boxplots show the range of SWE values within the surveyed extent. Each box represents the interquartile range with vertical bars for the 5th and 95th percentiles. Outliers are not shown for clarity. The horizontal line is the median value. The triangle is the mean SWE value and the circle is the RMSE of the model compared with the interpolated observations. The asterisk on each panel indicates the model with the lowest RMSE. OBS represents the range of 30 m pixels that were interpolated from the point observations from the snow surveys. OBS and the 2 models are limited to model pixels that are at least 50% covered by observations. "n" represents the number of samples per boxplot.
Figure 2.5: Histograms of the difference in percent bias between PHV-baseline and PHV-RCN. Model bias is relative to the SNOTEL measured SWE scaled by fSCA. Positive values indicate PHV-RCN had a lower bias than PHV-baseline. The red and blue lines indicate the mean and median differences, respectively. Median (med) and average (avg) values are shown.
Figure 2.6 maps the frequency at each SNOTEL station whereby PHV-RCN exhibited a smaller May 1 percent bias than PHV-baseline. All regions in the Upper Colorado River Basin include stations with overwhelming improvement in percent bias suggesting that there is no regional clustering in the performance of the model. The spatial statistic Global Moran’s I confirms this observation at the 0.1 significance level using neighbors defined by inverse distance weighting with a power of 0.5 and a binary neighborhood up to 300 km. Approximately 55% of stations (130/237) see a decrease in percent bias more than 50% of the time (>6 years) with PHV-RCN. Conversely, only 36% (86/237) of stations exhibit an increase in percent bias with PHV-RCN more than 50% of years; 9% (21/237) of stations are split with a decrease in percent bias six of twelve years and an increase in percent bias six of twelve years.

2.4.4 Spatial SWE Differences

2.4.4.1 SWE Depth

Figure 2.7 provides maps of SWE on April 1 2011 and March 1 2012, contrasting wet and dry years, respectively, for PHV-baseline and PHV-RCN. Despite the obvious difference in snow magnitude between 2011 and 2012, the locations of positive/negative differences in SWE between the models are similar in both the wet and dry year; note the purple areas in the difference map (Figure 2.7e,f). On average, PHV-RCN estimates 31% more SWE above 3000 m versus PHV-baseline, but 9% less SWE below 3000 m (Figure 2.8). Hence, these results show that use of the SWE reconstruction as an independent variable imposes a different hypsometric distribution of SWE in PHV-RCN than is present in PHV-baseline. For example, in the wet year of 2011, the average SWE below 3000 m is 0.22 m for PHV-baseline and 0.19 m for PHV-RCN. Above 3000 m, PHV-baseline estimates an average 0.54 m and PHV-RCN estimates 0.76 m. These are differences of 14% and 41%, respectively. Similar spatial differences occur in the dry year of 2012 in that the average SWE below 3000 m is 0.03 m for PHV-baseline and 0.02 m for PHV-RCN, a decrease of 33%. Above 3000 m, PHV-baseline estimates 0.28 m and PHV-RCN estimates 0.39 m, an increase
Figure 2.6: A map of the percent of years (2001-2012) at each SNOTEL station where PHV-RCN had a lower percent bias than PHV-baseline on May 1.
of 39%.

### 2.4.4.2 SWE Volume

The average April 1 volume of SWE for the Upper Colorado River Basin is 47.0 km$^3$ with PHV-baseline and 45.4 km$^3$ with PHV-RCN, a difference of 3.4%. The difference in volume between the models in any given year ranges from 6.7 km$^3$ or 6.8% less in PHV-RCN in 2008 to 1.8 km$^3$ or 14% more in PHV-RCN in 2012. The elevation zone from which PHV-baseline estimates the most SWE overall is 2000-2500 m (34% of the total 12 year volume) while PHV-RCN estimates the most SWE from 2500-3000 m (32% of the total 12 year volume). On average, PHV-RCN estimates an 11% lower contribution of April 1 SWE volume from elevations below 3000 m. Meanwhile, PHV-RCN estimates an average 18% greater contribution of SWE volume from above 3000 m. A comparison of the wettest and driest years, 2011 and 2012 respectively, shows that the distribution of SWE is more skewed to lower elevations in 2011 whereas 2012 is skewed to elevations above 3000 m; this is expected because larger snow years are more likely to have snow at lower elevations that encompass more area. As will be discussed in the next section, these spatial differences in SWE distribution have broad implications for the timing of runoff and for the management of water resources.

### 2.5 Discussion

The methods presented herein show distinct differences in SWE patterns. In this regard, we present PHV-RCN as a major improvement on PHV-baseline, the basis for which is well established in the literature [Elder et al., 1998; Erxleben et al., 2002; Fassnacht et al., 2003; Harshburger et al., 2010]. Our results compare favorably with previous works that have estimated the distribution of SWE across the mountain range scale. Guan et al. [2013], who retrospectively estimated SWE distribution, similarly reported a bias near 0 (i.e. 0.2%) and an RMSE of 0.205 m averaged over 17 snow surveys. Fassnacht et al. [2003] did not compare against independent snow surveys but
Figure 2.7: SWE for a wet and dry year, April 1 2011 and March 1 2012, respectively. The difference between PHV-baseline and PHV-RCN estimates are shown right. The polygons represent the different Hydrologic Unit Code 4 basins in the Upper Colorado River Basin labeled in Figure 2.1.
Figure 2.8: Elevation bands of average SWE depth on April 1 from 2001-2012 for each HUC4 basin by elevation band within the Upper Colorado River Basin. The basin numbers correspond to those in Figure 2.1. The error bars show the yearly April 1 minimum and maximum SWE depth averages for each sub-basin.
reported an average cross-validated RMSE of 0.14 m, higher than the 0.12 m average reported here for PHV-RCN. Rasmussen et al. [2011] simulated precipitation using the Weather Research Forecasting (WRF) model over the Colorado headwaters but did not report statistics about SWE estimates compared with individual SNOTEL stations. Notwithstanding, comparisons with 4 km gridded SWE from WRF indicate a bias of 23% relative to the snow surveys compared to the 3.4% with PHV-RCN. When resampled to the WRF grid, each have comparable RMSE of 0.2 m overall and in each terrain class. The correlation with the gridded snow surveys for PHV-RCN and WRF are 0.71 and 0.74, respectively. More detailed comparisons between the data products developed here and WRF, while beyond the scope of the current research, are warranted in future studies.

Finally, we compare the performance of PHV-RCN at our survey locations with the U.S. National Weather Service’s operational Snow Data Assimilation System (SNODAS), which produces a spatially distributed SWE product available at 1 km resolution since 2004. SNODAS assimilates a physically based model with SNOTEL observations and remotely sensing snow covered area images. It attempts to capture the snow dynamics across the entire coterminous United States and is the only high-resolution, gridded SWE product available at a daily time step for the continental United States; data is available from http://nsidc.org [Barrett, 2003]. Previous work suggests that SNODAS works well in environments similar to that of the SNOTEL stations it assimilates but suffers in alpine regions [Clow et al., 2012]. In this regard, we compare PHV-baseline and PHV-RCN to the SNODAS SWE product to evaluate if PHV-baseline and PHV-RCN provide improved SWE information above tree-line, an area previously demonstrated in section 2.4.2 to be substantially improved in PHV-RCN compared to PHV-baseline. Since we cannot extract the effect of an individual SNOTEL station from the SNODAS product, we compare PHV-baseline and PHV-RCN to the survey data without excluding the station that is at the center of the specific survey. The average RMSE at both forested and alpine surveys for PHV-baseline, PHV-RCN and SNODAS are 0.25 m, 0.23 m and 0.23 m, respectively. The similarity in these RMSE values is not entirely surprising given that all three models incorporate SNOTEL SWE values, which are co-located with 14 of the 18 snow surveys available after 2004. However, for the four snow surveys conducted in the Green
Lakes Valley far from SNOTEL stations, the average RMSE for SNODAS was 0.61 m compared to 0.42 m for PHV-RCN, confirming that the performance of SNODAS can suffer away from stations. Given that SNODAS attempts to capture the snow dynamics across the entire continental United States, the method has a distinct handicap compared to PHV-RCN, which was specifically developed for use in the mountains of the western United States. Nevertheless, implementation of the statistically based PHV-RCN is significantly more straightforward than SNODAS given the exhaustive number of model forcings used in the SNODAS model and the staffing required to conduct the real-time model assimilation of SNOTEL values into the SNODAS system [Barrett, 2003].

Both PHV-baseline and PHV-RCN rely heavily on point measurements of SWE reported from SNOTEL stations, but the SNOTEL network only represents a limited range of elevation within the domain. Elevation (along with latitude) is the most consistently utilized independent variable in the stepwise regression that constitutes PHV-baseline. Figure 2.9 shows that the centroid elevation of SNOTEL stations (2800 m) is 200 m above the centroid elevation of snow accumulating area (2600 m; defined by areas of the DEM that PHV-RCN estimates to have at least 10 cm average SWE on April 1 from 2001-2012) within the domain. Moreover, the highest SNOTEL station is at 3536 m, 724 m below the maximum elevation of the domain. This is relevant because, on average, PHV-RCN estimates 31% more SWE above 3000 m, but 9% less SWE below 3000 m. Nonetheless, the biases at the SNOTEL stations from PHV-baseline and PHV-RCN have a minimal relationship with elevation; this is expected because elevation is an independent variable in the regression, and therefore, based on first principles of the linear regression technique, biases should be uncorrelated with independent variables.

The centroid of the snow surveys (3100 m) used for validation is located 300 m above that of SNOTEL stations, thereby providing observations in an important area where there is greater SWE depth outside the extent of the SNOTEL network. It should be noted that validation data at elevations above that of the SNOTEL network were only available in one alpine location (Green Lakes Valley). In addition to the 113,582 observation points used for validation from Green Lakes Valley, further spatial SWE information from a diverse set of alpine areas are needed to confirm
Figure 2.9: The elevation distribution of the snow accumulating area in the domain based on the Digital Elevation Model (DEM), SNOTEL stations and the survey locations binned into 250 m bands. Snow accumulating area is defined by areas that are estimated to have at least 10 cm average SWE on April 1 from 2001-2012.
the results presented here. Nevertheless, we posit that increased benefit of the reconstruction in the alpine may be due to physical processes not captured by SNOTEL stations. Wind redistribution of snow is a dominant process for distributing SWE above tree-line in Colorado [Knowles et al., 2012]. In this regard, the reconstruction implicitly captures wind redistribution patterns with observations of fSCA and is able to provide additional information to the interpolation above tree-line. However, the increased error (and percent error) in the alpine suggests that more information is needed to properly scale measurements from the subalpine to the alpine; in particular, spatial measurements in basins spanning subalpine and alpine terrain types will provide insight to the driving processes without being influenced by the climatology specific to different basins.

Another important consideration for the efficacy of PHV-baseline and PHV-RCN is the scaling behavior of the SNOTEL stations with respect to the surrounding terrain. The poor representation of SNOTEL SWE observations relative to the SWE of surrounding terrain is a well-documented issue in the literature [Molotch and Bales, 2005; Meromy et al., 2012]. Of the 29 surveys used in this study, 22 were performed around SNOTEL stations (Table 2.1). The SNOTEL biases for each of the 22 surveys were calculated as per Meromy et al. [2012]. However, we cropped the 4km x 4km interpolated survey to an extent such that the model pixels used for comparison overlap at least 50% with interpolated pixels from the survey data; as such, the SNOTEL biases used here may not match with those previously published in Meromy et al. [2012]. The median station bias is 14%, and as a result one would expect PHV-baseline and PHV-RCN to have a positive average bias compared to the surveys as both models are based on the inherently biased observations at the SNOTEL stations. However, scaling the SNOTEL observed SWE by the observed fSCA, as was performed here, decreases the average positive bias exhibited by SNOTEL from 27% to 20% and reduces the cross-validated rRMSE (see section 2.4.1) from 2 to 1.2 for PHV-baseline and from 1.9 to 1.1 for PHV-RCN, both improvements on the order of 40%. The average RMSE with the snow surveys did not change. It should be noted that SNOTEL biases vary tremendously from station to station and throughout time. Hence, the evaluation of SNOTEL biases shown here are not exhaustive in that almost all of the snow surveys took place in Colorado.
(Figure 2.1), and importantly, none took place in Utah, Arizona or New Mexico.

All regions of the domain contain some stations with little or no improvement in percent bias, however the lack of visual and statistically significant clustering of these stations suggests that these stations are random in space. We recognize that some stations decrease in overall model performance. However, it is expected that regression-based techniques will inherently result in positive and negative biases, and therefore some station-years will increase in bias even when overall biases decrease. Given the complex mountainous climate that these stations exist in, it would also not be surprising if the stations with systemic deterioration in model performance experience distinct microclimates that do not translate to other physiographically similar stations. Storm track simulations [Stein et al., 2015] and high resolution weather models [Rasmussen et al., 2011] may provide mechanistic insight to the processes that generate these anomalous SWE patterns. High resolution remote sensing such as NASA JPL Airborne Snow Observatory may aid with interpolating these biased station observations, particularly those stations where modelling results deteriorated.

A source of uncertainty in PHV-RCN that is not present in PHV-baseline is error in the reconstruction product. That said, the simulation date of the reconstructed SWE is always from years prior to the simulation date of the regression. Consequently, the absolute errors in the reconstruction are not relevant in this application since one would never expect the absolute SWE to be the same between years. Rather, we are leveraging the similarity in the relative SWE pattern between years. Several previous publications have shown that reconstructed SWE contains relevant information about the spatial distribution of SWE. The mean absolute errors reported in these publications are 23% in the San Juan Mountains of Colorado [Molotch, 2009] and 47% in the Sierra Nevada of California [Guan et al., 2013]. The SWE error in two small headwater catchments in the Sierra Nevada and Colorado Rockies were reported to be between -37% and 34% [Jepsen et al. 2012]. We direct the reader to Appendix A for a more comprehensive discussion on the sources of uncertainty associated with the SWE reconstruction model. It is important to remember that only past reconstructed SWE patterns (i.e. from previous years) are used in the real-time regression, and the reconstruction is scaled to mean 0 and standard deviation 1. As such, only errors in the relative
SWE from pixel to pixel will impact PHV-RCN.

Given the computational cost of reconstructing SWE, we tested a third regression product that uses the simulation day’s fSCA from MODSCAG as a predictor (hereafter PHV-fSCA) instead of past reconstructed SWE. This regression is statistically invalid if we also scaled SNOTEL SWE by fSCA; as such, PHV-FSCA is simulated without scaling the SNOTEL SWE. The RMSE of PHV-fSCA against the independent snow surveys in alpine regions is 0.44 m and in forested terrain 0.17 m. The bias is 11.6% in the alpine region and 2.8% in forested areas. PHV-RCN had lower RMSE than PHV-fSCA in alpine terrain where the errors are generally greatest (RMSE = 0.41 m vs 0.44 m). PHV-fSCA exhibits a slightly lower RMSE than PHV-RCN in forested areas with a RMSE of 0.17 m versus a RMSE of 0.18 m, respectively. However, PHV-RCN had a lower bias than PHV-fSCA in both alpine and forested regions with biases of 3.4% and 0% vs 11.6% and 2.8%, respectively.

The results presented herein indicate that past patterns of SWE distribution contain information with regard to real-time SWE distribution. The important characteristic of the reconstructed SWE is that the spatial patterns are independent from ground observations. In this regard, any spatial SWE product that is independent from the SNOTEL network could be used in this regression framework. Future work should compare different spatial products of SWE in this context, e.g. those from numerical weather models [Rasmussen et al., 2011], LiDAR [Deems et al., 2006; Harpold et al., 2014], and emerging radar methods [Cline et al., 2009]. An unknown in this type of statistical modeling is the size of the ensemble of past patterns required for robust real-time estimation. This may vary by location and time period since the ensemble’s utility is limited by the degree to which the historical patterns reflect real-time SWE patterns. Further work utilizing atmospheric variables, e.g. geopotential heights, should investigate the physical processes behind the similarities in SWE patterns. The performance of PHV-fSCA suggests that remotely-sensed fSCA provides considerable information for the distribution of SWE within the regression framework, but further work beyond the scope of this paper is required to distinguish the relative contributions between fSCA and the calculated melt flux in the reconstructed SWE. Nonetheless, it should be
noted that the quality of operational “near-real-time” fSCA may hinder SWE estimates solely from fSCA. We expect that the value added in the context of regressions would only increase with more accurate distributed SWE products from the past, irrespective of the method used. Lastly, the results shown herein provide evidence that an optimal ground network, with the goal of estimating the distribution of SWE, may be one that can leverage both direct observations, physically-based modeling, and remotely sensed data. Future work should identify the necessary distribution of ground observations in this context to inform developing regions as they expand their ground observation network.

The results presented here have significant implications for water management. PHV-RCN estimates on average 31% more SWE depth than PHV-baseline at elevations >3000 m on April 1 which translates to 10% more SWE volume from areas above 3000 m. This increased SWE at higher elevation is important in the context of hydrologic vulnerability to climate change as the high elevation snowpack may be less sensitive to climate change [Clow, 2010]. These hypsometric differences will also affect the timing of runoff [Guan et al., 2013]. Nonetheless, PHV-RCN estimates an average volumetric difference of 1.6 km$^3$ less SWE than PHV-baseline, which is equivalent to 17% of the water allocated to the upper Colorado River basin states. Because of substantially higher SWE depth estimates at the highest elevations (Figure 2.7, Figure 2.8), PHV-RCN tends to estimate greater SWE volumes than PHV-baseline in years with less snow accumulation. This occurs because SWE accumulation is skewed to higher elevations in low snow years.

The percent difference in total volume between the two methods is inversely correlated to the magnitude of SWE (r=-0.88). In the driest year, 2012, both methods estimated less than 15 km$^3$ of SWE, with PHV-RCN estimating 14.1% more SWE. Conversely, in the highest year by volume, 2011, PHV-RCN estimated -7.4% less SWE than PHV-baseline. These differences are perhaps intuitive given that PHV-baseline is more sensitive to SNOTEL SWE values (as indicated in section 2.4.2 above) and SNOTEL values tend to under-estimate basin SWE during drought years because basin SWE is skewed toward higher elevations, which the stations do not represent. Similarly, in larger snow years, basin SWE volumes skew toward the lower elevations, which are
also not represented by the mid-elevation SNOTEL stations (Figure 9). In this regard, the SNOTEL network design-dependent biases have a larger influence on the SWE estimated with PHV-baseline compared to PHV-RCN.

Additionally, PHV-RCN has considerable less variability in estimation skill with 36% lower standard deviation in $r^2$ providing higher confidence in the SWE estimates. We also see the largest statistical improvements in PHV-RCN in April and May when snowmelt is transitioning from lower to higher elevations, thereby providing potentially more robust information for estimates of snowmelt timing, which are important for reservoir operations to balance flood risk, ecological needs and water supply for agriculture and domestic uses.

2.6 Conclusion

When compared to independent distributed snow surveys, PHV-RCN reduces the bias from -5.5% to 0.8% and RMSE from 0.25 m to 0.23 m. Notable improvements are seen with PHV-RCN when compared against snow surveys in alpine terrain, an area poorly represented by the SNOTEL network. The mean cross-validated $r^2$ increases by 50% from 0.22 with PHV-baseline to 0.33 with PHV-RCN. The largest increases in $r^2$ occur in May when monthly average $r^2$ improves 121% from 0.14 to 0.31. PHV-RCN estimates on average 31% greater SWE depth than PHV-baseline in areas above 3000 m elevation, which contributes up to 66% of annual SWE volume in the driest year. These results show that observations of snow cover depletion from previous years, inherent in the reconstructed SWE, provide important information for estimating SWE in real time. Given that previous works using SWE reconstructions were limited to retrospective analyses by necessity, the work presented here represents an important contribution in that it extends SWE reconstructions to real-time applications and illustrates that doing so significantly improves the accuracy of SWE estimates. This result is not unique to the western United States as similar processes associated with the interactions between topography and snow distribution govern the transferability of ground measurements to unsampled locations worldwide. Improved high resolution distributed SWE information will afford more accurate water supply forecasts and enable more efficient water
resource management. Such information will be increasingly needed as changes in climate and land cover alter the hydrology of mountainous regions.

2.7 Acknowledgements

This work was supported by NASA NNXLLAK35A, USDA 2012-67003-19802, NSF EAR1141764, NOAA Western Water Assessment and NSF Niwot Ridge Long-Term Ecological Research. D. Schneider was also supported by a NASA Earth and Space Science Fellowship. This work utilized the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794) and the University of Colorado Boulder. The Janus supercomputer is a joint effort of the University of Colorado Boulder, the University of Colorado Denver and the National Center for Atmospheric Research. The authors would like to thank Randy Julander for insightful conversations regarding SNOTEL data and B. Rajagopalan for statistical advice. We also thank GDAL developers [GDAL Development Team, 2015] and the R Core Team [R Core Team, 2015]; in particular authors of the raster [Hijmans, 2015], rgdal [Bivand et al., 2015], reshape2 and tidyr [Wickham, 2007], plyr and dplyr [Wickham, 2011], ggplot2 [Wickham, 2009] and MASS [Venables and Ripley, 2002] packages. All code and data is available upon request at dominik.schneider@colorado.edu.

2.8 Appendix A

The SWE reconstruction model is a physically based model and is therefore subject to structural, parameter and forcing uncertainty. For example, the Richardson parameterization used in this model can result in suppressed turbulent fluxes and reduced snowmelt in colder climates where the snow surface temperature is often lower than the air temperature, e.g. Colorado [Lapo et al., 2015]. This stability feedback has been previously hypothesized as a reason for snow model differences [Slater et al., 2001] and may behave differently in different parts of our modeling domain, e.g. by elevation and land cover. More generally, the impact of model parameterization is the topic of
ongoing research [Lapo et al., 2015; Mendoza et al., 2015].

Uncertainty in the radiation forcing of the SWE reconstruction model is inherent given that shortwave and longwave downwelling radiation are both downscaled from the relatively coarse-scale (i.e. 1/8° degree) North American Land Data Assimilation System v2 [Xia et al., 2012]. Lapo et al. [2015] report absolute biases in shortwave and longwave radiation from satellite and reanalysis products to be upwards of 80 W m⁻² and 56 W m⁻², respectively. They also report snowmelt to be about twice as sensitive to biases in longwave radiation than shortwave radiation, although opposing biases in the two can balance each other out [Lapo et al., 2015]. Moreover, differences in the snow surface temperature parameterization will result in different emitted longwave estimates. Different albedo parameterizations introduce uncertainty into the radiation partitioning [Molotch and Bales, 2006], but the parameterization used here [U.S. Army Corps of Eng., 1956] has been shown to match observed albedo by -1±3% at a research site within the domain [Molotch et al., 2010; Painter et al., 2010].

Another source of uncertainty lies in the fSCA retrievals, which are often occluded by cloud cover. However, there are gaps of more than 5 days only 5%-10% of the time [Slater et al., 2013]. Cloud masking introduces errors associated with falsely classifying clouds as snow or snow as clouds and potentially prohibits observation of the date of melt out. Slater et al. [2013] suggest that a five day delay in observing snow free conditions is equivalent to approximately 13 W m⁻² net radiation bias. The MODSCAG product used here is the most accurate available product at 500 m with an RMSE of 0.076 in Colorado versus 0.247 for the operational NASA MOD10A1.5 product and can detect snow down to about 15% snow covered area in a pixel [Rittger et al., 2013].

2.9 References

Barrett, A. P. (2003), National Operational Hydrologic Remote Sensing Center SNOW Data Assimilation System (SNODAS) Products at NSIDC, National Snow and Ice Data Center, Boulder, CO, USA.

Western Snow Conference.


Erickson, T. A., M. W. Williams, and A. Winstral (2005), Persistence of topographic controls on


Fassnacht, S. R. (2004), Streamflow Predictability in the Upper Versus Lower Colorado River Sub-Basins, Portland, Maine, USA.


GDAL Development Team (2015), GDAL - Geospatial Data Abstraction Library, Version 1.11.3.


Mizukami, N., and S. Perica (2012), Towards improved snow water equivalent retrieval algorithms for satellite passive microwave data over the mountainous basins of western USA, *Hydro-


Painter, T. H., K. E. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier (2009), Retrieval of subpixel snow covered area, grain size, and albedo from MODIS, *Remote Sensing*

R Core Team (2015), R: A Language and Environment for Statistical Computing, Vienna, Austria.


Slater, A. G. et al. (2001), The representation of snow in land surface schemes: Results from PILPS 2(d), J. Hydrometeor, 2(1), 7–25.


Chapter 3

Estimating relationships between snow water equivalent, snow covered area, and topography to extend the Airborne Snow Observatory dataset

3.1 Introduction

The spatial distribution of snow water equivalent (SWE) is important for plant ecology [Litaor et al., 2008] and controls the timing and magnitude of streamflow in snow-dominated catchments around the world [Stewart et al., 2004]. In these catchments, accurate assessments of snowpack water storage are critical for ensuring robust estimates of seasonal water supply. Nevertheless, SWE is poorly measured operationally with only sparse point measurements on the order of one per thousand square kilometers of snow covered terrain. Additional manual measurements in the form of snow courses add little information about the spatial variability and typically occur on the order of only three times a season.

Satellite remote sensing provides spatially explicit information about the snowpack but current satellites are unable to measure SWE directly at the scales relevant for water resources management in mountainous terrain (e.g. the western United States) [Dozier, 2011]. The need for improved information regarding the quantity and distribution of SWE has led to the development of new measuring techniques including the application of ground penetrating radar (GPR) [Marshall and Koh, 2008], Global Positioning Systems (GPS) [Gutmann et al., 2012; Koch et al., 2014], Light Detection and Ranging (LiDAR) [Schirmer et al., 2011; Deems et al., 2013], and photogrammetry [Bühler et al., 2015; Nolan et al., 2015]. Although these techniques are typically limited to snow depth, they can still capture the majority of the variability in SWE because snow depth varies
an order of magnitude greater than density [Mizukami and Perica, 2008]. Snow depth distributions can be converted to SWE using modeled snow density or in situ snow pit observations [Elder et al., 1991; Sturm et al., 2010; Painter et al., 2016].

Airborne systems have significantly increased the ability to measure SWE distribution at high resolution and at extents relevant to water resource management. Since 2013, the NASA Jet Propulsion Laboratory Airborne Snow Observatory (ASO) flies approximately weekly from peak SWE to melt out, in the Tuolumne Basin in Sierra Nevada California [Painter et al., 2016]. ASO measures snow depth by differencing the LiDAR-derived surfaces from a snow-on and snow-off flight and infers albedo and snow extent based on spectroradiometric measurements. An energy-balance model that can incorporate measured albedo is used to model snowpack density, and snow depth is converted to SWE. For the first time there exists a weekly time series of spatially explicit observations of SWE for multiple years over ~1000 km². Hence, repeated LiDAR observations, as provided by ASO, provide new opportunities for understanding the spatial and temporal dynamics of snow distribution. However, little work has been done to extend this dataset in time beyond the periods observed by ASO – a goal of the work presented here.

Statistical modeling has been extensively used to estimate the relationship between SWE point measurements and physiography [Elder et al., 1998; Balk and Elder, 2000; Fassnacht et al., 2003; Erickson et al., 2005; Molotch et al., 2005; López- Moreno and Nogués-Bravo, 2006; Schneider and Molotch, 2016]. At small scales, dedicated sampling of headwater catchments has led to models that explain between 20% and 65% of the variability in SWE based on physiographic variables at 30m resolution [Elder et al., 1998; Balk and Elder, 2000; Erxleben et al., 2002]. Regional-scale studies have been conducted whereby operational SWE observation networks have been used to explain up to 33% of the variability in SWE based on physiographic variables [Fassnacht et al., 2003; Schneider and Molotch, 2016]. These studies aimed to understand the processes controlling snow distribution and to apply this knowledge to interpolate point observations of SWE to regional scales for a single point in time. Importantly, Erickson et al. [2005] found persistence in the topographic controls of SWE distribution and successfully parameterized a multi-year model relating
the physiographic variables of a small headwater catchment to annual peak SWE by adjusting the mean based on in situ measurements. Later studies have since confirmed an inter-annual consistency in snow depth distribution in other basins based on high resolution LiDAR measurements [Deems et al., 2008; Trujillo et al., 2009; Schirmer et al., 2011] and the inter-annual persistence of topographic controls [Grünewald et al., 2013]. Little work has been done to date to extend LiDAR snow measurements in time using statistical relationships with topography – a focus of the work presented herein.

Remotely sensed snow covered area data has long been recognized to provide information with regard to snowpack water storage and consequently expected summer streamflow [Potts, 1937; 1944; Martinec and Rango, 1981; Good and Martinec, 1987]. Today, SCA is commonly estimated from SWE in hydrologic models through a depletion curve parameterization in order to constrain melt production to the areal extent of snow cover [Anderson, 1973; Luce and Tarboton, 2004; Livneh et al., 2010; Clark et al., 2011; Lawrence et al., 2011; Niu et al., 2011]. The utility of this relationship is predicated on the fact that snow depth distribution is extremely heterogeneous over complex terrain. Upon melt out, terrain features are progressively uncovered. This process varies only slightly each year because of similarities in the meteorology, e.g. wind direction, that drive accumulation patterns and solar exposure that drives melt out [Luce and Tarboton, 2004]. The utility of depletion curves to provide sub-model-scale information in physically-based modeling would suggest that fSCA may provide additional information in statistical models of SWE distribution. Snow covered area is also relatively easy to measure due to the distinctive appearance of snow compared to bare earth and vegetation. In fact, photography of fSCA has been utilized for seven decades within hydrologic applications [Potts, 1937; 1944; Parsons and Castle, 1959] and today robust observations of fSCA can be obtained from variety of ground-based, aerial and satellite optical imagers [Dozier et al., 1981; Bloschl et al., 1991; Kirnbauer and Bloschl, 1994; Rosenthal and Dozier, 1996; König and Sturm, 1998; Hall et al., 2001; Painter et al., 2009; Rittger et al., 2013].

Given that fSCA imagery is readily available, exploration of fSCA - SWE relationships for
The purpose of SWE estimation is warranted. The recent availability of ASO data represents an opportunity to develop these relationships given the unprecedented SWE information that ASO provides. Hence, we propose that fSCA could be a useful explanatory variable of SWE and snow depth distribution. The basis for this possibility is rooted in the fact that patterns of fSCA depletion are largely dictated by interactions between sub-pixel terrain variability, the mean snow depth, and the variability of snow depth. These relationships have been previously explored in the literature. In this context, Marchand and Killingtveit [2005] and Lopez et al. [2014] found significant correlations between sub-pixel terrain variability and the mean and standard deviation of pixel scale snow depth. In addition, [Lehning et al., 2011; Grünewald et al., 2013] showed that sub-pixel terrain variability and surface roughness yield controls on snow depth that exhibit elements of consistency from year to year. We are further motivated by the results of Helbig et al. [2015] who successfully parameterized the standard deviation of peak snow depth based on terrain roughness for use in a depletion curve parameterization to estimate fSCA from modeled mean snow depth.

The repeat observations from ASO provide a unique dataset of concurrent SWE and fSCA over multiple years with which to develop relationships between SWE and fSCA. Given that fSCA is widely observable from a variety of satellites, these relationships could then be used to estimate SWE for any date on which fSCA observations are available. Hence, the objective of this research is to use ASO-derived relationships between SWE (dependent variable), and fSCA and physiography (independent variables) to estimate SWE distribution for time periods when ASO data are not available. We aim to test how well statistical models of the relationship between SWE, fSCA and physiography transfer in time. We ask (1) Do relationships exist between fSCA and SWE in the Tuolumne River Basin? (2) Can statistical models of SWE distribution be transferred directly from one year to another? (3) How can we determine which SWE distribution from the ASO record best represents a date of interest?

We present our SWE distribution modelling framework and compare the results of two models that use fSCA as a predictor to a model that does not use fSCA. Further, we evaluate the impact of transferring models from one year to another. Lastly, we present a methodology for identifying
which models of SWE distribution, from the ensemble of historical ASO acquisitions, best repre-

sents the SWE distribution for unsampled dates of interest. We then discuss the results in the
context of extending ASO to unsampled dates.

3.2 Site Description

We used a SWE dataset from the Tuolumne River basin in the Sierra Nevada mountains in
California, USA (Figure 3.1). The basin is 1,175 km\(^2\) in area, consisting of 48% vegetation, 50% rock, 2% water, and small isolated areas with permanent snow/ice. The elevation range is 1127 m
to 3965 m, encompassing 4 distinct ecological zones ranging from lower montane forest to alpine
[NPS, 2016]. The lower montane forest ranges from 1127 m to 1800 m elevation and consists of a
diverse mix of coniferous and deciduous trees. The upper montane forest ranges from 1800 m to
2450 m elevation and primarily consists of coniferous species such as red fir and lodgepole pine.
Elevations from 2450 m to 2900 m are considered subalpine and consist of a mix of meadows and
coniferous forest. The highest elevation band above 2900 m is an alpine zone that is devoid of
tree cover and contains limited herbaceous vegetation. This alpine zone contains areas with large
granitic features, talus slopes and boulder fields. Snowmelt from the basin runs off into the Hetch
Hetchy reservoir, which is the main water supply for the City of San Francisco.

3.3 Data Sources

We utilize the National Aeronautics and Space Administration (NASA), Jet Propulsion Labo-
ratory (JPL), Airborne Snow Observatory (ASO) dataset in the Tuolumne River Basin in the Sierra
Nevada mountains of California. The ASO mission consists of airborne LiDAR in conjunction with
a hyperspectral spectroradiometer and high precision global positioning system (GPS) [Painter et
al., 2016]. The dataset consists of a 3 m resolution snow-free digital elevation model (DEM), 3 m
snow depth maps for which snow-free areas are masked using spectral information from the spec-
troradiometer, 3 m vegetation height map and 50 m SWE maps. The dataset is distributed in the
Figure 3.1: Land cover map of the Tuolumne Basin with snow pillows shown as blue dots. The land cover data was provided by ASO and was derived from spectral information from summer snow-off flight. The snow pillows locations were obtained from the California Department of Water Resources.
UTM zone 11 and WGS84 datum map projection system.

We mean-aggregated the 50 m SWE maps to 500 m. Subsequently, we converted the 3 m DEM and snow depth maps to 501 m using mean aggregation and then bilinearly resampled to the 500 m SWE grid. We chose the 500 m resolution because it is relevant to water resource management and is the scale for which daily satellite fSCA images are available from the Moderate Resolution Imaging Spectroradiometer, which is a commonly used fSCA data product [Salomonson and Appel, 2004; Painter et al., 2009]. We also used a binary aggregation of the 3 m snow depth maps to obtain 501 m fSCA maps, which were then resampled to 500 m by nearest neighbor to preserve snow free pixels. Lastly, we aggregated and bilinearly resampled the vegetation height map to 500 m.

The ASO flies approximately weekly from near peak SWE to the end of the melt season. As such, there were six flights in 2013, ten in 2014, eight in 2015, and eight in 2016. Only the final four flights of the 2016 season were available for this study resulting in a total of 28 SWE maps. Painter et al. [2016] report a mean absolute vertical snow depth error of 8 cm and a bias of 1 cm when compared with manually measured snow depths at the 15 x 15 m scale. Further details about the mission and processing can be found in Painter et al. [2016].

We also obtained daily SWE measurements from 54 snow pillows operated by the California Department of Water Resources that are within 20 km of the Tuolumne watershed boundary. The stations range from 2000 m to 3250 m. We downloaded the adjusted SWE records, which have been manually quality controlled, for the 2013-2016 water years. No further adjustments were performed. The data can be downloaded from http://cdec.water.ca.gov/.

3.4 Methods

A relationship between fSCA and SWE is well established in that depletion curves are commonly used to estimate areal snow cover based on SWE simulated from a physically based model. However, available remotely sensed images of fSCA present an opportunity to estimate the distribution of SWE directly since fSCA also decreases as SWE decreases and the sub-grid terrain
variability is increasingly exposed throughout the melt season. Notwithstanding, no one has explicitly explored the utility of fSCA for directly estimating the distribution of SWE. We use linear regression to model the distribution of SWE for every ASO flight. The explanatory variables we consider are ASO-observed fSCA and physiographic variables previously used in the literature [Table 3.1]. We compare three models: 1. ”FSCA” is a single variable regression with fSCA as the only independent variable. Note the capital ”F” in ”FSCA” is used to denote the model while lowercase ”f” in ”fSCA” is used to denote the variable; 2. ”PHV” is a multiple linear regression that consists only of physiographic variables as independent variables; 3. ”PHV-FSCA” is a multiple linear regression that includes both physiographic variables and fSCA as independent variables. These distinctions are necessary as we first illustrate the utility in the fSCA data (i.e. FSCA and PHV-FSCA models) with regard to SWE modeling in the context of models that only include physiographic variables (i.e. PHV). Conceptually, FSCA and PHV models are the components of the PHV-FSCA model and thus each provide general insight into the explanatory power of fSCA and topography, respectively, for the estimating the distribution of SWE. Once we illustrate the utility of the statistical models for characterizing ASO SWE distribution patterns at discrete points in time, we then show how these statistical models can be transferred to time periods without ASO observations. With regard to all statistical models, we report the squared Pearson correlation coefficient ($r^2$) as a measure of the relative spatial pattern between the modeled SWE distribution and ASO observed SWE distribution. We also report the mean absolute error as a percent of mean observed SWE (%MAE) as a measure of the accuracy of the modeled SWE distribution. Lastly, we report bias as a percent of mean observed SWE (%Bias) as a measure of the systematic over or under-prediction by the model.

3.4.1 SWE Models: Discrete Time

To examine the utility of fSCA as a predictor, we compared the SWE distributions modeled with FSCA, PHV, and PHV-FSCA using a split sampling strategy. We split each date into a training (80%) and test (20%) dataset to evaluate overall model performance on this date. This insures
Table 3.1: List of physiographic variables considered in the multiple linear regression to model SWE distribution. Source includes studies in which these variables have been used and the source of the algorithm, if applicable. Citations in the source column are by no means exclusive.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units/Derivation Specifics</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTM Northing</td>
<td>meters</td>
<td>Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>UTM Easting</td>
<td>meters</td>
<td>Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>zness</td>
<td>sine(slope); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>northness</td>
<td>cosine(aspect); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>eastness</td>
<td>sine(aspect); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>topographic position index (TPI)</td>
<td>elevation difference of a pixel from the mean of the surrounding pixels; meters</td>
<td>Revuelto et al. (2014); GDAL (2015)</td>
</tr>
<tr>
<td>vector ruggedness measure (VRM)</td>
<td>3-dimensional measure of the variation of slope and aspect; not correlated with slope or aspect; ranges 0-1; dimensionless</td>
<td>Veitinger et al. (2015); Sappington et al. (2007); Conrad et al. (2015)</td>
</tr>
<tr>
<td>standard deviation of slope</td>
<td>standard deviation of slope in 3x3 window around each pixel; shown to detect changes in slope at multiple scales; radians</td>
<td>Marchand and Killingveit (2005); Lopez et al. (2014); Grohman et al. (2007)</td>
</tr>
<tr>
<td>vegetation height</td>
<td>measured by ASO; used in place of forest canopy density from previous studies; meters</td>
<td>Molotch and Bales (2005, 2006); Painter et al. (2016)</td>
</tr>
</tbody>
</table>
that we are not evaluating the model with the same data used to create the model. Furthermore, this procedure is replicated 20 times to provide 20 different subsets with which to evaluate model performance; this is more representative than a single replication. More replications were computationally prohibitive. This split sample strategy is an important initial step in transferring ASO data in time as it is necessary to first show that fSCA and physiographic variables can be used to adequately model ASO-observed SWE on the date of acquisition. Once this is established, the transferability of the models in time can then be explored – as described in the next section.

3.4.2 SWE Models: Transferred in Time

We evaluated a second set of predictions whereby each date is modeled using all the data, i.e. not split, and then this model is used to predict SWE on dates that ASO flew in different years. In this manner, we simulate SWE on the date ASO flew using models from other years and then we use the ASO data on the date of interest strictly to evaluate the model estimates of SWE. Hereafter, we refer to the date of the model (i.e. the date of the ASO observation for which the model is developed) as the \textit{model date} and the date being predicted as the \textit{transfer date}. This results in 28 models of SWE distribution for FSCA, PHV, and PHV-FSCA each because there are a total of 28 ASO flights. Given our primary goal of estimating the SWE distribution for unsampled dates, we apply models developed for each ASO acquisition to all other dates except for dates within the same year as that in which the model was developed. For example, in 2013 there were a total of 6 ASO flights out of the total of 28 flights during the four-year study period. This leaves 22 flights from other years that can be used to develop statistical models of SWE that can be transferred to the dates in 2013. By conducting our model tests in this manner we are more robustly testing the transferability of models from a given \textit{model date} with regard to simulating SWE distribution on a given \textit{transfer date}. Due to a different number of ASO observations in each year, the prediction ensemble size differs for each year. As noted above, for each date in 2013 there are 22 potential models. For each date in 2014 there are 18 potential models; for each date in 2015 there are 20 potential models; and for each date in 2016 there are 24 potential models. These models are referred
to as *transferred models*, and for each transfer date we identify the best model from another year based on error statistics generated from the ASO data acquired on the transfer date. We refer to this model as the *best model*.

The ability of a model to transfer from the one date to another will vary based on how well the relationships between the dependent variable (SWE) and explanatory variables are captured and the similarity of the SWE distributions. Here we quantify the SWE similarity between dates using the mean absolute error (MAE) of SWE recorded at nearby snow pillows. For each transfer date, there is an ensemble of predictions from the model dates from the other years. Each of the model dates exhibit a similarity with the SWE distribution of the transfer date. In order to pick which model date exhibits the greatest similarity with the transfer date without having an ASO observation, we calculate the MAE of SWE at the snow pillows between each pair of model-transfer dates and select the model date with the lowest MAE. We compare the prediction performances from this model selection procedure (denoted *selected model*) with those of the best models.

### 3.4.3 Statistical Model

The single linear regression, i.e. FSCA, is performed using ordinary least squares. The multiple linear regression models described above, i.e. PHV and PHV-FSCA, are based on a regularized regression model applied in an elastic net framework as implemented in the glmnet package in R [Friedman et al., 2010]. The benefit of a regularized regression over standard regression is that it reduces overfitting while permitting all conceptual variables to be included, rather than removing potentially useful variables due to multicollinearity. Hence, there is no reason to use a regularized regression with a single variable, e.g. FSCA. Regularized regression increases the predictive ability of a model with multiple predictor variables by penalizing the objective function used to estimate the parameter set. The elastic net is an extension of ordinary least squares, which estimates parameter coefficients by minimizing the residual sum of squares (RSS) as the objective function (Eqn 3.1):
\[ RSS = \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 \]  

(3.1)

where \( y_i \) is the response variable at the \( i \)th observation, \( \beta_j \) is the coefficient for predictor variable \( j \), and \( x_{ij} \) is predictor variable \( j \) at each observation \( i \). The elastic net penalizes RSS by two different types of regularization techniques, known as L1 and L2, that have opposing properties (more on this below). In this regard, the elastic net estimates the regression parameters \( \beta \) by minimizing RSS in Eqn 3.2:

\[ RSS = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \alpha(\beta) \]  

(3.2)

where

\[ P_{\alpha}(\beta) = \sum_{j=1}^{p} \left[ \frac{1}{2} (1 - \alpha) \beta_j^2 + \alpha |\beta_j| \right] \]  

(3.3)

\( \beta_0 \) is the intercept, \( \alpha \) controls the magnitude of the penalty, and \( P_{\alpha} \) changes the relative influences of the L1 and L2 regularizations. In practice, this shrinks the coefficient values towards zero to account for multicollinearity and predictors with low explanatory power. Penalized coefficients have less variance and can select variables without resorting to a discrete selection procedure, e.g. a p-value threshold such as in step-wise regression. Consequently, resulting parameter sets are more robust predictors for independent data [Zou and Hastie, 2005a].

The elastic net has two tuning parameters, \( \alpha \) and \( \lambda \), which are determined through cross validation. When \( \alpha = 1 \), the penalty is composed completely of the L1 penalty and commonly known as Lasso regression. When \( \alpha = 0 \), the penalty is composed completely of the L2 penalty and is commonly known as Ridge regression. The advantage of the elastic net is that alpha can range between 0 and 1 and therefore inherits the properties of both L1 and L2 regularization. L1 regularization is commonly used for model selection because predictor coefficients can be shrunk to zero and effectively removed from the model. However, in the presence of correlated predictor variables, one predictor variable would be randomly selected while the others are removed. This
can result in decreased predictive performance since variables with some explanatory power are no longer in the model. With L2 regularization, regression coefficients will shrink towards zero but with an asymptote at zero. This is the preferred type of regularization in the presence of multicollinearity because all variables would remain in the model but with smaller coefficients. The elastic net provides a framework to choose the best compromise between the L1 and L2 penalties. For further details, we direct the reader to Zou and Hastie [2005b] and Hastie et al. [2009].

We do not directly treat spatial correlation in our models due to the large computational demands of fitting the covariance function for ~4000 pixels for 28 dates. Neglecting spatial correlation is another potential source for regression coefficients to be over fit and consequently we do not interpret them for physical meaning [Cressie, 1993; Erickson et al., 2005]. Nonetheless, we show utility with our methods without addressing spatial correlation and expect the results presented herein would improve if spatial correlation were explicitly treated [Carroll and Cressie, 1997].

3.5 Results

3.5.1 SWE Models: Discrete Time

Table 3.2 shows that the models that include fSCA as a predictor variable, i.e. FSCA and PHV-FSCA outperform models that do not include fSCA in all metrics in all years except %Bias (where all models exhibit close to 0 %Bias). The models that include fSCA as a predictor explain on average between 78% and 86% of the variance in SWE distribution in any given year whereas PHV only explains between 55% and 67%. We similarly see improvement in %MAE where models with fSCA as a predictor exhibit mean annual %MAE between 27% and 41% compare with PHV which yields mean annual %MAE between 50% and 71%. In summary, we note a substantial improvement in $r^2$ and %MAE for distributing SWE when fSCA is considered.
Table 3.2: Mean prediction performance for FSCA, PHV, and PHV-FSCA from each observation date using a split-sampling approach.

<table>
<thead>
<tr>
<th></th>
<th>PHV</th>
<th>FSCA</th>
<th>PHV-FSCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r^2 )</td>
<td>%MAE</td>
<td>%Bias</td>
</tr>
<tr>
<td>2013</td>
<td>0.55</td>
<td>71</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>0.67</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>2015</td>
<td>0.57</td>
<td>64</td>
<td>-3</td>
</tr>
<tr>
<td>2016</td>
<td>0.59</td>
<td>71</td>
<td>2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.6</td>
<td>61</td>
<td>0</td>
</tr>
</tbody>
</table>

3.5.2 SWE Models: Transferred in Time

Figure 3.2 shows March 23, 2014 as an example transfer date in 2014 that was predicted using a PHV-FSCA model from May 25, 2013. Overall, we see similar spatial trends between the observed and modeled SWE distributions, but we see darker purples in the observed map indicating higher SWE. The mean observed SWE is 0.23 m compared to 0.21 m modeled. The range of observed SWE is 0-0.75 m while the modeled SWE ranges from 0-0.39 m. The standard deviations of observed and modeled SWE are 0.13 m and 0.11 m, respectively. The difference map shows large areas of agreement to within 0.05 m SWE and the histogram shows the mean centered on zero. A qualitative comparison with Figure 3.1 suggests these are mainly forested areas. We see areas of under prediction (red) mostly above tree line in the north and areas of over prediction (blue) above tree line in the south. A comparison with Google Earth® aerial imagery confirms that the pixels that exhibit very large negative differences (bright red pixels) are areas with persistent snow for much of the year. The snow extent is very similar between the modeled SWE and observed SWE because only areas observed to have fSCA greater than 0 were predicted.

Figure 3.3 shows the range of \( r^2 \) for transferred models for FSCA, PHV, and PHV-FSCA on each date based on models created in other years. Additionally, we indicate the best model performance for each transfer date by a diamond. In this regard, we observe unanimous improvement across all dates with models that use fSCA as a predictor compared to the model that only uses topographic variables, i.e. PHV. PHV-FSCA yields highest mean best \( r^2 \) of 0.84 (mean of diamonds).
Figure 3.2: An example of SWE distribution from 2014-03-23. Observed SWE (top left); modeled SWE using PHV-FSCA (bottom left); the difference, modeled SWE – observed SWE, draped over a shaded relief map (bottom right).
followed by FSCA which exhibits a mean best $r^2$ of 0.81 and then PHV with a mean best $r^2$ of 0.6. We also observe a notable decrease in $r^2$ for PHV towards the end of the season while the models that contain fSCA as a predictor (FSCA and PHV-FSCA) exhibit a consistent $r^2$. The standard deviation of $r^2$ for FSCA is 0.07, PHV is 0.14, and PHV-FSCA is 0.05.

Figure 3.4 also clearly shows FSCA and PHV-FSCA to exhibit the best, i.e. lowest, %MAE with transferred models compared to the model that only used topographic variables, PHV. The mean best %MAE (mean of diamonds) for FSCA and PHV-FSCA is 33% each while the mean best %MAE for PHV is 63%. Particularly obvious in these panels is the upward distribution shift and larger range for PHV later in the season compared to only a minimal increase in %MAE for the models containing fSCA as a predictor. The standard deviations of the best %MAE are 10% for FSCA and PHV-FSCA and 26% for PHV.

Figure 3.5 shows that the best transferred models (diamonds) for both all models exhibit close to zero bias. The mean best %Bias for FSCA is 2%, for PHV is 2%, and for PHV-FSCA 1%. However, we note that the variability in %Bias increases more dramatically at the end of the season, especially for PHV, the model without fSCA as a predictor. The standard deviations of the best %Bias for FSCA, PHV, and PHV-FSCA are 12%, 15%, and 7%.

The unanimous improvement of models that include fSCA as a predictor motivates us to evaluate the performances of FSCA and PHV-FSCA more closely. We observe only minor differences in the $r^2$ and %MAE when we analyze each of the best predictions on each transfer date. In contrast, comparison of the best %Biases from each model on each transfer date shows that FSCA incurs higher annual mean and median %Biases in all years (not shown). Furthermore, the two largest errors in %Bias across all dates occurred with FSCA in 2015 on February 17 and April 9. These dates exhibited abnormal fSCA-SWE relationships after storms blanketed the basin and experienced little wind redistribution. As a single variable regression, FSCA is most sensitive to fSCA distribution so PHV-FSCA was more robust to these anomalous fSCA conditions. In this regard, the standard deviation of %Bias of the best predictions is 7% for PHV-FSCA compared to 12% for FSCA. Notwithstanding, we observe the relationships between SWE and physiography to
Figure 3.3: The range of $r^2$ for each simulation date from the transferred models. The diamond represents the best model for each transfer date. The boxplots represent the interquartile range with vertical lines to denote the 5th and 95th percentiles. Black dots are outliers.
Figure 3.4: The range of %MAE prediction errors for each transfer date from the transferred models. The diamond represents the best model for each transfer date. The y-axis is limited for clarity. The boxplots represent the interquartile range with vertical lines to denote the 5th and 95th percentiles. Black dots are outliers.
Figure 3.5: The range of %Bias prediction errors for each transfer date from the transferred models. The diamond represents the best model for each transfer date. The y-axis is limited for clarity. The boxplots represent the interquartile range with vertical lines to denote the 5th and 95th percentiles. Black dots are outliers.
Table 3.3: Mean model performance comparison with PHV-FSCA for the split sample model and the best transferred model for each simulation date.

<table>
<thead>
<tr>
<th>Year</th>
<th>$r^2$</th>
<th>%MAE</th>
<th>%Bias</th>
<th>$r^2$</th>
<th>%MAE</th>
<th>%Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>0.86</td>
<td>33</td>
<td>1</td>
<td>0.86</td>
<td>34</td>
<td>-1</td>
</tr>
<tr>
<td>2014</td>
<td>0.86</td>
<td>27</td>
<td>1</td>
<td>0.86</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>0.83</td>
<td>33</td>
<td>1</td>
<td>0.82</td>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td>2016</td>
<td>0.82</td>
<td>41</td>
<td>2</td>
<td>0.82</td>
<td>40</td>
<td>-1</td>
</tr>
<tr>
<td>Mean</td>
<td>0.85</td>
<td>32</td>
<td>1</td>
<td>0.85</td>
<td>33</td>
<td>1</td>
</tr>
</tbody>
</table>

break down towards the very end of the season while the relationship with fSCA does not (compare blue and purple for the last dates each year in Figures 3.3, 3.4, 3.5). Nonetheless, we suggest that this SWE-fSCA relationship is adequately captured within PHV-FSCA while still being robust to anomalous distributions of SWE and fSCA. We expect the use of satellite remote sensing of fSCA, which is of lower quality than that measured by ASO, to exacerbate this issue. As such, we proceed our analysis in the remainder of the study with PHV-FSCA.

We compare the best transferred models of PHV-FSCA presented in Figures 3.3, 3.4, 3.5 with the split-sample models of PHV-FSCA from the previous section, to assess the performance degradation one would expect due to transferring a model between years. We observe that PHV-FSCA models can be transferred to another year with little degradation in performance (Table 3.3). The yearly $r^2$ of the best transferred model are always within 1% of the split sample model and, on average, explains the same amount of variance in SWE distribution. The yearly mean %MAE of the transferred model is always within 6% of the split sample model with the mean 1% higher. The yearly mean magnitude of %Bias is actually lower, i.e. better, in two years with the best transferred model compared to the split sample model, and the overall mean is the same.

3.5.3 Evaluating the Selected Model Performance

Table 3.4 summarizes and compares the yearly mean statistics from PHV-FSCA for the best models and the selected models transferred from another year. The model selection procedure
Table 3.4: Yearly and overall prediction errors of PHV-FSCA for best transferred models and the selected models. Best transferred model errors involved fitting a model to all the data on each date and using these models to predict SWE on dates in other years. Only the best model date-simulation date pair is considered. The selected model errors are derived from the same ensemble of model date-simulation date pairs, but the model is selected based on the pillow SWE similarity described in the text.

<table>
<thead>
<tr>
<th>Year</th>
<th>Best Transferred Model</th>
<th>Selected Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^2$</td>
<td>%MAE</td>
</tr>
<tr>
<td>2013</td>
<td>0.86</td>
<td>34</td>
</tr>
<tr>
<td>2014</td>
<td>0.86</td>
<td>26</td>
</tr>
<tr>
<td>2015</td>
<td>0.82</td>
<td>39</td>
</tr>
<tr>
<td>2016</td>
<td>0.82</td>
<td>40</td>
</tr>
<tr>
<td>Mean</td>
<td><strong>0.85</strong></td>
<td><strong>33</strong></td>
</tr>
</tbody>
</table>

results in similar yearly $r^2$ but increases in both %MAE and %Bias. Yearly mean %MAE increases between 3% and 18% and yearly absolute %Bias increases between 3% and 37%. The years 2013, 2014, and 2015 yielded increases in %MAE of 18%, 17%, and 16%, respectively, while 2016 exhibited an increase of only 3%. The absolute %Bias increases 3% in 2013, 6% in 2014, 37% in 2015, and 24% in 2016.

Figure 3.6 shows the difference for each transfer date between the errors of the best models (diamonds in Figures 3.4, 3.5) and the errors of the selected models (these are shown as open circles in Figure 3.6). We focus on %MAE and %Bias from PHV-FSCA only because $r^2$ showed generally consistent performance for a given transfer date (Table 3.4; comparatively small vertical range of the purple boxplots in Figure 3.3 compared with Figures 3.4 and 3.5).

The mean difference in %MAE is 15% and the selected model was the same as the best model on only one date (Figure 3.6a). The range in performance difference is between 0% and 45% with a mean of 15% and median of 12%. The difference in errors were generally lowest in 2016 with a mean of 3% and standard deviation of 4%, respectively. In contrast, 2013, 2014, and 2015 exhibited both higher mean differences (18%, 16%, and 18%, respectively) and higher standard deviations (14%, 8%, and 14%, respectively).

The mean difference in the absolute magnitude of %Bias is 29% and the selected model was
Figure 3.6: The increase in prediction error for %MAE (a) and %Bias (b) between the best model and the selected model.
the same as the best model on only one date (Figure 3.6b). The range in performance difference is between 0% and 73% with a mean of 29% and median of 26%. The yearly mean difference in error was consistently higher for %Bias than %MAE, with means of 25% in 2013, 30% in 2014, 34% in 2015, and 20% in 2016. The standard deviation in the error difference was lowest in 2014 (9%) compared to 16% in 2013, 22% in 2015, and 11% in 2016.

We also tested regressions whereby SWE and fSCA measurements from the ASO flights were aggregated into 2-week and monthly datasets to test the sensitivity of transferring individual observations, i.e. to test if the methodology might be more robust to the wrong model selection. However, the performances of the selected models did not change significantly, hence the ability to estimate the distribution of SWE accurately at any given time will largely depend on how well the observed years exhibit the same general distributions. Therefore, we evaluate the prediction errors from different model years to see if there are any systematic differences in the predictions generated and thus how sensitive predictions for transfer dates will be given the existing ensemble of observations. In this context, we note distinct differences in the predictive ability of models from different years, which has important implications for the ability of a model to transfer.

Figure 3.7a shows that, on average, models from 2015 produced the lowest %MAE and models from 2016 produce predictions with the largest %MAE. The mean %MAE in 2016 is 118% compared to 79% with 2013 models, 69% with 2014 models and 50% with 2015 models. We observe consistency in a year’s ability to predict another year relative to the overall distribution, i.e. the colored dots are typically clustered within the range of the boxplots. We also note in 2016 that the best models always came from 2013, but in 2013 the best models only came from 2016 for the first three flights. In 2014, the bulk of the models around the median performance were from 2015 and vice-versa in 2015. In these two years, the poorest predictions were from 2013 and 2016.

Figure 3.7b show that models from 2016 also produce the largest %Bias with a mean of 97% compared with 56% from 2013, 21% from 2014, and -26% from 2015. We again see consistency in the location of the colored dots relative to the boxplots thus suggesting years will consistently model the SWE distribution of certain other years better. In this regard, similar to %MAE, 2013
produces the lowest %Bias in 2016, but in 2013 the inverse is only true for the first three flights as the distribution shifts up.

3.6 Discussion

3.6.1 SWE Distribution Modeling

The relationship between snow covered area and SWE is well established as evident by the common use of depletion curves in hydrologic modeling [Anderson, 1973; Liston, 1999; Luce et al., 1999; Clark et al., 2011]. We invert the idea of the depletion curve in this study by using the spatial distribution of fSCA to predict the spatial distribution of SWE. The basis for this approach is relatively well established given that fSCA is sensitive to topographic complexity and the spatial distribution of SWE [Donald et al., 1995; Niu and Yang, 2007; Fassnacht et al., 2016]. In this regard, the statistics for PHV-FSCA reported in this study compare favorably to SWE distribution statistics reported previously. Headwater catchment scale studies based on intensive field data have been able to achieve $r^2$ values from as low as 0.18 to 0.65 [Elder et al., 1991; 1998; Balk and Elder, 2000; Erxleben et al., 2002; Molotch and Bales, 2005; López-Moreno and Nogués-Bravo, 2006] compared to the mean $r^2$ of 0.83 for the selected model in this study. These papers, which cover only a few square kilometers, represent a far more simplistic problem with regard to characterizing relationships between snow accumulation and physiographic variables.

The results presented herein compare favorably with larger scale studies (i.e. > 1000 km$^2$) of snow distribution. Fassnacht et al. [2003] reported average yearly RMSE between 0.12 and 0.16 m and 0 m bias when cross-validated with snow pillows. Harshburger et al. [2010] reported an average $r^2$ of 0.82 and RMSE of 0.05 m when cross-validated with snow pillows. Schneider and Molotch [2016] reported a mean RMSE of 0.23 m and %Bias of 0.8% from snow surveys in the Upper Colorado River Basin. This is compared to a mean RMSE of 0.07 m and mean %Bias of 11% for the selected model in this study. We expect the errors reported in this study to be very robust because we compare against spatially explicit observations over 1000 km$^2$ compared with
Figure 3.7: The transferred prediction errors for PHV-FSCA for each date; %MAE (a) and %Bias (b). The boxplots represent the distribution of prediction errors for each date. The boxes represent the interquartile range with vertical bars for the 5th and 95th percentiles. Small black dots are statistical outliers. The colored dots are the prediction errors coded by model year and jittered to prevent over plotting. Open circles represent the selected model.
point observations in Fassnacht et al. [2003] and Harshburger et al. [2010] and 28 snow surveys covering ~100 km² in Schneider and Molotch [2016].

Bair et al. [2016] compared a retrospective SWE reconstruction to the same ASO observations (2013-2015) and reported yearly mean %MAE between 20% and 31% and yearly mean %Bias between -11% and 10%. These yearly statistics are lower than those reported in this study (Tables 3, 4), but are the result of a much more complicated energy balance model that can only be run after the snow has disappeared. The selected model in this study is a simple linear regression that can be applied in real-time, thus we consider our results valuable for applications where real-time estimates of SWE distribution are needed. Furthermore, we compare our selected model results to SWE estimates from the U.S. National Weather Service’s operational Snow Data Assimilation System (SNODAS). SNODAS produces spatially distributed SWE estimates for the coterminous United States at 1 km by assimilating a physically based model with SNOTEL observations and remotely sensed snow covered area. The SWE product from SNODAS is the only high-resolution, gridded SWE product available at a daily time step for the continental United States and is available from http://nsidc.org [Barrett, 2003]. Previous work has shown the physically based model to perform well at the point scale [Rutter et al., 2008] but suffer in alpine zones because it does not consider wind redistribution [Clow et al., 2012]. The yearly mean r² between ASO and SNODAS ranges from 0.04 in 2016 to 0.36 in 2015 with a mean of 0.17. The yearly mean %MAE ranges from 120% in 2014 to 274% in 2013 with a mean of 199%. The yearly mean %Bias ranges from -10% in 2016 to 236% in 2015. We refer the reader to “Selected Model” of Table 3.4 for the PHV-FSCA error summary. In this regard, SNODAS exhibits a mean %MAE 4 times that of PHV-FSCA and a mean %Bias 8 times higher than PHV-FSCA. Both models struggled to predict the anomalous conditions in 2015 but SNODAS struggled even more in 2013. Errors for SNODAS in 2013 were 30% and 74% higher than the multi-year mean SNODAS %MAE and %Bias, respectively. Errors for PHV-FSCA in 2015 were 18% and 364% higher than the multi-year mean PHV-FSCA %MAE and %Bias, respectively. While it is clear that the errors with PHV-FSCA are considerably lower than with SNODAS, PHV-FSCA struggled more in 2015 relative to its mean.
Further, SNODAS is a complex system that attempts to capture the snow dynamics across the entire United States compared with PHV-FSCA which was trained using a very specialized data set in the study region.

All dates improved when fSCA was used as a predictor, and, on average, the %MAE decreases from 61% with PHV to 32% with PHV-FSCA while the %Bias slightly increases from 0% to 1%, with PHV-FSCA. Further, even the model that only used fSCA as a predictor without any topographic variables yields improvements in $r^2$ and %MAE, but degrades with %Bias compared to the model that only includes physiographic variables (Table 3.2, Figures 3.3, 3.4, 3.5). The improvements seen with fSCA should not be surprising since they are not totally unprecedented. König and Sturm [1998] showed in situ measurements of SWE could be linked to relationships with topography and snow cover patterns from aerial photographs taken during the arctic melt season to estimate SWE distribution in unsampled locations. We suggest that repeated observations of fSCA for the same locations implicitly provide a measure of the relative magnitude of SWE by providing an estimate of the sub-grid variability of SWE. Marchand et al. [2005] suggested that the sub-grid standard deviation of physiography could improve regression models of SWE distribution because it accounts for sub-grid snow depth variability. Remotely sensed fSCA provides a means of capturing this sub-grid variability in SWE without requiring higher resolution data regarding the variance of physiography for each pixel.

We also show that SWE distributions can be related to fSCA and physiography in one year and applied to another year. The performance of PHV-FSCA was quite similar when applied to the date at which the model was trained (i.e. split sample) versus applying the model to other years (i.e. transferred models). In this regard, we see a minimal decrease in prediction skill and minimal increase in prediction error when we compare split sample models with the best transferred models. Recall that the split sample model was trained and tested on the same day and the transferred model was trained in a different year from which the prediction was applied. Table 3.3 shows the mean $r^2$ in the split sample model to be equivalent to that of the best transferred model. Moreover, the mean %MAE of the best transferred model exhibits only a 1% difference from the split sample
model and the mean %Bias exhibits no difference. Thus, if we are able to identify the best model we would see minimal degradation in predictive ability. However, we see significant differences in predictive ability from models of different years (Figure 3.7) and conclude that relationships between SWE and physiography are only similar between specific years, not as uniformly as put forth by previous studies [Erickson et al., 2005; Grünewald et al., 2013]. Also, it is unclear the impact climate non-stationarity will have on the ability to transfer models to future years. Even so, for each year in this dataset there exists a corresponding year from which accurate predictions can be made.

The benefits of using fSCA as a predictor variable in the transferred models are particularly large at the end of the season when the errors are highest (Figures 3.3, 3.4, 3.5). The average difference in %MAE between the best PHV model and best PHV-FSCA model for the last 2 dates of each year is 54% compared to a mean difference of 20% for the other dates. The improvements seen by including fSCA as a predictor are noteworthy because the ensemble of models trained using ASO observations could then be applied using remotely sensed fSCA from satellites. The degree to which satellite-based fSCA will improve model performance toward the end of the snowmelt season will be partially dictated by the accuracy of the fSCA data, which is subject to increasing uncertainty at low fSCA values [Painter et al., 2009; Rittger et al., 2013]. Optical fSCA products such as MODSCAG also suffer in forested areas since snow cover is occluded by the canopy [Raleigh et al., 2013; Rittger et al., 2013]. A viewable gap fraction correction is typically used to extrapolate fSCA to the occluded portions of a pixel by assuming that fSCA is the same under the canopy [Molotch and Margulis, 2008], and future studies should evaluate the sensitivity of PHV-FSCA to this assumption. Cloud cover can also obscure a satellite’s view of the snow and therefore making it difficult to estimate the snow extent. However, Slater et al. [2013] showed that gaps of 5 or more consecutive days are rare with MODIS thus suggesting that weekly estimates of SWE would be feasible. Lastly, omission and commission errors of cloud identification can provide erroneous estimates of fSCA although Parajka and Blöschl [2008] report an overall filtering accuracy of 96% in the Alps. Nonetheless, this is still an active area of research [Dozier et al., 2008; Parajka and
3.6.2 Considerations for Extending the ASO Record

The value of ASO during the year flown is significant for water management because it provides high resolution SWE information with low uncertainty compared to traditional estimation methods and therefore facilitates more confident water supply forecasts. The downside to ASO is the relatively high cost of operation. The work presented here provides a first step in realizing the value of ASO subsequent to active operations. We find that the number of flights within a year affects the mean and variance of the predictions in other years by <1%; this was quantified by iteratively selecting between 1 and the number of observations in a given year 100 times and assessing the change in error. In other words, it is better to perform ASO once per year for 10 years than 10 times in one year.

It is clear from our results that flights in one year do not necessarily transfer well to another year, e.g. 2016 to 2014 and 2015 (Figure 3.7). The model selection procedure relies on operational snow pillows to identify similar patterns of SWE between historical dates and the date of interest, but these may not always represent SWE of the surrounding terrain [Molotch and Bales, 2006; Rice et al., 2011; Meromy et al., 2012]. However, the minimal degradation in prediction performance when considering the best transferred model should motivate improvements for identifying the dates from the past with the most similar SWE distribution. As shown here, if these dates can be identified, the SWE can be modeled at accuracy levels that are equivalent to models generated from ASO data collected on the day of interest. We also tested the similarity in remotely sensed fSCA as a method for selecting the most similar distribution, but found anomalous SWE and fSCA distributions to make this method less robust. However, a completely remote sensing based approach would be useful in data sparse regions where ground stations do not exist. A potentially robust remote sensing based method could be to track fSCA through time and select a similar SWE distribution based on the trajectory of fSCA rather than a single snapshot of fSCA. It is also important to note that flying ASO during an anomalously dry year such as 2015 in California does not
necessarily reduce the predictive capacity of models for future years. In our case, 2015 provided useful estimates of SWE distribution in all the other years, including 2016, even though the models from the relatively wet year of 2016 did not transfer well to the very dry year of 2015 (Figure 3.7).

Intuitively, it is not surprising that the models from 2016 did not transfer to other years because there was much more snow than in 2014 and 2015, which were very low snow years. Less intuitive, however, is why the inverse worked well, i.e. the models from 2014 and 2015 did transfer relatively well to 2016. In this regard, Veitinger et al. [2014] showed that that deeper snow packs show greater redistribution and more smoothing. Accordingly, we see lower mean fSCA in 2016 for the same mean SWE in 2015. This means that for the same mean SWE, there are deeper pockets of snow covering a smaller area remaining at the end of a higher snow year than a lower snow year. Therefore, we would expect the relationships between topographic variables and SWE to become increasingly disparate from the underlying terrain with a deeper snow accumulation. Thus, the regression coefficients from the deeper snowpacks can no longer represent the shallow snowpacks even though the distribution of the deeper snowpack still inherently contains some signature of the shallow snowpack. However, this does not mean that the model from the last date of the season (which is now also a shallow snowpack) transfers well because this SWE distribution still largely represents the dominant patterns from the peak SWE distribution. Multiple studies have shown that each year’s specific relative peak SWE distribution controls the melt path and the relative depths throughout [Liston, 1999; Luce et al., 1999; Egli and Jonas, 2009]. Notwithstanding, Veitinger et al. [2014] studied snow depth at 1 m resolution and quantified the variability of the snow pack up to 25 m resolution so our results are not directly comparable. However, we posit that the inclusion of fSCA as a predictor variable is able to capture some of the variability in SWE below the 500 m model resolution, similar to including terrain roughness as done is previous studies [Lehning et al., 2011; Grünewald et al., 2013; Helbig et al., 2015].
3.7 Conclusion

We estimated the relationships between SWE, physiography, and fSCA and show that the temporal consistency in these relationships can be used to estimate SWE in years beyond the ASO observation record, with a mean $r^2$ of .85, mean %MAE of 33%, and mean %Bias of 1%. The relationships transfer robustly in time with no degradation in $r^2$ or %Bias and only 1% in %MAE when comparing predictions between models fit on the same day and models from a different year. Models with fSCA as a predictor transfer better than those without, and we suggest that the inclusion of fSCA provides information with regard to the variability of the SWE resulting from different accumulation dynamics due to differences in terrain roughness. In this regard, the availability of satellite images of fSCA facilitate the transfer of modeled relationships based on ASO observations to dates when no airborne snow depth measurements exist. The crux of this proposition is in selecting the model to transfer. We offer a method with which to select a model from another year based on the similarity in SWE distribution at existing snow pillows in the area. Comparison of the best predictions and the selected predictions results in a mild decrease in $r^2$ (0.02) and moderate increases in %MAE (15%) and %Bias (10%). The results presented above motivate further refinement in the technique used to select the best model because if these dates can be identified then SWE can be modeled at accuracy levels equivalent to models generated from ASO data collected on the day of interest. Lastly, although SWE distributions simulated in years with anomalous SWE distributions (2014, 2015) had the highest errors, models from these years still yielded good performance in 2013 and 2016. Thus, the benefit of ASO in anomalously dry years is two-fold: water managers receive accurate information during a year that is difficult to model, but also these observed SWE distributions can be used to simulate SWE distributions in future, less anomalous years. Overall, ASO provides an unprecedented observation of the relationships between SWE, fSCA, and physiography. The ASO dataset facilitates improved understanding of these relationships in both time and space and should lead to better information for water managers.
3.8 References


GDAL Development Team (2015), *GDAL - Geospatial Data Abstraction Library, Version 1.11.3*.


Kirnbauer, R., and G. Bloschl (1994), Wie ähnlich sind Ausaperungsmuster von Jahr zu Jahr (How similar are snow cover patterns from year to year)? *Deutsche Wässerkundliche Mitteilungen*, 37(5/6), 113–121.


Parsons, W. J., and G. H. Castle (1959), Aerial reconnaissance of mountain snow fields for maintaining up-to-date forecasts of snow melt runoff during the melt season, Western Snow Conference, Reno, Nevada.

Potts, H. L. (1937), Snow surveys and runoff forecasting from photographs, *Proceedings of the 5th Annual Western Interstate Snow Survey Conference*.


Revuelto, J., J. I. López-Moreno, C. AzorÃ¡n-Molina, and S. M. Vicente-Serrano (2014), Topo-


Chapter 4

Topographic Controls on Depletion Curves Observed from Airborne LiDAR Snow Depth Data

4.1 Introduction

The existing body of literature regarding depletion curves covers a variety of spatial scales from point-based derivations over hillslopes [Luce et al., 1999; Luce and Tarboton, 2004], physically-based gridded models applied to the watershed scale [Clark et al., 2011] and regional to global-scale land surface models [Niu and Yang, 2007; Swenson and Lawrence, 2012]. One of the challenges for these models is the impact of scale, i.e. the scale at which the heterogeneity of snow is consistent. In this regard, Clark et al. [2011] concluded that probability distributions of snow depth can be used to parameterize subgrid variability in snow distribution at the hillslope scale (~100 m). They note, however, that elevation gradients larger than 200 m would not be adequately represented by a single probability distribution. Essery and Pomeroy [2004] similarly point out that the utility of a single probability distribution could be extended by stratifying heterogeneous landscapes.

Multiple studies have related fractional SCA (fSCA) to the subgrid standard deviation of snow depth at peak accumulation [Essery and Pomeroy, 2004; Helbig et al., 2015]. Other works have related the standard deviation of snow depth to the variability of the underlying terrain [Marchand and Killingtveit, 2005; López- Moreno et al., 2014; Helbig et al., 2015]. Hence, the variability of subgrid snow depth will vary in space. However, Lopez et al. [2014] observed a consistent spatial pattern of the coefficient of variation (CV) of snow depth across multiple LiDAR acquisitions in time. Current parameterizations of subgrid depletion curves based on probability distributions
rely on an assumed subgrid CV of SWE. Estimates of the CV of SWE and snow depth for different environments have been provided in the literature based on field studies [Liston, 2004; Clark et al., 2011]. However, these are broad classifications that may not hold true across a basin of mixed environments, even when modeled at the hillslope scale.

Multiple studies have investigated subgrid depletion curves based on terrain characteristics [Donald et al., 1995; Niu and Yang, 2007; Swenson and Lawrence, 2012; Helbig et al., 2015; Fassnacht et al., 2016]. In particular, Niu and Yang [2007] observed that snow cover depletion patterns within 1°x1° grid cells varied substantially between cells with differing sub-grid variances in elevation. Accordingly, they provided a depletion curve parameterization that they recommended be tuned based on scale and an observed SCA time series. Swenson and Lawrence [2012] parameterized a similar depletion curve with an empirical factor they suggested could be partially explained by subgrid elevation variability, but that the remaining discrepancy between the observed and modeled depletion curves was the result of other variables affecting the shape of the depletion curve. In this regard, Fassnacht et al. [2016] evaluated the use of topographic variables for modeling the threshold of point-measured SWE for which SCA is at a maximum. They found that the slope of SWE vs SCA was similar each year even though the threshold SWE was predominantly controlled by peak SWE over an eight-year period. These studies suggest that the shape of depletion curves is largely controlled by the terrain that affects both the snow accumulation and ablation processes.

The previous studies provide insight into the relationships between snow depth and SCA. Notwithstanding, these previous studies have been limited to theory or to poorly constrained measurements of the snowpack. Lacking has been studies that utilize spatially explicit measurements of snow depth in the context of depletion curve characteristics. Hence, we still lack understanding of the spatial variability in depletion curve behavior and how individual terrain elements at scales less than 100 m affect this behavior. The U.S. National Aeronautics and Space Administration (NASA), Jet Propulsion Laboratory (JPL), Airborne Snow Observatory (ASO) dataset offers a new opportunity to observe and investigate spatial differences in depletion curves that result from topographic influences on snow accumulation and ablation processes. The ASO dataset provides high spatial
resolution time series of Light Detection And Ranging (LiDAR)-derived snow depth and hyperspectral measurements of snow extent over multiple years from which relationships between SCA and snow depth can be inferred. Our objective is to improve understanding of the physiographic controls of depletion curve shape based on subpixel, pixel, and inter-pixel terrain characteristics. Specifically, we ask: What are the primary physiographic controls on the characteristics of snow depletion curves?

4.2 Methods

4.2.1 Site Description

This study was conducted in the Tuolumne River basin in the Sierra Nevada mountains in California, USA (Figure 4.1). The basin is 1,175 km$^2$ consisting of 48% vegetation, 50% rock, and 2% water (with $<<1$% permanent snow/ice). The elevation range is 1127 m to 3965 m, encompassing 4 distinct ecological zones ranging from lower montane forest to alpine. The lower montane forest ranges from 1127m to 1800 m elevation and consists of a diverse mix of coniferous and deciduous trees. The upper montane forest ranges from 1800 m to 2450 m elevation and primarily consists of coniferous species of red fir and lodgepole pine. Elevations from 2450 m to 2900 m elevation are considered subalpine and consist of a mix of meadows and coniferous forest. The highest elevation band above 2900 m is an above-timberline alpine zone that contains limited herbaceous vegetation. It consists of large granitic features, talus slopes and boulder fields [NPS, 2016]. Snowmelt from the watershed runs off into the Hetch Hetchy reservoir in the western corner of the basin.

4.2.2 Data Sources

We use the NASA-JPL, ASO dataset in the Tuolumne River Basin in the Sierra Nevada mountains of California. The ASO mission consists of LiDAR in conjunction with a hyperspectral spectroradiometer [Painter et al., 2016]. The dataset consists of a 3m pixel snow-free digital eleva-
Figure 4.1: Landcover and location map of Tuolumne basin, California, USA.
tion model (DEM) and 3-m snow depth digital surface models (DSM). The data are distributed in the UTM zone 11 and WGS84 datum map projection system. The ASO flies approximately weekly from near peak SWE through the snowmelt season. There were six flights in 2013, ten in 2014, eight in 2015, and eight in 2016. Painter et al. [2016] report a mean absolute vertical error of 8 cm and a bias of 1 cm when compared with manually measured snow depths at the 15 × 15 m scale. Further details about the mission and processing can be found in Painter et al. [2016].

4.2.3 Depletion Curves

To explore sub-grid variability in depletion curves, the 3m resolution gridded snow depth data are aggregated to two new 501 m (nominally 500 m) resolution datasets. The first is the mean snow depth for each 500 m pixel and the second is the standard deviation of snow depth for each 500 m pixel. Similarly, the 3 m gridded snow depth data are aggregated based on a binary snow-free/snow-on schema to derive 500 m fractional snow covered area (fSCA) maps. Pixels at the edge of the domain were not included if the 500 m pixel extended past the boundary of the basin and pixels containing water bodies were also removed. We chose 500 m because it is the scale at which daily satellite observations of snow covered area exist from the MODerate-resolution Imaging Spectroradiometer (MODIS) and hence our results are theoretically transferable to applications using MODIS data. Furthermore, the 500-m scale is applicable for large scale hydrologic modeling which rely on depletion curves to accurately compute the energy balance. We consider the snow-free classification to be robust because the original 3m gridded snow depth data were masked for snow free areas based on spectral data from the spectrometer. ASO SCA has also compared well against SCA derived from 0.3m resolution images from World View [Bair et al., 2016].

We characterize the depletion curve for each 500 m pixel with a bilinear regression between snow depth (dependent variable) and fSCA (independent variable) through time using the 33 flights available from ASO. Theoretical considerations would motivate a curvilinear fit because fSCA can, by definition, only increase to a value of one whereas snow depth can increase to infinity [Liston,
However, we observed a distinctly linear relationship between snow depth and fSCA for low fSCA and snow depth values in each pixel, which is consistent with previous studies [Swenson and Lawrence, 2012]. Further, the objective of this work is not to parameterize a universal depletion curve but rather to evaluate the topographic controls on snow depletion dynamics. Hence, we quantify the depletion curve for each pixel based on a segmented bilinear regression. This results in two slopes and a breakpoint between the two linear segments.

We tested the statistical relationship for a significant breakpoint (p<0.1) in the bilinear regression based on a Davies’ Test [Muggeo, 2003; 2008]. The x-coordinate of the breakpoint represents the fSCA at which the slope of the depletion curve changes. The y-coordinate of the breakpoint represents the snow depth at which the slope of the depletion curve changes. Pixels without a significant breakpoint were fit with an ordinary least squares linear regression and the slope tested for significance. Pixels with significant slopes from the linear regression were included in our analyses under the assumption that this slope reflects the same process as the first slope of the bilinear regression (example: Figure 4.2e). Consequently, we examined the topographic controls based on the first slope of the bilinear regression or the slope of the ordinary linear regression and denote this the ”depletion slope”.

4.2.4 Topographic Controls on Depletion Curves

We analyze 2979 pixels (500-m resolution) for the effect of topographic controls on the depletion slope. Focusing on the depletion slope is particularly revealing with regard to process. From first principles of snow distribution, areas with steeper depletion slopes are indicative of relatively heterogeneous snow depths; i.e. an incremental decrease in depth results in a relatively small decrease in fSCA. In this regard, reductions of fSCA in areas with heterogeneous snow depth, in which snow is accumulated into large snow drifts, will occur relatively slowly even when snowmelt rates and depth reductions are relatively high. Conversely, shallow depletion slopes are indicative of relatively uniform snow depth because an incremental decrease in snow depth results in a large decrease in fSCA. It should be noted that the depletion slope is directly defined by the x and y
coordinates of the breakpoint of the bilinear regression since the y-intercept of the regression is assumed to be zero. Significant differences from zero (p<0.05) are rare and we consider this to be a function of uncertainty in the measurements. The upper slope of the bilinear regression is not explicitly addressed in this study since it was found to be uncorrelated with topography.

We use recursive partitioning, commonly known as a regression tree, to identify the physiographic controls (independent variables) on the slope of the depletion curve (dependent variable). We consider a range of independent variables that represent a variety of topographic controls on snow depth under the auspice that the snow depth-fSCA relationship is controlled both by the magnitude and variability of snow depth. We investigate three different scales of control including pixel-scale physiography, inter-pixel effects, and sub-pixel scale variability within each pixel [Table 4.1].

First, the pixel-scale physiographic variables represent the controls on snow depth associated with the pixel elevation, slope, aspect, and vegetation cover at 500 m. Second, the inter-pixel topographic variable represents the topographic variability within a 3x3 window around each 500 m pixel. Third, variables representing sub-pixel scale variability describe terrain roughness based on elevation, slope, and curvature within each 500 m pixel.

We evaluated two scales of sub-pixel variability at 33 m and 3 m. The first scale, quantified by the standard deviation of elevation (stdelev), the standard deviation of slope (stdslope), and the standard deviation of maximum curvature (stdmaxcurv), are calculated from a 33 m DEM because it factors evenly into the 501 m grid and 30 m is the finest DEM available worldwide (e.g. NASA ASTER and SRTM). The second scale explored represents the ratio of the typical height of a terrain feature to the typical width of a terrain feature; hereafter we refer to this ratio as "Δh/Δw" [Helbig et al., 2015]. "Δh/Δw" is proportional to the ratio of the standard deviation of elevation to the correlation length of the DEM and larger ratio values indicate a greater relative protrusion of the terrain. We calculated "Δh/Δw" for each 501 m pixel from the 3 m ASO DEM using equation 1 from Helbig et al. [2015]:
Table 4.1: Variables used in a recursive partitioning regression to explain depletion curve slopes. Variables were chosen based on their existing use in the literature but the sources listed are by no means exhaustive.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Resolution</th>
<th>Derivation Specifics/Units</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pixel-Scale Physiography</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zness</td>
<td>501 m</td>
<td>sin(slope); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>northness</td>
<td>501 m</td>
<td>cos(aspect); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>eastness</td>
<td>501 m</td>
<td>sin(aspect); ranges 0-1; dimensionless</td>
<td>Balk and Elder (2000); Erxleben et al. (2002); Fassnacht et al. (2003)</td>
</tr>
<tr>
<td>vegheight</td>
<td>501 m</td>
<td>vegetation height measured by ASO; used in place of forest canopy density from previous studies; meters</td>
<td>Molotch and Bales (2005, 2006); Painter et al. (2016)</td>
</tr>
<tr>
<td><strong>Interpixel Variability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tpi</td>
<td>501 m from a 3x3 window around each</td>
<td>topographic position index; elevation difference of a pixel from the mean of the surrounding pixels; meters</td>
<td>Revuelto et al. (2014); GDAL (2015)</td>
</tr>
<tr>
<td><strong>Subpixel Variability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stdelev</td>
<td>501 m from 33 m</td>
<td>standard deviation of elevation from 33m DEM; meters</td>
<td>Marchand and Killingveit (2005)</td>
</tr>
<tr>
<td>stdslope</td>
<td>501 m from 33 m</td>
<td>standard deviation of slope from 33m DEM; shown to detect changes in slope at multiple scales; radians</td>
<td>Fassnacht et al. (2016); Grohman et al. (2007)</td>
</tr>
<tr>
<td>stdmaxcurv</td>
<td>501 m from 33 m</td>
<td>standard deviation of max curvature from 33m DEM; max curvature from SAGA GIS; dimensionless</td>
<td>Marchand and Killingveit (2005); Lopez et al. (2014); Conrad et al. (2014)</td>
</tr>
<tr>
<td>Δh/Δw</td>
<td>501 m from 3 m</td>
<td>terrain variability metric defined in Eqn 1 in Helbig et al. (2015); computed from the 3m DEM;</td>
<td>Helbig et al. (2015)</td>
</tr>
</tbody>
</table>
\[
\frac{\Delta h}{\Delta w} = \left\{ \frac{\left[ (\partial_x z)^2 + (\partial_y z)^2 \right]}{2} \right\}^{1/2}
\]  

(4.1)

where \( \partial_x z \) and \( \partial_y z \) are the first partial derivatives of elevation, i.e. orthogonal slope components. The slope components were computed using the r.slope.aspect tool in the Geographic Resources Analysis Support System (GRASS) v6.4 [Neteler et al., 2012].

Regression trees are commonly found in the snow hydrology literature although typically for estimating the spatial distribution of snow [e.g. Elder et al., 1998; Balk and Elder, 2000; Molotch et al., 2005]. A regression tree allows us to model how the non-linear interactions between topography and snow depth distribution affect the rate of change between snow depth and fSCA. In this regard, we use the rpart package [Therneau et al., 2015] in R [R Core Team, 2015] to grow trees until nodes no longer improve the cross-validated \( r^2 \) by 0.001. We subsequently pruned the trees by selecting the number of splits corresponding to the minimum mean squared error, which is related to deviance.

### 4.3 Results and Discussion

#### 4.3.1 Depletion Curves

Figure 4.2 shows example depletion curves and highlights that the relationship between snow depth and fSCA is consistent in all four years, i.e. points from different years plot along the same line. The depletion slope in Figures 4.2a-d is the segment from the bottom left to the breakpoint in the red line, and Figure 4.2e shows an example of a pixel where no significant breakpoint is observed. Most of the snow depth values in Figure 4.2e are greater than 0.5 m, suggesting that this pixel experiences a snowpack that persists beyond the last seasonal ASO flights. We hypothesize that this pixel did not melt out significantly within the yearly observation windows due to very deep snowpacks, possibly due to preferential wind redistribution. We hypothesize that more observations later in the year as the snow depth decreases would result in a significant breakpoint in the depletion curve.
For pixels with a significant breakpoint, we typically observe that the depletion slope is less than the slope of the second segment. As such, if we start from peak snow depth, vertical melting of the snowpack will result in small changes of fSCA because the underlying terrain remains mostly covered until the breakpoint is reached. At this point, the slope decreases and we see larger decreases in fSCA per decrease in snow depth. The differences in the slope of the later stage of melt out, i.e. the depletion slope, is informative with regard to process. Given the conceptual model described above, we suggest that steeper slopes result from more heterogeneous snowpacks that have deeper, narrower piles of snow. As such, for an equivalent decrease in snow depth, the decline in fSCA for pixels with these deep pockets of snow is relatively low relative to pixels with shallower, more uniform snow distributions.

The range of the depletion slopes is 0.22 to 4.26 with a mean of 0.77 (Figure 4.3a). The spatial pattern of pixels that had statistically significant depletion slopes conform to higher elevation pixels (mean 2844 m with a range of 1234-3639m). At lower elevations in the western part of the basin, shallow snow depths and a lack of clear depletion dynamics rendered the regressions statistically insignificant (grey, Figure 4.3a); this may have occurred because of the drought conditions during this study period. The highest depletion slope values (yellow, Figure 4.3a) occur on the western edge of the southern tip and in the north of the basin. The lowest values (purple, Figure 4.3a) are clustered in the southeastern portion of the basin.

Figure 4.3b shows that the temporally averaged standard deviations of snow depth for each pixel exhibits a similar spatial pattern as the depletion slope (Figure 4.3a); the average standard deviations were computed based on the standard deviations of the 3 m snow depth observations within each 501 m pixel on each date.

Figure 4.3c shows that a strong, positive relationship between depletion slopes and snow depth variability exists across the basin($r^2 = 0.6; p<0$). This supports our interpretation of the depletion slope put forth in the previous section that pixels with steeper depletion slopes have more heterogeneous snowpacks.

The spatial patterns seen in Figure 4.3a suggest a physiographic control on the depletion
Figure 4.2: Example depletion curves from the Tuolumne for intervals of the depletion slope. All depletion slopes are statistically significant (p<0.05).
Figure 4.3: The depletion slope values and temporally average standard deviations of snow depth. a) Colors represent the depletion slope and b) colors represent the temporally averaged standard deviation of snow depth. The value of each 501 m pixel was computed based on the standard deviations of the 3 m snow depth observations within each 501 m pixel on each date. Subfigure b) is the average standard deviation of snow depth for each pixel. The values in subfigures a) and b) are divided into 20% quantiles so that each color represents the same number of observations. c) The relationship between the depletion slopes shown in a) and the average standard deviations of snow depth shown in b). The best fit linear relationship is shown in blue and the $r^2$ and p-value of the line are noted.
Table 4.2: The single variable correlations between each of the topographic variables and the depletion slope. p-values are shown. The correlation test was performed using a two-sided Kendall’s Rank Correlation test.

<table>
<thead>
<tr>
<th>Topographic Variable</th>
<th>Kendall Rank Correlation</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>stdmaxcurv</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>stdslope</td>
<td>0.26</td>
<td>0.00</td>
</tr>
<tr>
<td>∆h/∆w</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>vegheight</td>
<td>-0.24</td>
<td>0.00</td>
</tr>
<tr>
<td>elev</td>
<td>0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>stddem</td>
<td>0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>northness</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>tpi</td>
<td>0.12</td>
<td>0.00</td>
</tr>
<tr>
<td>zness</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>eastness</td>
<td>0.02</td>
<td>0.16</td>
</tr>
</tbody>
</table>

curve. However, Table 4.2 shows that none of the topographic variables have a strong individual relationship with the depletion slope. The highest magnitude correlation with depletion slope is displayed with the standard deviation of maximum terrain curvature \(r=0.27\) followed by standard deviation of terrain slope \(r=0.26\) and the \(\Delta h/\Delta w\) parameter \(r=0.25\). The prevalence of three different sub-pixel topographic variables indicates that sub-pixel terrain variability exerts substantial control on the depletion slope. All correlations are statistically significant except that for eastness. In this regard, we expect non-linear interactions between the topographic variables to dominate the results from the binary regression tree.

4.3.2  Topographic Controls on Depletion Slopes

The recursive partition regression tree analysis explains 29\% (95\% CI: +/- 0.2\%) of the variance in the depletion slope (Figure 4.4). Vegetation height exerts the main control on the depletion slope with taller vegetation causing smaller depletion slopes and smaller vegetation causing larger depletion slopes. This implies that taller vegetation leads to more homogeneous snowpacks and shorter vegetation leads to more heterogeneous snowpacks, which agrees with previous works that
have shown that areas with taller vegetation have more homogenous snow distribution due to reduced wind redistribution [Hiemstra et al., 2002; López-Moreno and Latron, 2008]. We interpret this split in the regression tree to be indicative of the difference in depletion dynamics above and below timberline. We note that northness, elev and multiple sub-pixel variables are used for subsequent splits both above and below tree-line. The repeated appearance of the same topographic variables in different places of the regression tree indicates non-linear interactions between variables are affecting the depletion slope.

We observe increased terrain roughness at sub-pixel scales (stdmaxcurv and ”∆h/∆w”) results in larger depletion slopes at multiple nodes in the regression tree. Lower sub-pixel terrain variability is associated with lower depletion slope values, consistent with expectations of relatively homogenous snow distribution. The observation that relatively homogenous terrain and snow distribution results in lower depletion slopes is intuitive given that these areas have shallower terrain hollows in which to trap snow. This implies piles of snow that are less deep and therefore SCA will decrease more quickly as snow depth decreases, resulting in a relatively shallow depletion slope.

Additionally, we observe northness to occur three times in the regression tree and in each case larger northness values lead to increased depletion slopes. This means that more north-facing pixels exhibit more heterogeneous snow distributions and may be a result of preferential snow accumulation on northeast-facing slopes due to a dominant wind direction from the southwest [Erxleben et al., 2002; Molotch et al., 2005].

Lastly, we note lower elevations are associated with smaller depletion slopes and therefore a more homogeneous snowpack. The physical process linking the depletion slope and elevation is not immediately apparent but one possibility is increased snow heterogeneity at higher elevations that comes from increased wind speeds and lower snowfall density that is more easily redistributed by wind compared to snow at lower elevations.
Figure 4.4: The regression tree for the controls on the depletion slope.
4.3.3 Implications

We observe that the relationship between snow depth and fSCA is consistent across four years during the depletion period and therefore the implications of this study are that remotely sensed fSCA contains a large amount of information regarding snow distribution. Repeat LiDaR scans of snow depth distribution are critical to developing fSCA-snow depth relationships and furthering our ability to estimate snow distribution in unsampled basins using available data such as fSCA and a DEM. Moreover, we suggest improvements in the depletion curve parameterization in hydrologic models could be obtained by further analysis of the relationship between fSCA and topography. In this regard, incorporating physically representative empirical factors that address accumulation processes at multiple scales would result in improved spatial representation of snow free and snow covered areas thus improving energy balance calculations.

The findings presented in this paper are, in general, conceptually consistent with the previously outlined global classification system for the CV of SWE for parameterizing probability distribution depletion curves based on topographic variability, wind speed, and air temperature [Sturm et al., 1995; Liston, 2004]. Using a new high resolution spatio-temporal dataset we similarly found more heterogeneous snowpacks with greater topographic variability. Further, increased vegetation height and decreased elevation are typically associated with decreased wind speeds and generally yielded less heterogeneous snowpacks. Temperature may have affected redistribution because colder, lower density snow at higher elevations is easier to redistribute, but the temperature in our study area is relatively uniform compared to global differences. However, we see a significantly larger range in CV of snow depth than previously reported with the mean observed subgrid CV (3m snowdepth within each 501 m pixel) ranging from 2.4 to 33.8 across all dates. This is compared with field values of CV for mountainous regions of 0.1-3 [Clark et al., 2011]. We consider this inconsistency to be a function of scale. The 3m snow depth dataset from ASO samples a much wider range of terrain and snow depths than the studies reviewed in Clark et al. [2011].
Notwithstanding, the results of our regression tree analyses are specific to the Tuolumne basin for two very dry years and two dry to average years, which may not transfer to other basins in other climates and at other latitudes. A similar analysis of repeated LiDaR measurements of snow depth distribution from multiple basins would strengthen the interpretation of our results by showing which results generalize to all basins as opposed to those specific to the Tuolumne Basin. The regression tree analysis misses the extremes of the range of depletion slope values (comparing the values in Figure 4.4 to Figure 4.3a, c), which is not totally surprising given an $r^2$ of 0.29. There is a lot of variability in the observed depletion slopes (Figure 4.3a) that may have resulted from local conditions that our predictor variables could not capture. This suggests that there were either not enough pixels with very high depletion slopes for a meaningful statistical relationship at the extremes of the observed parameters or a key variable is missing from our analysis. Further, pixels without a breakpoint (e.g. Figure 4.2e) exhibit slopes that are steeper than if we had more data from which we could calculate significant breakpoints. We suggest this could lead to an improved characterization of the relationship between depletion slope and topography.

The work presented here elucidates the controls on depletion curves which are largely affected by the distribution of snow depth. Hence, further analysis of remotely sensed fSCA may improve our understanding about the processes controlling snow distribution. In particular, the x and y components of the depletion slope (fSCA and snow depth values, respectively) may be analogous to the fractal dimension reported in previous studies that investigated the scale break at which snow distribution and depletion processes change [Deems et al., 2006; 2008; Trujillo et al., 2009; Schirmer and Lehning, 2011]. We suggest a detailed analysis of the x and y components could be fruitful in the context of modeling where depletion curves with different shapes could be parameterized with respect to topography.

We also performed the same analysis with SWE and did not find a consistent bilinear relationship. However, we consider the snow depth analysis to be more robust since it is rooted in direct observations from LiDAR, whereas SWE also consists of a modeled density with greater uncertainty. Notwithstanding, we might expect differences in the shapes of depletion curves derived
from SWE rather than snow depth because density increases in the ablation season, thus offsetting decreases in snow depth. Hence, we leave this analysis for future work.

4.4 Conclusions

We derived depletion curves from a high resolution spatio-temporal dataset of observed snow depth over the 1,175 km$^2$ Tuolumne River Basin and analyzed the physiographic controls on these depletion curves. We found that relationships between snow depth and fSCA (i.e. depletion slope) were robust over the four years of study. These depletion slopes exhibited significant variability across the watershed, ranging from 0.22 to 4.26 and averaging 0.77. In this context, we show that a strong, positive relationship between depletion slopes and snow depth variability exists across the basin ($r^2 = 0.6; \ p<0$). We also show that sub-pixel and pixel scale terrain variables explain 29% of the spatial variability in the depletion slope. In particular, increased vegetation height and decreased sub-pixel terrain variability were associated with more homogeneous snowpacks and lower depletion slopes. These results illustrate that repeat LiDAR-based snow measurements can improve process-level understanding of snow distribution. Such understanding has important implications for developing parameterizations of snow cover depletion curves across physiographic gradients. Given that parameterizations of snow cover depletion underpin sub-grid representation of energy and water fluxes across a range of earth system models, the approach presented herein has potentially broad applicability.

4.5 References


Essery, R., and J. Pomeroy (2004), Implications of spatial distributions of snow mass and melt rate


GDAL Development Team (2015), *GDAL - Geospatial Data Abstraction Library, Version 1.11.3*.


Kirnbauer, R., and G. Bloschl (1994), Wie ähnlich sind Ausaperungsmuster von Jahr zu Jahr (How similar are snow cover patterns from year to year)? *Deutsche Wässerkundliche Mitteilungen*, 37(5/6), 113–121.


Muggeo, V. M. R. (2008), segmented: an R package to fit regression models with broken-line relationships,


R Core Team (2015), *R: A Language and Environment for Statistical Computing*, Vienna, Austria.


Res., 117(D21), n/a–n/a, doi:10.1029/2012JD018178.

Therneau, T., B. Atkinson, and B. Ripley (2015), Recursive Partitioning and Regression Trees. version 4.1-10,

Chapter 5

References


Egli, L., and T. Jonas (2009), Hysteretic dynamics of seasonal snow depth distribution in the Swiss


Fassnacht, S. R. (2004), Streamflow Predictability in the Upper Versus Lower Colorado River Sub-Basins, Portland, Maine, USA.


GDAL Development Team (2015), *GDAL - Geospatial Data Abstraction Library, Version 1.11.3*.


Harpold, A. A. et al. (2014), LiDAR-derived snowpack data sets from mixed conifer forests across


Kimbauer, R., and G. Bloschl (1994), Wie ähnlich sind Ausaperungsmuster von Jahr zu Jahr (How similar are snow cover patterns from year to year)? *Deutsche Wässerkundliche Mitteilungen*, 37(5/6), 113–121.


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


Marshall, H.-P., and G. Koh (2008), FMCW radars for snow research, *Cold Regions Science and
Technology, 52(2), 118–131.

Martinec, J. (1975), Snowmelt-runoff model for stream flow forecasts, Nordic Hydrology.


Muggeo, V. M. R. (2008), segmented: an R package to fit regression models with broken-line relationships,


Nolan, M., C. F. Larsen, and M. Sturm (2015), Mapping snow-depth from manned-aircraft on landscape scales at centimeter resolution using Structure-from-Motion photogrammetry,*The
NPS (2016), Plants - Yosemite National Park, Available from: 
https://www.nps.gov/yose/learn/nature/plants.htm (Accessed 5 October 2016)

Pagano, T., and D. Garen (2005), A recent increase in western US streamflow variability and per-
sistence, J. Hydrometeor, 6(2), 173–179.

Colorado River runoff to dust radiative forcing in snow, Proceedings of the National Academy 
of Sciences of the United States of America, 107(40), 17125–17130,

Painter, T. H., K. E. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier (2009), Re-
trieval of subpixel snow covered area, grain size, and albedo from MODIS, Remote Sensing 

Painter, T. H. et al. (2016), The Airborne Snow Observatory: Fusion of scanning lidar, imaging 
spectrometer, and physically-based modeling for mapping snow water equivalent and snow 

Parajka, J., and G. Bloeschl (2008), Spatio-temporal combination of MODIS images - potential for 

Parsons, W. J., and G. H. Castle (1959), Aerial reconnaissance of mountain snow fields for main-
taining up-to-date forecasts of snow melt runoff during the melt season, Western Snow Con-
ference, Reno, Nevada.

Potts, H. L. (1937), Snow surveys and runoff forecasting from photographs, Proceedings of the 5th 
Annual Western Interstate Snow Survey Conference.

Potts, H. L. (1944), A photographic snow survey method of forecasting runoff, Proceedings of the 
12th Annual Western Interstate Snow Survey Conference.
R Core Team (2015), *R: A Language and Environment for Statistical Computing*, Vienna, Austria.


Rutter, N., D. Cline, and L. Li (2008), Evaluation of the NOHRSC snow model (NSM) in a one-


Sturm, M., J. Holmgren, and G. E. Liston (1995), A Seasonal Snow Cover Classification System


Therneau, T., B. Atkinson, and B. Ripley (2015), Recursive Partitioning and Regression Trees. version 4.1-10,


Wickham, H. (2007), Reshaping Data with the reshape Package, *Journal of Statistical Software*, 


