

Spatially Modeling Snowmelt Using Intra-Annual Patterns with Remote Sensing
in the Upper Snake River Basin

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science

with a

Major in Biological Engineering

in the

College of Graduate Studies

University of Idaho

by

Craig D. Woodruff

Major Professor: Russell J. Qualls, Ph.D., P.E.

Committee Members: Dev Shrestha, Ph.D.; Karen Humes Ph.D.

Department Administrator: Ching-An Peng, Ph.D.

December 2018

Authorization to Submit Thesis

This thesis of Craig D. Woodruff, submitted for the degree of Master of Science with a major in Biological Engineering and titled, "Spatially Modeling Snowmelt Using Intra-Annual Patterns with Remote Sensing in the Upper Snake River Basin," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: _____ Date: _____
Russell J. Qualls, Ph.D., P.E.

Committee Members: _____ Date: _____
Dev Shrestha, Ph.D.

_____ Date: _____
Karen Humes, Ph.D.

Department

Administrator: _____ Date: _____
Ching-An Peng, Ph.D.

Abstract

Snowmelt represents an important component of the hydrologic cycle with implications for water managers. In this research, we assess the applicability of intra-annual patterns for use in spatially modeling snowmelt. This is done through remote sensing which offers daily imagery of the watershed for the time period between 2000 and 2018. This thesis assesses the applicability of intra-annual patterns and develops two spatial models describing snow extent changes throughout a melting season.

In the first chapter we identify whether a pattern exists using spatial statistics and spatial comparisons. This fundamental step is the foundation for developing a model. We then address a critical assumption in the second chapter. Modeling the snowmelt progression is often done with depletion curves. We use narrowed error bounds to fit a curve, which solidifies our choice of curve to represent melt. This curve can be used to represent any snowmelt season. It is called a dimensionless curve and can be applied to any year.

The heart of this research is developing a spatial model which combines years of spatial data to describe the melting patterns. This is accomplished in chapter 3 with a principal component analysis. The model developed uses information from 2000 through 2016. It is applied to the snowmelt seasons of 2017 and 2018 to verify the spatial accuracy.

The final chapter creates another spatial model using a fundamentally different statistical approach. Not only are the results very similar using these two methods, but both can be accomplished quickly. This opens the door to further research into the widespread use of this type of model. Applications of both spatial models are real-time modeling, climate change modeling, cloud removal, and producing spatially distributed information in ungauged watersheds.

Acknowledgments

First and foremost I would like to thank Dr. Russell Qualls for his guidance and the opportunity he has given me. He has helped me develop professionally and academically. I cannot thank him enough for the support and care he has shown throughout my degree. This research was supported in part by USDA-NIFA Hatch project No. IDA01584. It has truly been a blessing to work with someone as excited, intelligent, and thoughtful as Dr. Qualls. I would also like to thank my committee members Dr. Shrestha and Dr. Humes. They have given me valuable guidance in writing my thesis, coursework, and directed me to helpful resources. Many of the chapters in this thesis are based on skills I acquired in coursework suggested by the committee members. Lastly, I would like to thank the Department of Biological Engineering for guidance throughout my time and the supportive atmosphere conducive to successful completion of this degree.

Table of Contents

Authorization to Submit Thesis	ii
Abstract	iii
Acknowledgments	iv
List of Figures	vii
List of Tables	ix
List of Equations	x
Chapter 1: Exploring the Existence of Snow Melt Patterns	1
Introduction	1
Study Area and Data Sources.....	4
Study area	4
Data Sources	5
Methods.....	5
Join Count Statistic.....	6
Results	8
Discussion	11
Limitations.....	12
Conclusion	13
References	15
Chapter 2: Developing Dimensionless Depletion Curves	31
Introduction	31
Study Area and Data	32
Study area	32
Data Sources	33
Methods.....	33
Model development	33
Identifying Effects the of Cloud Cover	35
Developing Dimensionless Depletion Curves.....	36
Results	37
Conclusion	39
References	41

Chapter 3: Derivation and Application of a Spatial Model (PCA)	50
Introduction	50
Study Area and Data Sources	54
Study area	54
Data Sources	54
Methods	55
Model Development	55
Combining the Data	56
Results	58
Climate Modeling Results	58
Model Accuracy	59
Cloud Removal Results	64
Discussion	66
Limitations	72
Conclusion	73
References	75
Chapter 4: Spatial Model Development (Clustering)	97
Introduction	97
Data Sources and Study Site	99
Study area	99
Data Sources	100
Methods	100
Model Development	100
Combining the Data	102
Results	105
Discussion	110
Conclusions	112
References	114
Chapter 5: Concluding Remarks	129

List of Figures

Figure 1.1: Study Area.....	20
Figure 1.2: Join Counts SCA 85-55%	21
Figure 1.3: Join Counts SCA 45-15%	22
Figure 1.4: Spatial Comparison 85% SCA	23
Figure 1.5: Spatial Comparison 75% SCA	24
Figure 1.6: Spatial Comparison 65% SCA	25
Figure 1.7: Spatial Comparison 55% SCA	26
Figure 1.8: Spatial Comparison 45% SCA	27
Figure 1.9: Spatial Comparison 35% SCA	28
Figure 1.10: Spatial Comparison 25% SCA	29
Figure 1.11: Spatial Comparison 15% SCA	30
Figure 2.1: Study Area.....	46
Figure 2.2: Max SWE Base Camp SNOTEL	46
Figure 2.3: Fitted Depletion Curve 2013.....	47
Figure 2.4: Dimensionless Depletion Curve.....	48
Figure 2.5: Fitted Depletion Curve 2001	48
Figure 2.6: Fitted Depletion Curve 2011	49
Figure 3.1: Study Area.....	85
Figure 3.2: LDS and FDL Output 2013.....	86
Figure 3.3: Histograms of Cloud 2013 and 2007	87
Figure 3.4: PCA Raster Output.....	88
Figure 3.5: Contour Map	89
Figure 3.6: Model Fit 4/28/18.....	90
Figure 3.7: Forest Influence on Accuracy	91
Figure 3.8: Available Cloud Free Imagery	92
Figure 3.9: Proof of Concept – Cloud Removal	93
Figure 3.10: Cloud Removal Accuracy	94
Figure 3.11: SWE by SNOTEL Station.....	95
Figure 3.12: Average Measured Cloud per Pixel	95
Figure 3.13: Depletion Curves for 2011 and 2007	96

Figure 4.1: Study Area.....	121
Figure 4.2: Clustering Model Raster	122
Figure 4.3: Model Fit 76% SCA.....	123
Figure 4.4: Model Fit 33% SCA.....	124
Figure 4.5: Model Fit 22% SCA.....	125
Figure 4.6: Two Ocean Plateau SNOTEL SWE.....	126
Figure 4.7: Effects of Forest on Accuracy.....	127
Figure 4.8: Depletion Curve for 2013.....	128

List of Tables

Table 1.1: Images Used for Join Counts.....	19
Table 2.1: DOY of Max SWE at Base Camp SNOTEL.....	44
Table 2.2: Cloud Cover Reduction	45
Table 3.1: Principal Component Weights.....	79
Table 3.2: Model Error Using PCA.....	80
Table 3.3: Average Model Error by SCA Percent.....	81
Table 3.4: Portion of Model Error Related to Forest Cover	81
Table 3.5: Melt Out Dates by SNOTEL Station.....	82
Table 3.6: Correlation Matrix of Principal Components	83
Table 3.7: Relationship between Principal Components and Cloud	84
Table 4.1: Model Error Using Clustering.....	119
Table 4.2: Adjusted Errors.....	119
Table 4.3: Model Fit between Represented SCA Percentages	120
Table 4.4: SWE by SNOTEL Site for 2017 and 2018.....	120

List of Equations

Equation 2.1: Fractional Snow Covered Area	36
Equation 2.2: Z Score	37
Equation 3.1: Matrix Multiplication for PCA.....	57
Equation 4.1: Z Score	110

Chapter 1: Exploring the Existence of Snow Melt Patterns

Introduction

Water stored as snow serves as an important component of available freshwater. Well over half of the world's potable water supply is extracted from rivers [Barnett et al., 2005]. In the Western United States 50-80% of the annual streamflow comes from snowmelt [Stewart et al., 2004]. In snowmelt dominated areas such as the example of the Western United States, water managers and legal systems rely on accurate modeling of discharge amount and timing [Qualls et al., 2013]. Remote sensing of snow has been available since the 1970's with the Advanced Very High Resolution Radiometer (AVHRR) with daily imagery available at a spatial resolution of 1.1 kilometers at nadir [Jensen, 2016]. This provides spatially distributed imagery of snow on a daily time step, an ideal temporal resolution for monitoring snow coverage. This information has been used by many research groups to derive information which can be used for modeling purposes [Martinec & Rango, 1986]. The spatial resolution of 1.1 kilometers may not accurately depict snow coverage in complex terrain [Rango, 1993]. With the launch of the MODerate Resoluituion Imaging Spectroradiometer (MODIS) on board the Terra satellite, and a second sensor on board the Aqua satellite, worldwide daily imagery of snow cover at a spatial resolution of 500 meters is available from 2000 through to the present [Jensen, 2016]. Microwave remote sensing has developed into an exciting data source for snow monitoring. Remotely sensed data in the microwave wavelengths along the electromagnetic spectrum can detect snow water equivalent (SWE) and snow depth even on cloudy days. The Advanced Microwave Scanning Radiometer- Earth Observing System (AMSR-E) collects SWE and snow depth information which can be used in snow melt modeling, but the spatial resolution of 25 kilometers leaves this data at a disadvantage in complex terrain [Wulder et al., 2007; Pulliainen, 2006]. The spatial resolution of AMSR-E is the best available for collecting data on snow depth and SWE. These data sources provide spatially distributed data with applicable temporal resolution for monitoring snow. With this information, snowmelt modeling may become more accurate. One statistical test has been used on spatially distributed information which can lead to further understanding of the nature of space on measurement values: autocorrelation. Autocorrelation is a global statistic which identifies

whether the assumption that the spatial distribution of measurements is independent of the others [Brunsdon & Comber, 2015]. This measure can indicate whether location is a factor in determining a value of a particular measurement.

Remote sensing of snow is a great tool for modeling snowmelt runoff; however, each type has complications. SWE and snow depth measurements are measured by means of microwave remote sensing. The low energy output over large ranges of wavelengths in the microwave spectrum forces this remote sensing to have large spatial resolutions from 25 kilometers and up. The coarse resolution hinders the use of this data in operation hydrological modeling [Kelly et al., 2003; Dressler et al., 2006]. The high variability in SWE across a landscape may also lead to incorrect representation of the spatial distribution by remote sensing. Spatial distribution of SWE is dependent on accumulation, topography, slope, aspect, vegetation, and a number of other factors [Jost et al., 2007]. This combination of large spatial resolution and high variability in SWE is a significant setback for using microwave remote sensing. On the other hand, snow extent is measured in the visible and mid-infrared wavelengths. The high amount of measureable energy over small ranges in wavelengths in the visible spectrum leads to finer spatial resolution, 500 meters with MODIS. The Normalized Difference Snow Index (NDSI) has been applied in the MODIS products to help differentiate between snow and cloud cover [Hall & Riggs, 2015]. The NDSI is a combination of two bands, one associated with the visible wavelengths (Red) and the other with the mid-infrared wavelengths. Since a portion of the calculation of the NDSI relies on visible wavelengths, snow extent detection is hindered by cloud cover. The MODIS product is also subject to uncertainty and errors at the pixel level due to viewing conditions, land cover, the amount of snow, and cloud cover or masking [Riggs et al. 2017]. On cloud free days the MODIS product showed 93% agreement with ground measurements, therefore cloud cover is a persistent challenge [Hall & Riggs, 2007]. Although the measurement of snow extent is less desirable, the finer resolution along with a daily temporal resolution makes it quite useful for research.

A substantial amount of research has been focused on addressing the issues associated with cloud cover and identification of snow in varied land cover. Many research groups have tackled the issue of cloud cover with good success. Combinations of the MODIS sensors on board the Aqua and Terra satellites, which have different pass over

times, have been used as a temporal filter to remove cloud cover [Dietz et al., 2012]. Other temporal filters have used imagery from preceding and succeeding days to identify whether a pixel is cloud covered [Parajka & Blöschl, 2008; Wang & Xie, 2009]. Spatial interpolation has also been used, which relies on adjacent pixels to inform the value of a cloud covered pixel [Gafurov & Bárdossy, 2009]. Elevation models have also been used. These define an elevation range that is considered to be all snow, a transition elevation range where pixels can be snow or land, and an elevation range which are all land [Parajka et al., 2010; Krajčí et al., 2014]. Step by step methods that incorporate the simpler methods outlined above with other more technical statistical methods to remove cloud cover. Examples of these are the MODSNOW-tool, and the Adaptive Spatio-Temporal Weighted Model (ASTWM) [Gafurov et al., 2016; Li et al., 2017]. These advanced methods boast cloud free daily imagery with agreement as good as 98.92% up to 99.8%. Other research groups have focused on accurate classification of pixels to address the variation in NDSI values based on land cover. MODSCAG uses a combination of spectral responses on a pixel by pixel basis to let each pixel independently define what spectral response is associated with snow cover [Painter et al., 2009].

Changes in snow pack and extent are an area of significant research in climate modeling. If the intensity in precipitation does not change, a warming climate could lead to less precipitation coming as snow leading to earlier peak discharge in rivers and streams [Barnett et al., 2005]. Spatial statistics offer a way to analyze these changes. A vital assumption of a linear regression model is independence of values. For spatial data this includes spatial dependence. Many research groups have developed regression models to estimate SWE over a basin [Mizukami et al., 2011; Bocchiola, & Groppelli, 2010]. A fundamental step to verifying the validity of the model is a measure for spatial autocorrelation. Chang et al., [2014] found that maps of runoff ratio and coefficient of variation of the runoff ratio showed strong positive spatial autocorrelation and application of a Geographically Weighted Regression (GWR) model more accurately predicted runoff ratio's than an ordinary least squares (OLS) model. Both used the same predictor variables with adjusted R-squared values increased from 0.793 to 0.828. The application of a GWR model reduces the spatial autocorrelation in the residuals of the model preserving the independence of predictor variables. To our knowledge, measures of autocorrelation have

not been widely applied to snow extent, for good reason. Snow extent is by nature a pattern, exhibiting high degrees of autocorrelation. Areas of snow cover are more likely to be close to other areas of snow cover, rather than completely spatially random (CSR). It has long been accepted snow exhibits a pattern in snow covered area (SCA) across multiple years; therefore the use for autocorrelation is not needed. The pattern of snow melt has been noted to exist independent of snow depth [Adams, 1976; König & Sturm, 1998; Luce et al., 2004; Sturm & Wagner, 2010]. The purpose of this research is to conduct a formal assessment of the existence of a repeatable pattern in the spatial distribution of snow as melt occurs for the period of 2000 through 2018. Spatial autocorrelation does not specifically address this; however the degree of spatial autocorrelation may prove to be useful in tandem with an assessment of the location of the SCA.

Study Area and Data Sources

Study area

Our study area, the Upper Snake Basin (Fig. 1.1), covers an area of 3,465 mi² (8,894 km²) and elevation ranges between 5,799 -13,760 feet (1,737 – 4,194 meters). The extent of the watershed has a Northing of 4752708.5 to 4938033.6 meters and an Easting of 262199.4 to 828367.5 meters in the Albers 1927 datum and UTM zone 12 projection. The area serves as an important water resource for Idaho agriculture, power generation, in stream fish requirements, as well as municipal, commercial, and industrial uses. The watershed is bordered to the West by the Teton Mountains and includes the hydrologic units known as the Snake Headwaters, Gros Ventre, and Greys-Hobock. Snow seasonally covers the watershed making remote sensing a useful tool for collecting snow cover information. The study site also has 11 ground based monitoring stations (SNOTEL sites), which measure a 2 meter by 2 meter snow pillow and collect snow water equivalent (SWE) measurements. The study site land cover is variable. Forest accounts for 47.9 percent of the total area with 46.6 percent of the total area classified as evergreen forest. Shrub and grass land account for 45.4 percent of the total area with the remaining 6.7 percent belonging to agriculture, developed land, and wetlands. Percentages calculated are based on the 2011 National Land Cover Database [Homer et al. 2015].

Data Sources

The MODerate Resolution Imaging Spectoradiometer (MODIS) on board the Terra satellite was used. The version six data product, MOD10A1 (MODIS Terra snow cover daily L3 global 500 m Sin GRID V006) distributed by the National Snow and Ice Data Center (NSIDC) was used. This data product comes with NDSI snow cover, NDSI snow cover basic QA, NDSI snow cover algorithm flags QA, NDSI, and snow albedo data. In this study only the NDSI snow cover was used because it gives information about snow and cloud extent. MOD10A1 was chosen as it has consistently shown less cloud obstruction [Parajka & Blöschl, 2008]. The calculation of NDSI is a combination of the red band (0.545-0.565 μm) and a middle infrared band (1.628-1.652 μm). The NDSI equation is the difference between the red and middle infrared bands divided by the sum of the two. The MODIS sensor aboard Aqua had a failure in the mid infrared band forcing a slightly different NDSI calculation. Since no temporal combination of Terra and Aqua was applied, only MOD10A1 product from the Terra satellite was used. To convert the data to fractional snow covered area (SCA) a constant threshold of 0.4 was used, as documented in the version six user guide [Hall & Riggs, 2015]. Two MODIS swaths were needed to encompass the study region, h09v04 and h10v04. All available MODIS imagery from 2000 through 2018 was used in this study.

Methods

Two methods were used to assess the existence of a pattern of the spatial distribution of snow at different SCA across multiple years. A join count statistic was applied, which measures the degree of autocorrelation. Positive and negative autocorrelation signify spatial dependence where no autocorrelation represents complete spatially randomness. In the context of this research spatial dependence is expected since snow cover is by nature distributed based on physical characteristics of a watershed which are not random. Identifying the existence of a pattern requires a comparison of the join count statistic and a comparison of the location of snow cover. The join count statistic is a global autocorrelation statistic (one value describing the entire image), which is the reason an analysis of location of the pattern is required.

Join Count Statistic

The join count statistic is a method for measuring spatial autocorrelation on binary data. Where other more frequently applied statistics for measuring spatial autocorrelation (Moran's I) are applicable to continuous data the join count statistic is defined explicitly for binary datasets. In the case of the MODIS daily snow cover products an NDSI is given. While it is tempting to apply Moran's I to this data, which is on a scale of 0 to 100, it would be incorrect because the data is binary in nature not continuous. Snow exhibits an NDSI value of greater than zero [Riggs et al., 2017]. The complication is an NDSI value for a pixel which is 50% snow covered varies based on land cover (forested or not), viewing angle of the sensor, illumination of the pixel, and the slope of the pixel [Riggs et al., 2017]. Where one NDSI on a given day may mean a certain percentage of snow cover, it may not mean the same snow cover on a different day, or even at a different location. For this reason the Moran's I is not applicable to this data without significant processing of the imagery and ground based verification. The join count statistic was applied to deal with this issue. A constant NDSI threshold of 0.4 (40 on the scale of 0 to 100) was applied to the data consistent with the MODIS user guide recommendations. The threshold converts the NDSI imagery to a binary snow or no snow image where cloud cover is also included in the no snow portion. This also may be subject to the errors in NDSI; however processing time is reduced greatly for a comparison of this size.

All imagery with cloud cover less than 5% of the total area were converted into the binary snow and no snow. The cloud free imagery (<5% cloud cover by area) from 2000 through 2018 during the melting period or day of year (DOY) 1-200 were used. Images at SCA values of 85, 75, 65, 55, 45, 35, 25, and 15 percent were identified and converted into the binary format. Images rarely measure exactly at a particular SCA; therefore at a given SCA a $\pm 3\%$ buffer was used. If for example, an image measures at 78% SCA it would be used for the 75% SCA comparison. This means theoretically if both the image allowable SCA error and allowable cloud cover error were maxed out a difference of 8% could be seen in snow covered area. The average cloud cover was 2.6% with a maximum of 4.8%. A total of 53 images fit these criteria.

Performing a join count statistical measure of spatial autocorrelation requires the definition of a neighborhood. The test visits each pixel in the image and analyzes the

neighboring pixels to calculate the statistic. Three methods can be used for the neighborhood rook's case, queen's case, and k-nearest neighbor. Queen's case was used in this analysis as it is the most theoretically fitting. The rook's case only analyzes the pixels bordering the pixel in question, a neighborhood of 4 pixels. The queen's case analyzes the bordering and adjacent pixels, a neighborhood of 8 pixels. The k-nearest neighbor is most applicable to non-uniformly sized data such as counties within a state. Queen's case was chosen since snow melt is driven by a number of drivers such as elevation, slope, aspect, and more suggesting adjacent pixels are of interest when calculating the statistic. The edge of a melting snowpack is complex in its distribution throughout a watershed which can be captured by the queen's case neighborhood, but might be misrepresented by the rook's case. Row standardization was also applied, which normalizes the dataset. The Join Count statistic provides an expectation value and a measured value. A join is measured when pixels of differing values are adjacent to one another. The expectation is the calculated number of joins which would be expected if the distribution of points is completely spatially random. For an in depth discussion of the derivation of the join count statistic, refer to Cliff & Ord [1981]. With the expectation value and measured value both calculated for each image, plotting the two gives a method for comparing the statistics at a given SCA for images from multiple years. The combination of these two in fact describes the interface between snow and land. This is because a measured join is a location where a pixel within the queen's case neighborhood has a different value than the pixel being analyzed. If one snow pixel is surrounded by 2 no snow pixels, and subsequently 6 snow pixels, 2 joins would be recorded. The interface between land and snow is where all the joins are measured in theory.

The location of the extent of snow and interface between land and snow was identified through a spatial comparison of all the cloud free images at a given SCA. The binary imagery of snow or no snow was assigned a value of 0 for snow and 1 for no snow. By adding the images at a given SCA together and dividing by the total number of images used, the number of times a pixel differed at a across all the images at a given SCA can be shown. If there are five images to compare and one pixel has a value of one image disagreement, this indicates one of the five images had a different value, whether it was measured as snow once and land on the other 4 images or vice versa. The higher the number of disagreement the less consistent the pixel is in the pattern of melt.

Results

Snowmelt is driven by a number of energy fluxes. Spatial autocorrelation is a measure of the randomness of the spatial distribution of values. High degrees of spatial autocorrelation are expected for melting snow on a watershed scale. It is unexpected for snow to melt in a random fashion because snow accumulation, slope, aspect, elevation, all influence melt. With only accumulation varying, the physical drivers (watershed physical characteristics) of snowmelt are consistent therefore areas of snow are more likely to be near areas of snow in a watershed as melt occurs. What is of interest in this research is the pattern across multiple years. The join count statistic produces two values: expected joins and measured joins. In this context the expected value would be the number of times a snow pixel joins a land pixel based on the number of snow pixels in the image. The expected value is based on the value that has the lowest count in an image, which is why once SCA switches from being the majority to the minority the expected value declines. At high snow covered area percentages the expected value is based on the no snow pixel count and vice versa at low SCA percentages. This information can be compared for multiple years helping to identify the existence of a pattern in snowmelt by explaining the interface between snow and land or the “snow line”.

As expected, high degrees of spatial autocorrelation were recorded at every image. A high degree means the expectation is much larger than the joins value. Each was highly significant with p-values of less than $2.2e-16$ for every image. Figure 1.2 shows the comparison of join count statistic versus expected value at SCA values of 85, 75, 65, and 55 percent. Each point represents an image. The points are grouped together based on SCA percentage. Each group of points is distinctly different than the next especially along the x-axis (expected value). There is also a difference in the y axis. As the SCA decreases the Joins increase. It is expected that there would be a difference in the expected value, but very meaningful that the y axis or join counts also follows a trend. This shows that the interface between land and snow follows a trend from year to year. Figure 1.3 shows the SCA values of 45, 35, 25, and 15 percent. As snow covered area decreases the expectation decreases, which follows the definition of how the expected value is defined as discussed above. The trend in the join counts in this plot follows an increasing trend as SCA increases. The percentages of 35 and 45 appear quite similar in join counts. This may be due to the limited

number of images used for these two percentages, four for 35 and six for 45 percent. The expectation value at 15 percent SCA is also interesting falling on 2 distinct expectation values rather than the mixture of expectation values seen at all the other SCA values. One can note the two figures produce similar expectation values. When all the image used are plotted on one plot 85 and 15 percent SCA plot on top of one another, as does 75 and 25, 65 and 35, and 55 and 45. This is because of how the expectation and join count statistic are defined. It is interesting though the join counts produce similar results at inverse SCA value such as 85 and 15. This suggests the length of the interface at these two snow covered areas are similar. At the mid-range SCA values the expectation and join counts are closer together. Even as melt occurs between 35 and 65 percent SCA the length of the snow line changes very little, an interesting finding.

The interface between snow and land at a constant SCA across multiple years shows a trend; however without the spatial comparison of the snow extent this alone does not identify a year to year pattern. At each of the SCA values used in the join count analysis the associated images were compared through raster algebra (addition of the pixel values with the same geographic location). At each SCA one image was created by adding all the images available at that SCA. To normalize the data, the final image was divided by the number of images used. This resulted in pixel values ranging from 0 to 1. With five images used the range of possible values would be 0, 0.2, 0.4, 0.6, 0.8 and 1. A zero or one indicates no difference at that particular pixel between all 5 images used. A 0.2 and 0.8 indicate one image differed from the other four at the pixel level. Subsequently a 0.4 and 0.6 indicate 2 images differed. The figures use black to indicate agreement between all the images used. Black represents agreement between snow and no snow. A grey scale was used to show image disagreement on the pixel level. Some disagreement is expected since no image has exactly the same SCA as another. In all the figures the interface between land and snow can be tracked by the edge of the disagreement values or grey to white pixels. The thicker disagreement line is the more spatial difference was recorded. A very thin grey line shows a very similar snow line across all the images compared while a thick one shows a bigger spatial difference in the distribution of snow. Ideally the images would be all black.

Figures 1.4 through 1.11 show the agreement of the spatial distribution of snow at different SCA percentages. It is important to note the cloud is considered no snow. Raw

imagery was used in this analysis and no attempt to interpolate the data was made. Figure 1.4 represents 85% snow cover and each subsequent figure is a 10% reduction in snow cover, presented in the same progression as snow melts. As melt begins (Fig. 1.4) the interface between land and snow is shifted in one image. The northern light grey colored area is a one image disagreement. Very few pixels show a 2 image disagreement. This shows that the pattern is mostly steady across the 5 images compared. At 75% SCA (Fig. 1.5) the Southeastern portion of the watershed is less consistent showing 3 image disagreement. The main bulk of melt, which is encircled by the white and grey, is the same and the line is similar in many locations. Figure 1.6 shows the stabilization of the pattern in the southern end at 65% SCA. The disagreements in Figure 1.5 are all but gone in Figure 1.6. The North and Eastern edge of melt is now blurred with two and three image disagreement. The main area of melt is once again stable. At 55% SCA the majority of the snow line is extremely consistent across all the images used (Fig. 1.7). The northern end of the main melt still shows multiple image differences. It is between 55% SCA and 45% SCA that a significant increase in spatial agreement is seen.

The agreement of all the images used at 45% SCA (Fig. 1.8) can be seen by the very thin line representing the snow line. Even with disagreement it is so little around the snow line at this percentage it can be safely stated the pattern is stable at this percentage especially since 6 images were compared. This trend continues as the SCA decreases. Figures 1.9 through 1.11 all show good agreement in the spatial distribution of snow. The relationship between the spread of join count statistics and spatial agreement is evident when both are considered. The SCA with the worst spatial agreement 75-65% also show a great spread in the join count values in Figure 1.2. In fact at 75% SCA there appear to be two groups for the join count statistic. Three of the seven images used group together, this is represented in Figure 1.5 as 3 images showed disagreement in the spatial extent of snow. While the join count statistic does not describe explicitly the spatial extent of snow, it is related. This is further backed up by the tight groupings of join count statistics in Figure 1.3 with a range of 315 at 45% SCA compared with the range of 661 at 55% SCA and 620 at 75% SCA in Figure 1.2. In terms of the statistic as snow melts the join count statistic decreases meaning a more stable snow line. This is also seen in the spatial comparison of the imagery.

Discussion

The implications of a repeatable pattern in snow melt are wide spread and underutilized. This long accepted fact has remained a novelty for remote sensing. In an aggregated way dimensionless depletion curves have been proposed, using the pattern in a non-spatial way [Luce & Tarboton, 2004]. These curves represent the spatial change in snow extent through a melting season with time. The dimensionless part of the curve is added by shifting the curve and altering the rate or steepness which can be fit to any snow year. Depletion curves have long been developed as input into snow melt models particularly the snowmelt runoff model (SRM) [Martinec & Rango, 1986]. In these cases a depletion curve is derived for each year independent of previous years. A repeatable pattern reinforces the applicability of the dimensionless depletion curve which fits each year. A repeatable pattern in SCA changes across multiple years is a relationship that can be used in real-time modeling, cloud removal, climate modeling, and can be used as spatial information in ungauged watersheds.

Modeling snowmelt has used remote sensing as an input as early as 1982 with the AVHRR sensor [Baumgartner et al., 1987]. Currently remote sensing of snow cover is only used as input when cloud cover is below a certain threshold or interpolated [Steele et al., 2017]. This preliminary work shows a significant stability in the pattern of melt across a 19 year period. With this information accurate depiction of the spatial coverage of snow at multiple snow covered areas can be developed. A very important note about the pattern can be seen in Table 1.1. The DOY and year of each image used in this analysis is shown sorted by SCA. While differences in snow packs exist this information shows the variability in the timing that each SCA is achieved. The most surprising value is the range in the DOY at each SCA. The largest range in DOY of 39 is measured at an SCA of 25%. The smallest of 16 days is associated with 15% SCA. All SCA percentages except for 15% had more than 20 days difference. This information shows the high variability in the snow melt timing. A criticism of this work may be that years of similar snow and melting conditions were used; however this shows significant shifts in the timing of melt across all SCA percentages and years. The patterns exist regardless of the snow depth, max SWE, and melt timing. A SCA value of 45% is the same spatially in any year. This assessment shows the stability of the pattern even with some variability at SCA percentages of 55 and 65.

The most exciting application of a repeatable pattern is ungauged watersheds. Spatial information is extremely useful for real time modeling, cloud removal, and climate modeling, but ungauged watersheds have relied on interpolation for distributed data [Teutschbein et al., 2018]. A repeatable pattern, which is shown to exist here with remotely sensed data, is a new exciting application for spatial information in ungauged watersheds. The pattern is a product of the physical characteristics of a watershed, which is available through remote sensing. Where data in these locations has relied on interpolations, real spatial data is available. Since the pattern is applicable across multiple years and snow pack conditions data from years passed is useful.

Remote sensing of snow offers a variety of data, some of which are more important for modeling than the spatial extent of snow. Remote sensing of SWE and depth of snow have both become an area of interest. The question remains, why is the spatial extent of snow useful if SWE, a much more desirable variable, can be interpolated? Two factors determine the reason snow covered area is presently more valuable. One, the spatial resolution of SWE measurements is too large to capture its variability in complex terrain. Most importantly year to year SWE measurements are driven by precipitation and temperature which may not show a significant year to year pattern in spatial distribution. Snow extent on the other hand shows a stable pattern across multiple years and varying snow melt timing. While the pattern in this watershed may be slightly more variable between 75-55% SCA, once snow cover is less than 50% the spatial pattern and interface is exceptionally stable. This is shown in terms of the spatial coverage and the join count statistics which are more clustered together at the lower SCA values. This is shown to be even more valuable with the wide range in DOY at 25% SCA of 39 days for the images used. The pattern exists even if melt timing is significantly different.

Limitations

In this study a constant threshold was applied to the NDSI to transform the imagery into a binary snow or no snow image. The fact is any NDSI value greater than zero may indicate some snow in the pixel [Riggs et al., 2017]. With this watershed nearly 50% covered by evergreen forest errors in the spatial distribution of snow in the forested areas are unavoidable using the constant threshold. NDSI values of 0.2 or greater have been shown to indicate complete coverage in forested areas, which is substantially lower than the 0.4 used

in this study [Rittger et al., 2013]. While this use of a constant threshold limits this study it does not disprove the pattern, but may underestimate snow coverage. Using a varied threshold also may not fix the issue as viewing angle of the sensor, illumination, and age of the snow all play a part in defining the NDSI value. One positive NDSI value may represent a wide variety of snow coverage [Riggs et al., 2017]. With this in mind a constant threshold was applied for ease and because the solution to identifying snow is not as simple as a varied threshold based on land cover. As identification of snow with remote sensing becomes more accurate the pattern may become even more stable.

Conclusion

The goal of this research was to investigate the existence of a pattern in snowmelt across multiple years. This idea has long been accepted in the literature [Adams, 1976; König & Sturm, 1998; Luce & Tarboton, 2004; Sturm & Wagner, 2010]. No formal assessment of the pattern to our knowledge has been conducted, merely noted by other research groups. Two methods were applied to identify the stability of the pattern, one a global statistic describing the interface between snow and land and the other a spatial comparison of the location of the pattern. Both analyses showed separation between SCA percentages and a stable pattern. Even with the variability in the spatial location of the pattern at the higher SCA percentages the pattern was significant even with a high variability in the snowmelt timing, or the DOY associated with the images used. These findings support the long accepted idea of patterns at the watershed scale. The pattern is a result of the watershed physical characteristics which do not change such as elevation, slope, and aspect which are all important to the radiation a snow pack receives. For this reason a year to year intra-annual pattern exists, which is stable due to these unchanging physical characteristics. Applications of the pattern are widespread with a particularly exciting area of ungauged watersheds. Spatial data on a daily time step at a spatial resolution of 500 meters is available from 2000 to the present. This research shows the imagery from years passed can inform future years with spatial representation of snow extent. While other remote sensing technologies develop such as microwave remote sensing the spatial extent of snow offers a finer resolution to observe daily snow cover changes through a melting season.

The patterns in melt offer a new untapped area of research in snowmelt modeling. This information is applicable to real time modeling, climate modeling, cloud removal, and offers extremely valuable information in ungauged watersheds. Further research is required into developing a methodology for applying this information to hydrological models. Next steps in this research are to develop an application of this information. The old idea of patterns may prove to be useful information as we manage our water in a changing climate.

References

- Adams, W. P. (1976). Areal differentiation of snow cover in east central Ontario, *Water Resources Research*, 12(6), 1226–1234, doi:10.1029/WR012i006p01226.
- Barnett, T. P., Adams, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438, 303-309.
- Baumgartner, M., Seidel, K., & Martinec, J. (1987). Toward Snowmelt Runoff Forecast Based on Multisensor Remote-Sensing Information. *Geoscience and Remote Sensing, IEEE Transactions on, GE-25(6)*, 746-750.
- Bocchiola, D., & Groppelli, B. (2010). Spatial estimation of snow water equivalent at different dates within the Adamello Park of Italy. *Cold Regions Science and Technology*, 63(3), 97-109. doi:10.1016/j.coldregions.2010.06.001.
- Brunsdon, C., & Comber, L. (2015). *An introduction to R for spatial analysis & mapping*. Los Angeles: SAGE.
- Chang, H., Johnson, G., Hinkley, T., & Jung, I. (2014). Spatial analysis of annual runoff ratios and their variability across the contiguous U.S. *Journal of Hydrology*, 511(C), 387-402. doi:10.1016/j.jhydrol.2014.01.066.
- Cliff, A., & Ord, J. K. (1981). *Spatial processes : Models & applications*. London: Pion.
- Dietz, A., Wohner, C., & Kuenzer, C. (2012). European Snow Cover Characteristics between 2000 and 2011 Derived from Improved MODIS Daily Snow Cover Products. *Remote Sensing*, 4(8), 2432-2454. doi:10.3390/rs4082432.
- Dressler, K. A., Leavesley, G. H., Bales, R. C., Fassnacht, S. R., Pomeroy, J., Klein, A. G., & Elder, K. (2006). Evaluation of gridded snow water equivalent and satellite snow cover products for mountain basins in a hydrologic model. *Hydrological Processes*, 20(4), 673-688. doi:10.1002/hyp.6130.
- Gafurov, A., & Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrology and Earth System Sciences*, 13(7), 1361-1373.
- Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., & Merz, B. (2016). MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environmental Earth Sciences*, 75(14), 1-15. doi:10.1007/s12665-016-5869-x.
- Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12), 1534-1547. doi:10.1002/hyp.6715.

- Hall, D.K. & Riggs, G. A. (2015). *MODIS Snow Products Collection 6 User Guide*. doi: <http://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf>. [July 12, 2018].
- Hall, D. K. & Riggs, G. A. (2016). *MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6*. [NDSI snow cover]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: <https://doi.org/10.5067/MODIS/MOD10A1.006>. [January 2017 - March 2018].
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., (2015), [Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information](#). *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354
- Jensen, J. (2016). *Introductory Digital Image Processing: A Remote Sensing Perspective* (4th ed.).
- Jost, G., Weiler, M., Gluns, D. R., & Alila, Y. (2007). The influence of forest and topography on snow accumulation and melt at the watershed-scale. *Journal of Hydrology*, 347(1), 101-115. doi:10.1016/j.jhydrol.2007.09.006.
- Kelly, R. E., Chang, A. T., Leung, T. & Foster, J. L. (2003). A prototype AMSR-E global snow area and snow depth algorithm. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(2), 230-242. doi:10.1109/TGRS.2003.809118.
- König, M. & Sturm, M. (1998). Mapping snow distribution in the Alaskan Arctic using aerial photography and topographic relationships, *Water Resour. Res.*, 34(12), 3471–3483, doi:10.1029/98WR02514.
- Krajčič, P., Holko, L., Perdigão, R. A. P., & Parajka, J. (2014). Estimation of regional snowline elevation (RSLE) from MODIS images for seasonally snow covered mountain basins. *Journal of Hydrology*, 519(PB), 1769-1778. doi:10.1016/j.jhydrol.2014.08.064.
- Li, X., Fu, W., Shen, H., Huang, C., & Zhang, L. (2017). Monitoring snow cover variability (2000–2014) in the Hengduan Mountains based on cloud-removed MODIS products with an adaptive spatio-temporal weighted method. *Journal of Hydrology*, 551, 314-327. doi:10.1016/j.jhydrol.2017.05.049.
- Luce, C. H., & Tarboton, D. G. (2004). The application of depletion curves for parameterization of subgrid variability of snow. *Hydrological Processes*, 18(8), 1409-1422. doi:10.1002/hyp.1420.
- Martinec, J., & Rango, A. (1986). Parameter values for snowmelt runoff modelling. *Journal of Hydrology*, 84(3), pp.197–219. doi:10.1016/0022-1694(86)90123-X.

- Mizukami, N., Perica, S. & Hatch, D. (2011). Regional approach for mapping climatological snow water equivalent over the mountainous regions of the western United States. *Journal of Hydrology*, 400(1), 72-82. doi:10.1016/j.jhydrol.2011.01.019.
- Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4), 868-879. doi:10.1016/j.rse.2009.01.001.
- Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. *Water Resources Research*, 44(3), N/a. doi:10.1029/2007WR006204.
- Parajka, J., Pepe, M., Rampini, A., Rossi, S., & Blöschl, G. (2010). A regional snow-line method for estimating snow cover from MODIS during cloud cover. *Journal of Hydrology*, 381(3), 203-212. doi:10.1016/j.jhydrol.2009.11.042.
- Pulliainen, J. (2006). Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sensing of Environment*, 101(2), 257-269. doi:10.1016/j.rse.2006.01.002.
- Qualls, R. J., Taylor, R. G., Hamilton, J., & Arogundade, A. B. (2013) Climate change opportunities for Idaho's irrigation supply and deliveries. *Journal of Natural Resources Policy Research*, 5(2-3), 91-105, DOI: 10.1080/19390459.2013.811856.
- Rango, A. (1993). II. Snow hydrology processes and remote sensing. *Hydrological Processes*, 7(2), 121-138. doi:10.1002/hyp.3360070204.
- Riggs, G.A., Hall, D. K., & Roman, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, 9(2), pp.765–777. doi:10.5194/essd-9-765-2017.
- Rittger, K., Painter, T. H., & Dozier, J. (2013). Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, 51(C), 367-380. doi:10.1016/j.advwatres.2012.03.002.
- Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A., & Bleiweiss, M. (2017). Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the Rio Grande headwaters. *International Journal of Applied Earth Observations and Geoinformation*, 63, 234-243. doi:10.1016/j.jag.2017.08.007.
- Stewart, I., Cayan, T., & Dettinger, D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a 'Business as Usual' Climate Change Scenario. *Climatic Change*, 62(1), 217-232. doi:10.1023/B:CLIM.0000013702.22656.e8.

- Sturm, M., & Wagner, A. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*, 46(12), N/a. doi:10.1029/2010WR009434.
- Teutschbein, C., Grabs, T., Laudon, H., Karlsen, R. H., & Bishop, K. (2018). Simulating streamflow in ungauged basins under a changing climate: The importance of landscape characteristics. *Journal of Hydrology*, 561, 160-178. doi:10.1016/j.jhydrol.2018.03.060.
- Wang, X., & Xie, H. (2009). New methods for studying the spatiotemporal variation of snow cover based on combination products of MODIS Terra and Aqua. *Journal of Hydrology*, 371(1), 192-200. doi:10.1016/j.jhydrol.2009.03.028.
- Wulder, M., Nelson, T., Derksen, C., & Seemann, D. (2007). Snow cover variability across central Canada (1978–2002) derived from satellite passive microwave data. *Climatic Change*, 82(1), 113-130. doi:10.1007/s10584-006-9148-9.

Tables

Table 1.1: Summary of imagery used for the analysis with year and DOY presented at different SCA percentages.

SCA %	85		75		65		55		45		35		25		15	
	Year	DOY														
	2003	98	2003	99	2004	103	2001	116	2002	133	2003	143	2001	132	2006	156
	2003	100	2005	106	2006	119	2002	124	2003	136	2012	134	2003	147	2010	169
	2004	89	2013	116	2008	137	2003	134	2004	116	2015	121	2005	146	2014	157
	2006	110	2014	123	2010	114	2004	114	2006	135	2017	151	2006	151	2014	162
	2018	111	2015	103	2013	122	2004	117	2009	138			2006	152	2017	170
			2017	124	2015	110	2006	133	2016	124			2008	168	2018	154
			2018	116	2018	117	2007	118					2008	171		
							2008	138					2017	158		
							2015	107								
							2017	130								
Range in																
DOY	22		25		34		31		22		30		39		16	

Figures

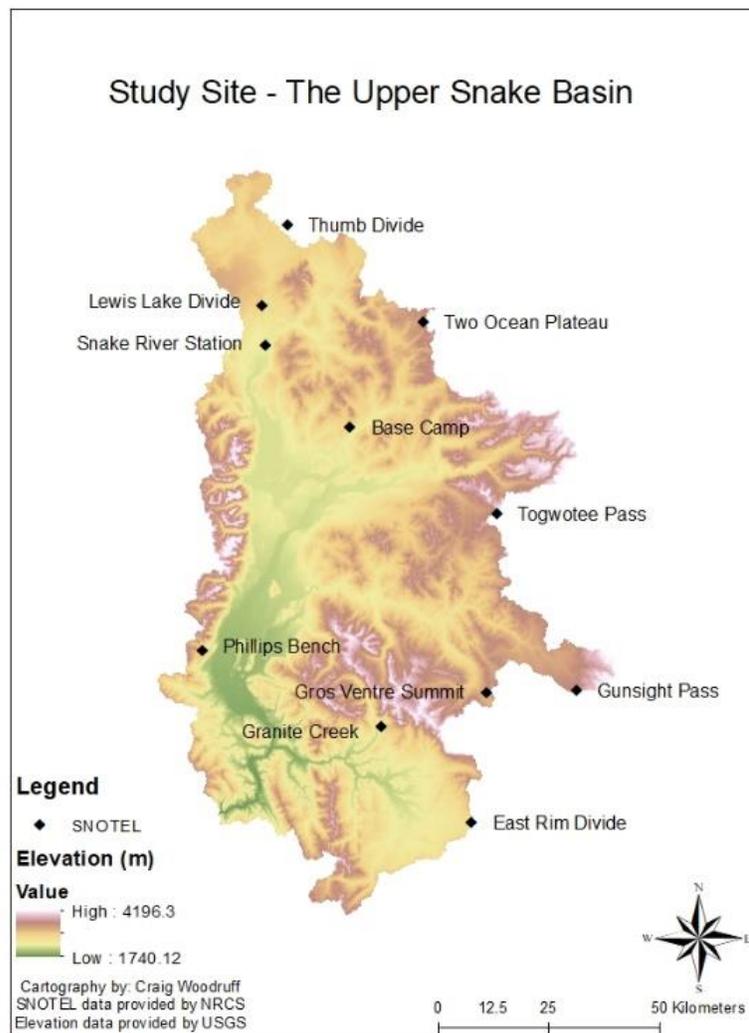


Figure 1.1: Upper Snake Basin with associated SNOTEL stations.

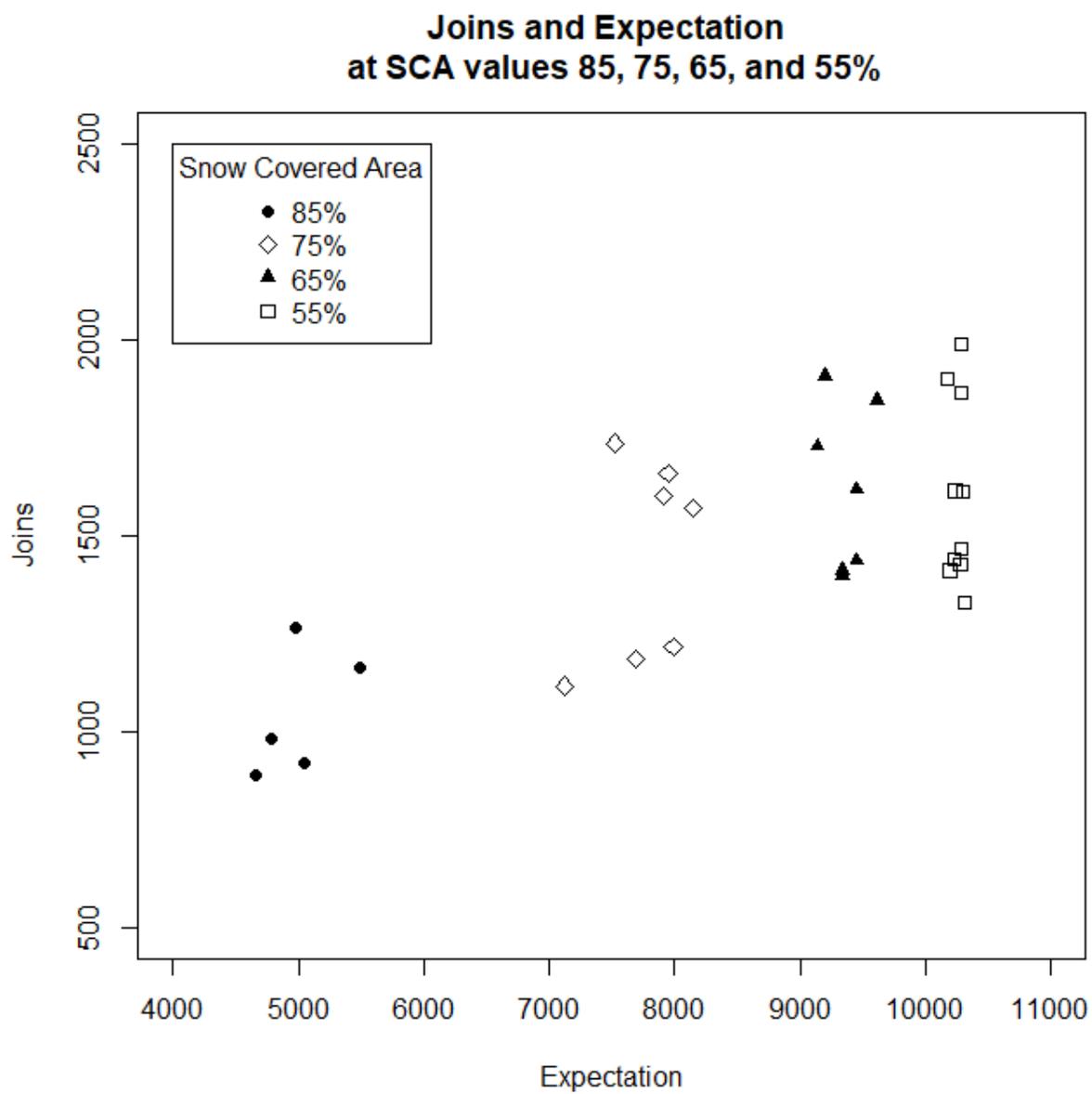


Figure 1.2: Comparison of join count statistic versus expected at snow covered areas between 85 and 55% across multiple years at varied snow covered areas. Each point represents one image.

Joins and Expectation at SCA values 45, 35, 25, and 15%

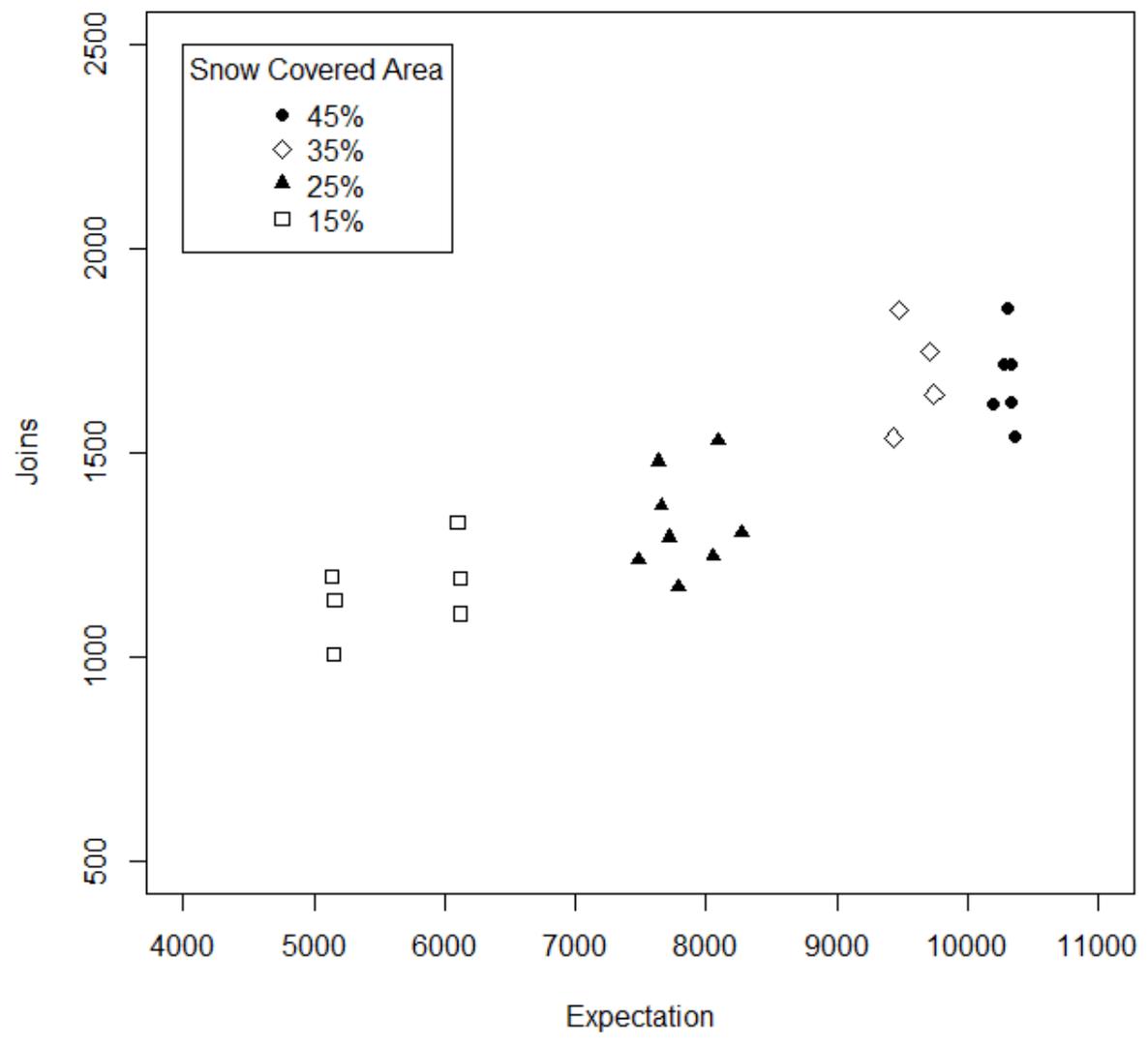


Figure 1.3: Comparison of join count statistic versus expected at snow covered areas between 45 and 15% across multiple years at varied snow covered areas. One point represents the values calculated for one image, while shape point shape differentiates snow covered area.

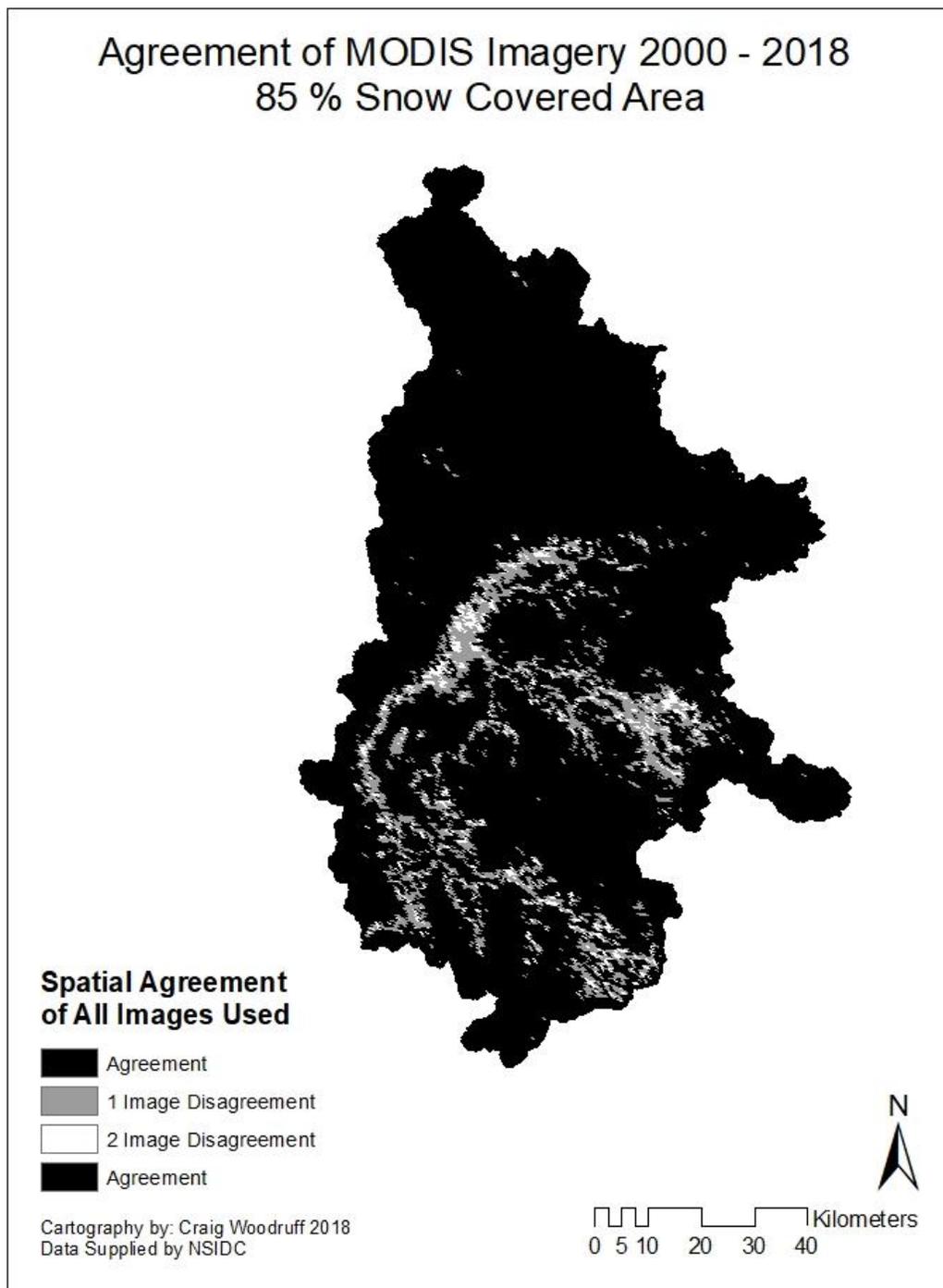


Figure 1.4: Agreement of all the images used at 85% snow covered area.

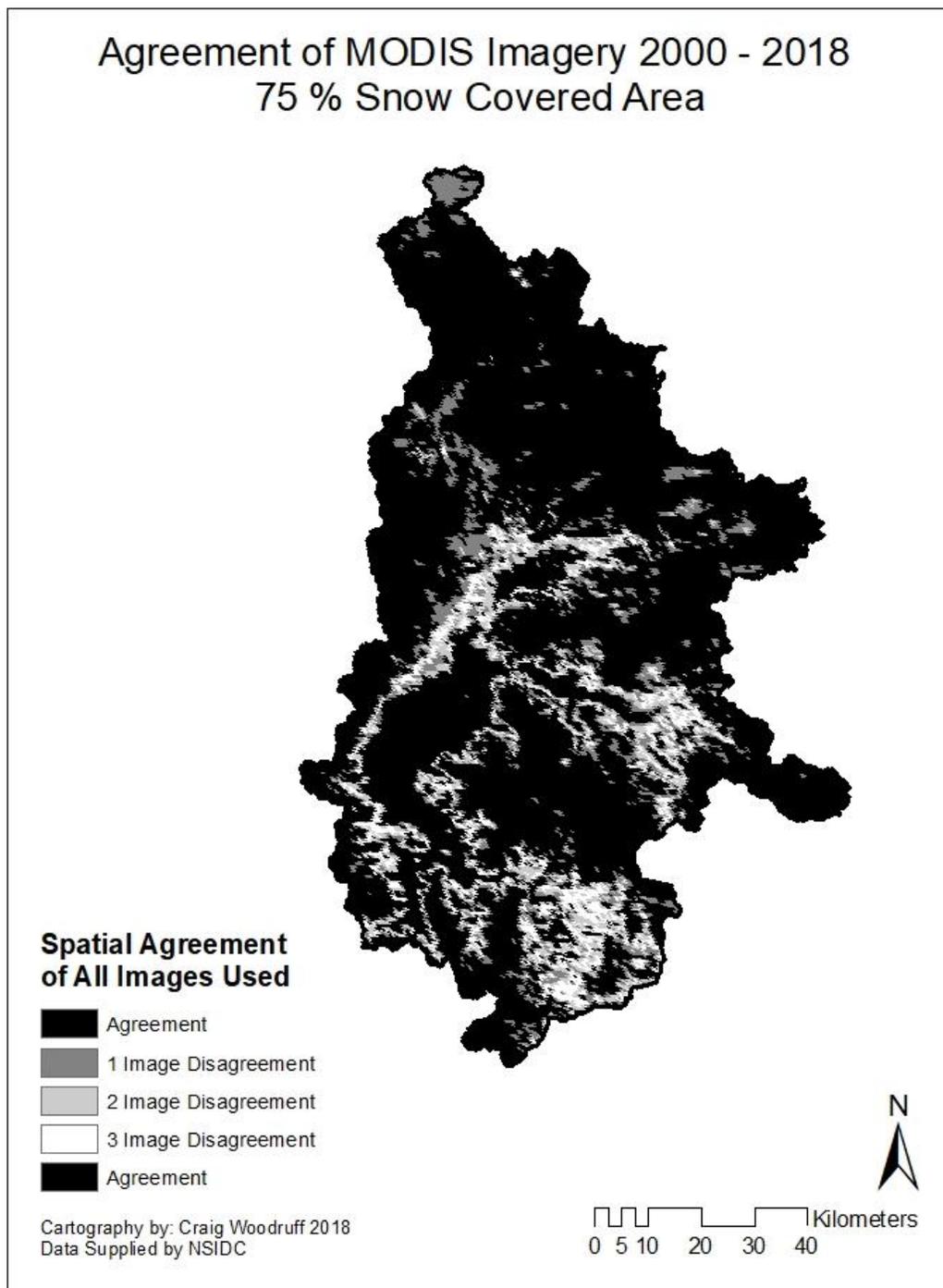


Figure 1.5: Agreement of all the images used at 75% snow covered area.

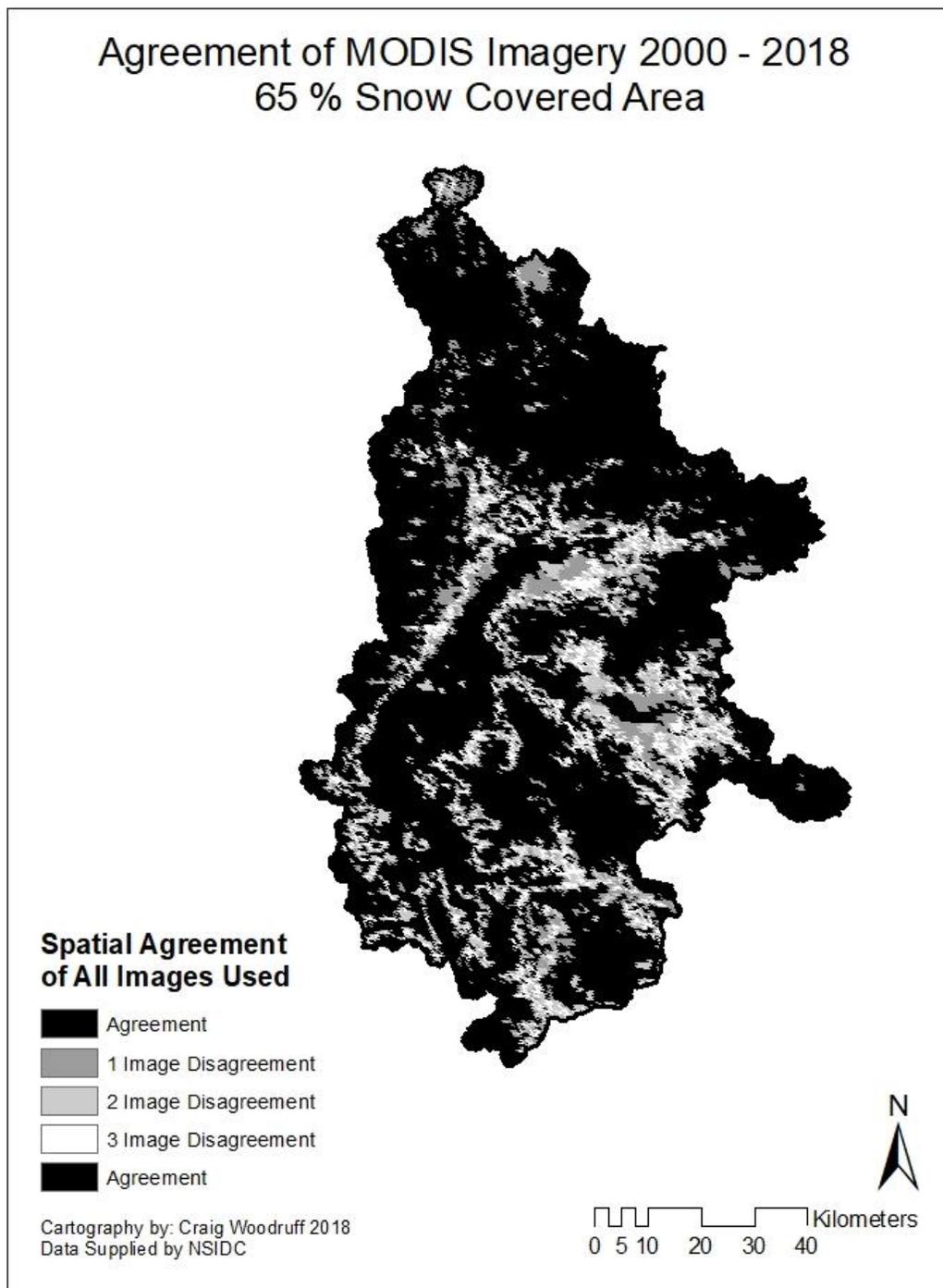


Figure 1.6: Agreement of all the images used at 65% snow covered area.

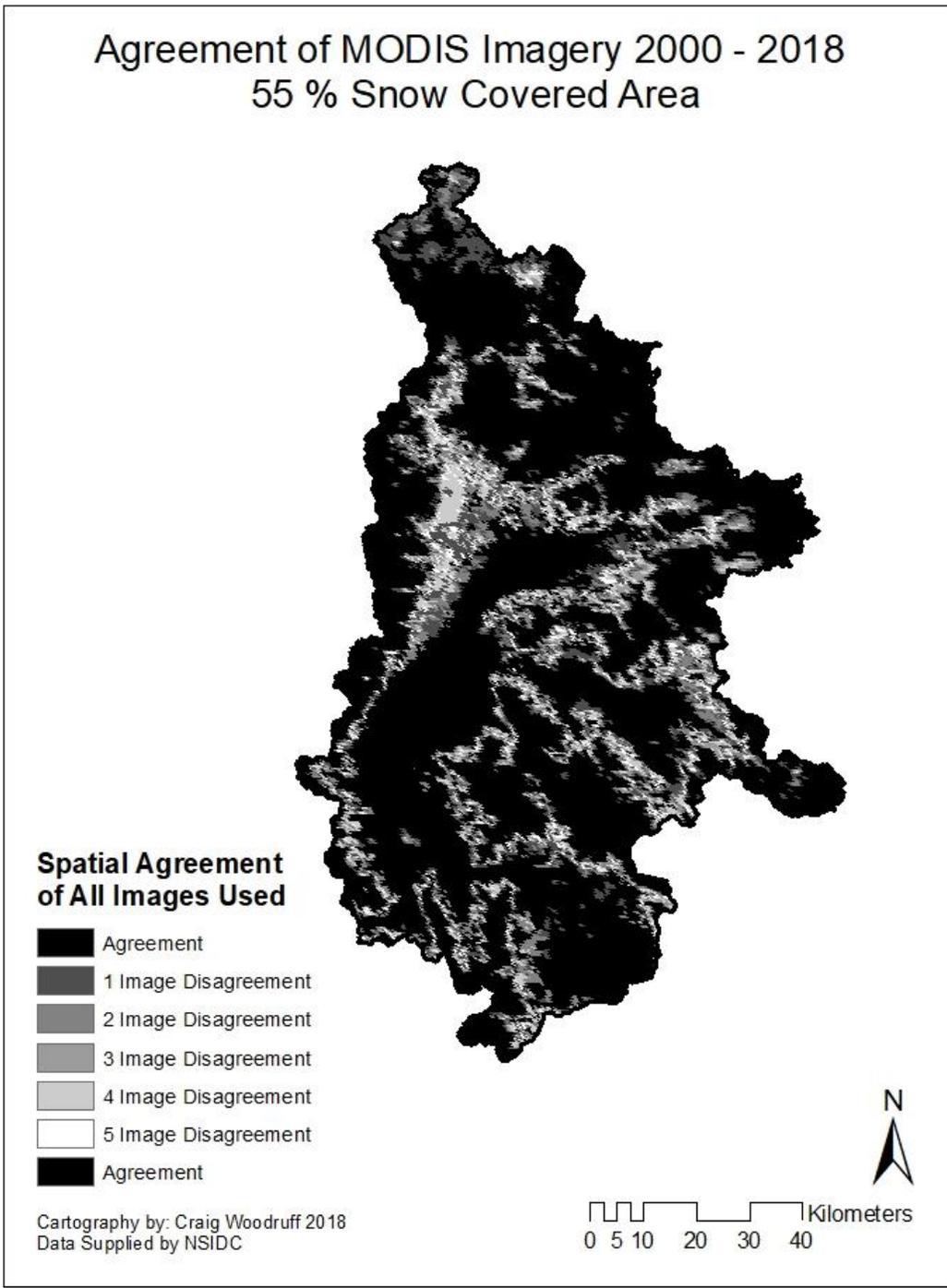


Figure 1.7: Agreement of all the images used at 55% snow covered area.

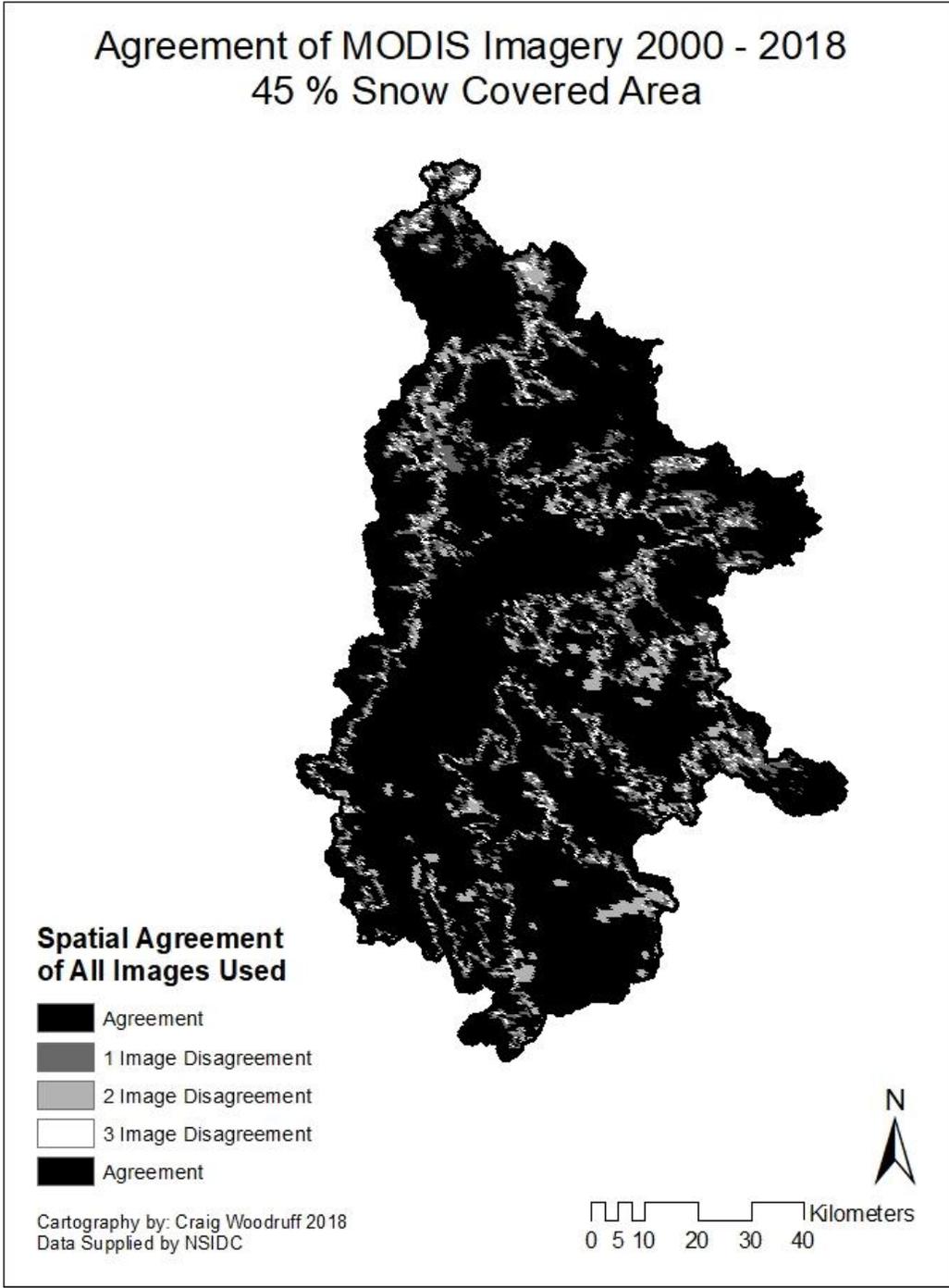


Figure 1.8: Agreement of all the images used at 45% snow covered area.

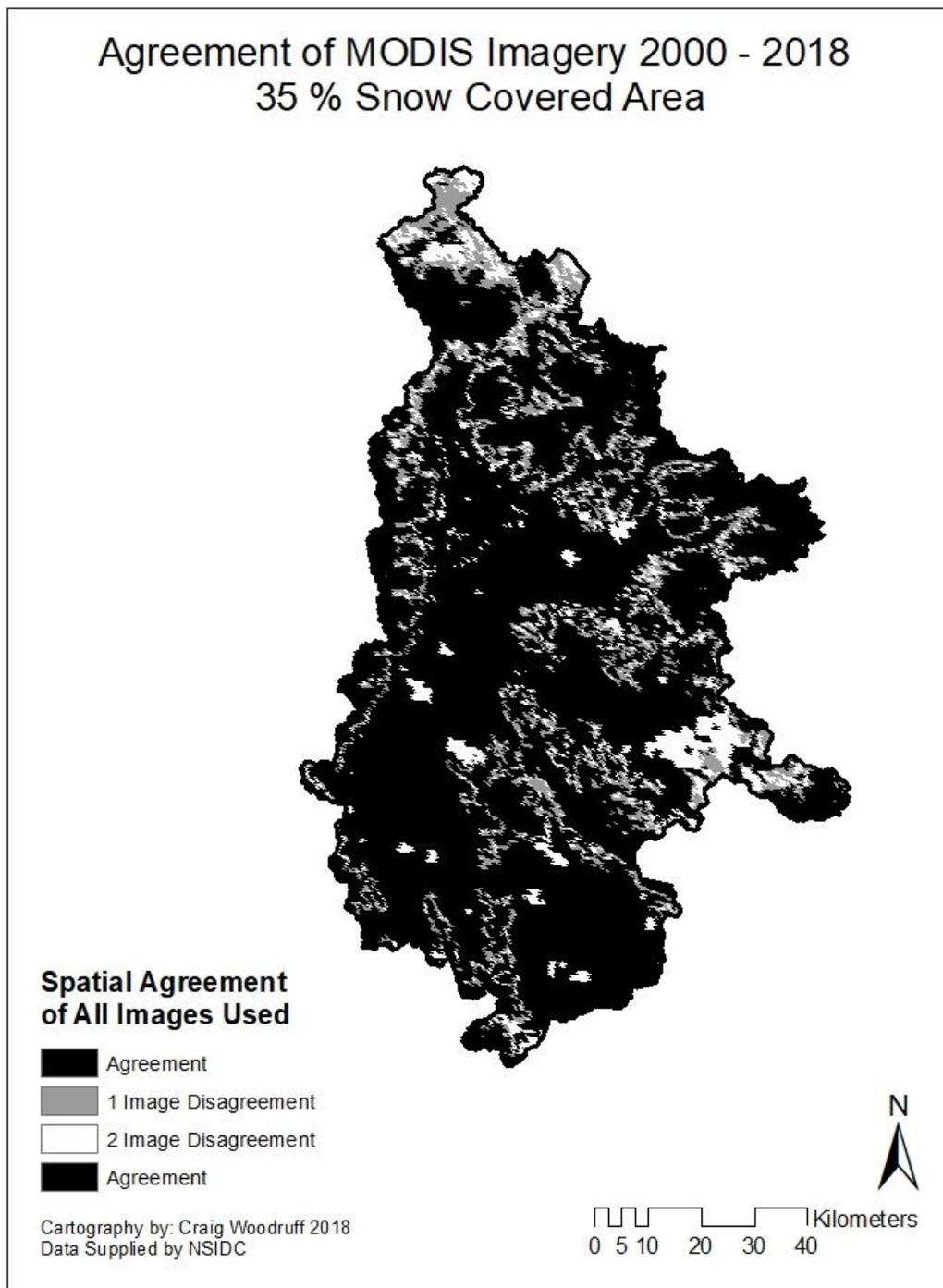


Figure 1.9: Agreement of all the images used at 35% snow covered area.

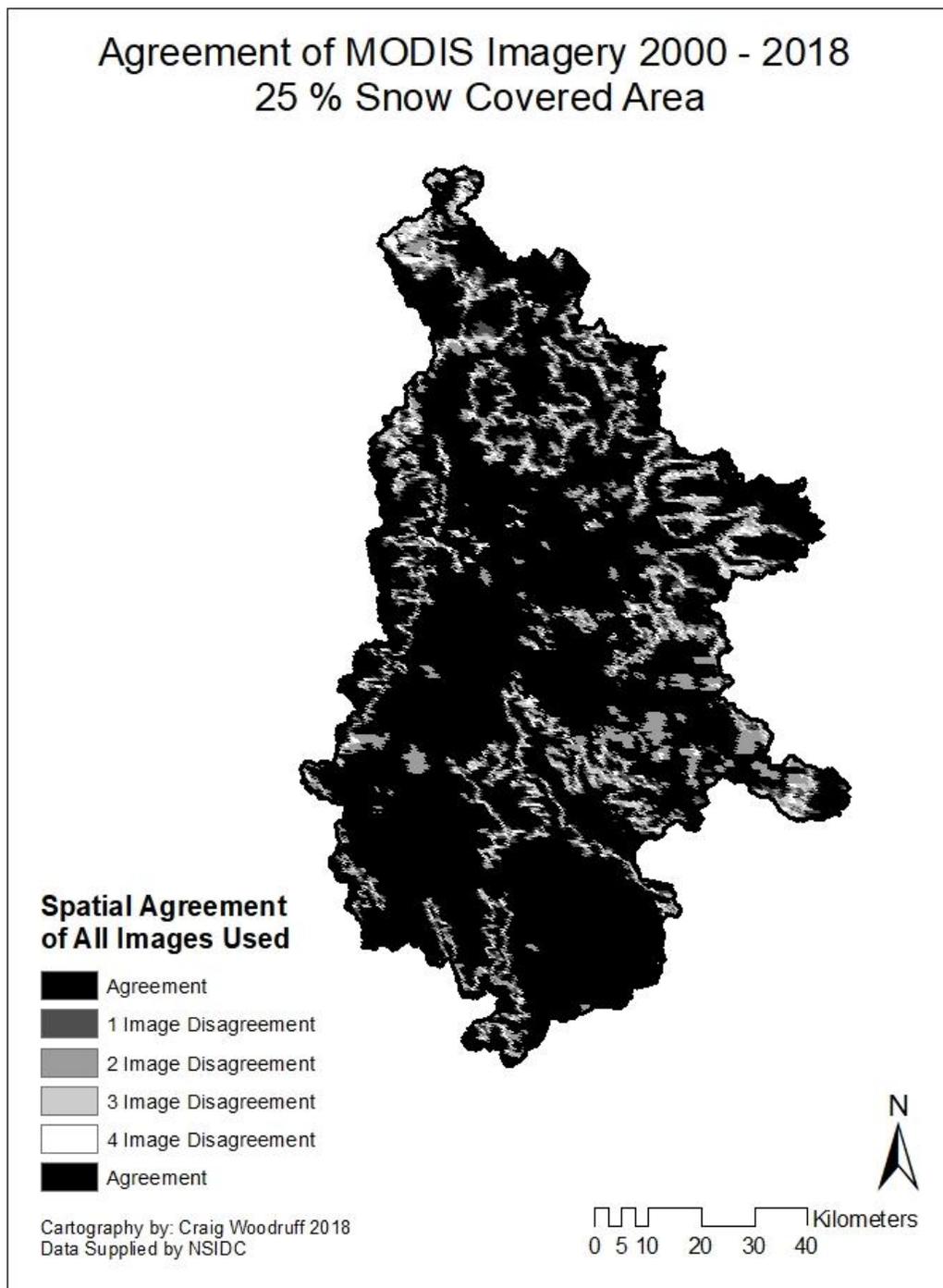


Figure 1.10: Agreement of all the images used at 25% snow covered area.

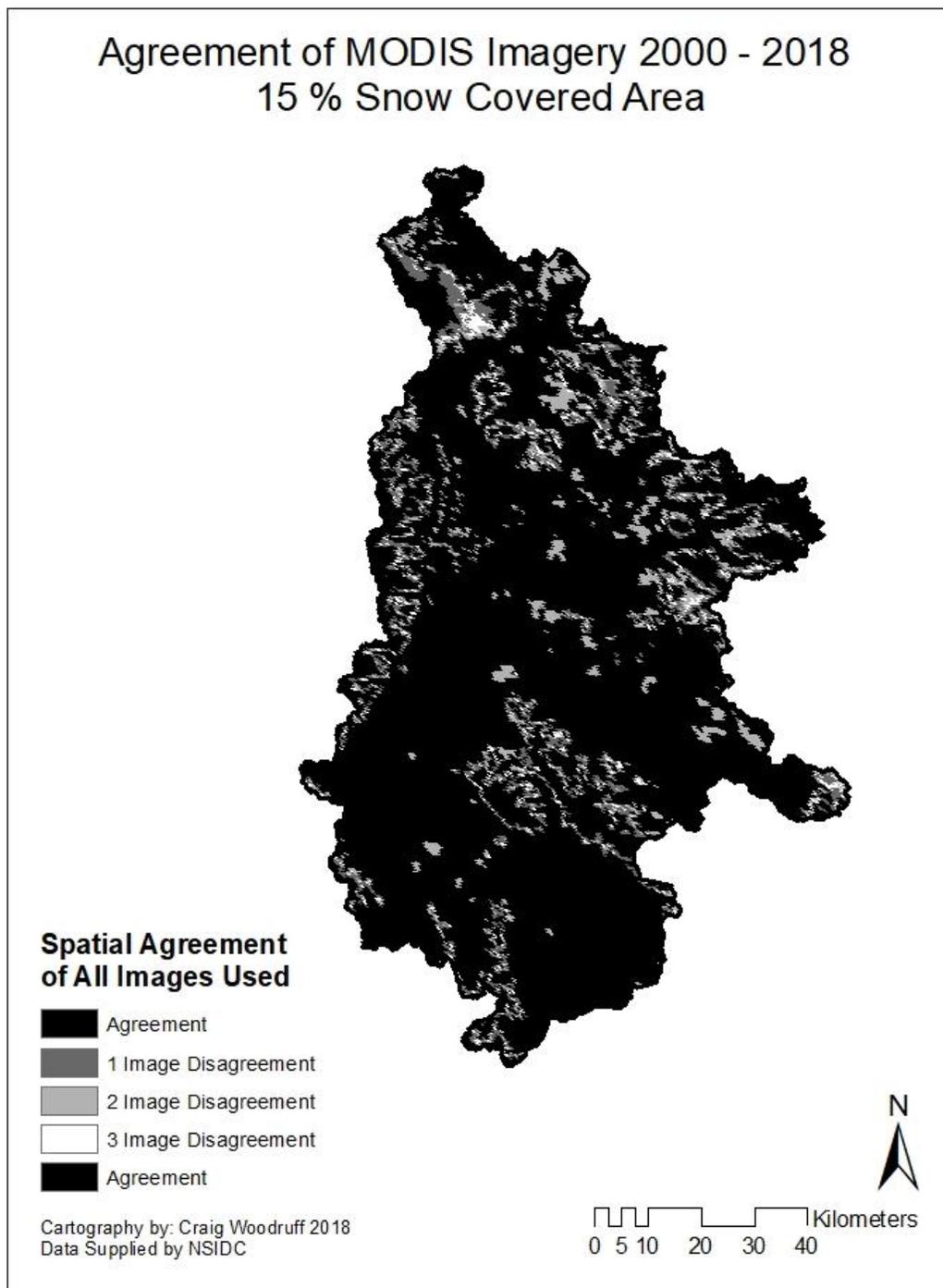


Figure 1.11: Agreement of all the images used at 15% snow covered area.

Chapter 2: Developing Dimensionless Depletion Curves

Introduction

Snow melt is a vital component of available fresh water in the Western United States. Between 39% and 67% of the annual precipitation in the North American West comes as snowfall, and between 50–80 % of the annual flow is contributed by snowmelt [Serreze et al., 1999; Stewart et al., 2004]. The timing and amount of discharge is of the utmost importance especially in areas such as Southern Idaho where legal authority for water management relies heavily on the timing of water availability and storage [Qualls et al., 2013]. Since the 1970's remote sensing has offered spatially distributed information of snow. In 1999 the Moderate Resolution Imaging Spectroradiometer (MODIS) was launched on board the Terra satellite. MODIS offers daily worldwide coverage at a spatial resolution of 500 meters making it ideal for monitoring snow. Unfortunately, snow detection errors and uncertainty at the pixel level are affected by viewing conditions, land cover, the amount of snow, and cloud masking [Riggs et al. 2017]. The MODIS product showed 93% agreement on cloud free days for detection of snow and thus cloud cover poses a persistent challenge as shown by Hall & Riggs [2007].

Cloud removal techniques have been developed for remotely sensed imagery using interpolation methods. Most of these methods use a form of spatial interpolation, temporal interpolation, a combination of these, or complex multistep procedures to produce cloud free imagery [Gao et al., 2010; Krajčič et al., 2014; Painter et al., 2009; Parajka & Blöschl, 2008; Parajka et al., 2010; Gafurov & Bárdossy, 2009; Gafurov et al., 2016]. Significant cloud reduction has been achieved however; real time modeling of the spatial coverage of snow requires the ability to forecast. For this application a depletion curve that can be fit early in the melt season is ideal. An example of the application of a depletion curve is input into the Snowmelt Runoff Model (SRM) [Martinec & Rango 1986; Rango & Martinec, 1994]. The utility of depletion curves can be extended to ungauged catchments using MODIS as well, but to overcome cloud masking interpolation methods are used [Safari et al., 2014; Qui et al., 2014]. Often days that are too cloud covered (>10% by area) are simply thrown out [Steele et al., 2017]. With the magnitude of imagery collected by MODIS, identifying the effects of cloud cover during a melt season on a pixel by pixel basis can be accomplished.

Doing so increases the available data points by reducing the effects of cloud cover while maintaining the original data.

The method developed in this research takes advantage of the large amount of available MODIS imagery, ground based data, and the ideas used in temporal interpolation to add spatial and temporal error bounds to depletion curves. With smaller error bounds and more available days, a depletion curve can be developed. A dimensionless depletion curve can be derived from this, which is applicable to all years. A dimensionless depletion curve is of similar shape for each year altering only specific parameters and is dimensionless with time. To fit a dimensionless depletion curve to any given melt season, only a couple parameters need to be identified. This idea is based on the consistent snowmelt patterns across multiple years. Repeatable patterns of snowmelt have been recorded and accepted for many years [Adams, 1976; König & Sturm, 1998; Luce & Tarboton, 2004; Sturm & Wagner, 2010]. Relying on the year to year pattern, a dimensionless depletion curve which is derived from the original data with reduced error bounds could be a useful tool for real-time snow melt modeling.

Study Area and Data

Study area

Our study area, the Upper Snake Basin, covers an area of 3,465 mi² (8,894 km²) and elevation ranges between 5,799 -13,760 feet (1,737 – 4,194 meters) (Fig. 2.1). The area serves as an important water resource for Idaho agriculture, power generation, in stream fish requirements, as well as municipal, commercial, and industrial uses. The watershed is bordered to the West by the Teton Mountains and includes the hydrologic units known as the Snake Headwaters, Gros Ventre, and Greys-Hobock. Snow seasonally covers the watershed making remote sensing a useful tool for collecting spatial information. The study site also has 11 ground based monitoring stations (SNOTEL sites), which measure a 2 meter by 2 meter snow pillow and each collects a number of physical measurements. The study site land cover is variable. Forest and shrub/grass land account for 47.9 and 45.4 percent of the total area respectively, with the remaining 6.7 percent belonging to agriculture, developed land, and wetlands as calculated using the 2011 National Land Cover Database [Homer et al. 2015].

Data Sources

Snow cover data used in this analysis was collected by MODIS on board the Terra Satellite. The version six data product, MOD10A1 (MODIS Terra snow cover daily L3 global 500 m Sin GRID V006) distributed by the National Snow and Ice Data Center (NSIDC) was used. This product comes with five layers relating to snow extent and albedo. Of the five layers, only the NDSI snow cover was used. Since no temporal filtering was applied there was only a need to use one sensor, MOD10A1. Fractional snow coverage was calculated using an NDSI threshold. Two MODIS swaths were needed to capture the study region, h09v04 and h10v04. All available MODIS imagery from 2000 through 2016 was used in this study.

SNOTEL data distributed by the Natural Resource Conservation Service (NRCS) was also implemented in the algorithm. There are eleven sites in the watershed; however for the purposes of this study snow water equivalent (SWE) measurements from one site were collected: Base Camp, station ID 314. Base Camp is located at one of the lower elevations at 7060 feet, but the snow pillow at this station melts completely before the other ten stations in most years. Figure 2.2 shows peak SWE measured at Base Camp for each year. The largest peak SWE was measured in 2011 with a value of 26.3 inches and the smallest in 2001 at 8.7 inches. Over the seventeen years the average value for peak SWE is 17.2 inches.

Methods

Model development

Analyzing spatial and temporal error bounds on depletion curves requires identification of the days melt occurred and the effect of cloud cover during the melt season. To do this a data reduction algorithm was developed. The algorithm takes advantage of the great deal of spatial information available through MODIS daily imagery, and on a pixel by pixel basis records dates of interest in two separate rasters which can be opened in ArcGIS, python, and many other platforms making it useful for many research applications. Ground based SNOTEL data is incorporated into the algorithm to increase accuracy. For each year of available data, 2000 through 2016 in this study, two output files are created. For example, 2000 would have two rasters of spatial resolution and geolocation equal to the MODIS daily imagery. The two rasters are made up of dates. The “snow” raster contains the last day snow was seen for each pixel, and the “land” raster the first day land is seen for each pixel. Ideally

if cloud cover posed no issue during melt the land raster would be exactly one day larger for each pixel than the snow raster. Since this is not the case, the difference between these rasters is cloud persistence during the time period in which we know melt has occurred for each pixel.

Three basic steps are incorporated in this algorithm: a land, snow, and cloud protocol. To increase processing speed the algorithm works by sub-setting pixels as it works through the large amount of data. Before the algorithm is initiated a starting date is identified. Snow cover is seasonal in this watershed and the timing of peak snowpack varies year to year. In order to best capture the peak snow coverage of the watershed the SNOTEL station with the earliest melt out date over the 17 years (Base Camp) is used to define the start date for each year. Melt out refers to the complete melt of the snow pillow at that station, or the date snow water equivalent is equal to zero. The day of peak snow water equivalent (SWE) for each year was used as the start date. This helps avoid capturing secondary snowfall events after melt has already occurred. Secondary snowfall would be considered a snow fall event which occurs after melt over the peak snow coverage has occurred. The dates of peak SWE used for each year are included in Table 2.1. It is evident from the large spread in the date of peak SWE that the years used vary greatly in the snowpack with peak SWE dates ranging from day 68 to day 113. To coincide with the MODIS nomenclature dates are listed on a scale of 1 to 365 (366 in leap years) where day one is January 1st. Since this date tracking is the same as MODIS it can be used to identify imagery for a given year.

The algorithm extracts the last day of snow and first day of land values by iterating through all the available imagery one year at a time using the day included in the MODIS nomenclature as a guide for processing. The day on a scale of 1 to 365 is included in each image along with the year. For the following description “day” describes the MODIS image on that specific day. A few scenarios can take place during processing. Given a pixel that is identified as snow on the start day the algorithm begins forward processing, known as the snow protocol. Forward processing means the next day of MODIS imagery is analyzed and the next for only the pixels identified as snow on the first day, saving the day that snow was last seen in one raster and the first day land was seen in another until all the pixels have a value for each raster. For the pixels identified as land on the start day, the algorithm back

processes, analyzing the previous day of MODIS imagery until snow is seen for each pixel, known as the land protocol. If a pixel was never snow covered or cloud cover never allowed for snow to be seen the date of last day of snow is saved as day 0. The final scenario is pixels that are identified as cloud on the start day. For these the algorithm begins back processing. As back processing proceeds, if pixels are identified as snow first they are sent to the snow protocol beginning on the day after the start day. For pixels identified as land first they are sent to the land protocol. If a pixel is cloud covered for the entire period from the start day to January 1st the algorithm begins forward processing from the day after the start day. For these pixels if land is seen first the day is only saved to the first day of land raster. The last day of snow raster is saved as a zero meaning it was never seen as snow before it is identified as land.

This method brings in ground based data to inform the timing of the algorithm methodology to efficiently reduce the large amount of imagery available to only two images per year. The output images describe the spatial melt progression through a melting season. The depletion curve for a given year can be derived by fitting the curve to the snow covered area reduction with time. Unlike the other methods for creating a depletion curve, which rely on the limited cloud free days or interpolation methods, this algorithm can be used to create a depletion curve that still incorporates the cloud free days but does not interpolate in between. While the depletion curve in essence is interpolating with time, a substantial amount of raw data is used with spatial and temporal error bounds since the output is a collection of dates of interest.

Identifying Effects the of Cloud Cover

With all of the imagery available for each year condensed down into two images the effect of cloud cover over the melt season on a pixel by pixel basis, can be quantified. The difference between the land and snow raster describes the cloud persistence of each pixel during this time. The number of pixels on a given day that had the “potential to melt” can be extracted. Potential to melt refers to a pixel on a given day of year (DOY) whose last day of snow seen pixel value is greater than the given DOY, but less than the first day of land. The pixel would be considered cloud obscured during its specific period of melt. For each day the number of pixels that had the potential to melt can be quantified. This number acts as a

spatial error bound with temporal meaning. The count of pixels is spatial while the temporal component comes from the potential to melt pixel count, which varies each day.

A depletion curve describing snow covered area with much smaller error bounds based on cloud cover during the melt season on a pixel by pixel basis can be developed. This curve is superior because it is a product of the original data rather than a spatial interpolation of cloud cover on a daily basis, but it still produces daily values with error bounds. Effectively the data is less aggregated than using an interpolation method due to the tighter error bounds. From here a dimensionless depletion curve can be derived.

Developing Dimensionless Depletion Curves

Extending the research conducted by Qualls & Arogundade [2012], a Gaussian Decay was applied to model the melt trend of the watershed. It is important to note that this particular watershed was well described by this function while other watersheds may not be. The curve was fit using a numerical solver to minimize the sum of squared errors by adjusting two parameters: the rate of melt and date of 50% snow covered area. Before the curve can be fit fractional snow covered area (FSC) and the error bounds must be defined. The first day of land raster was used to calculate the fractional snow covered area because it defines the day there is confidence the pixel has melted. Fractional snow covered area is calculated according to Equation 2.1. The error bounds were defined using the number of pixels with the potential to melt on a given day. This value is calculated in the same way as FSC where the numerator is the number of pixels with the potential to melt.

$$FSC = \frac{\text{Snow Covered Pixels}}{\text{Total Pixels}} \quad \text{Equation 2.1}$$

The error bounds were added to the FSC. Since the first day land is seen raster was used, the errors are only associated with pixels that have not melted on each DOY. To fit the curves, an error was only assessed if the curve fell outside the region defined by FSC area plus the fractional representation of the pixels with the potential to melt. With this approach a spatial and temporal error bound is added, as the number of pixels with the potential to melt fluctuates day to day. The cloud free days a still represented using this method. To create the dimensionless depletion curve a simple equation is applied (Equation 2.2). A z score is calculated at each snow covered area (SCA). The X_{50} value is the DOY 50% SCA is achieved and σ is the rate of melt or the steepness of the Gaussian decay function. Since

each depletion curve for any given year is numerical fit to minimize the sum of squares error by altering X_{50} and σ the z value is quickly calculated.

$$z = \frac{X_{50} - DOY}{\sigma} \quad \text{Equation 2.2}$$

Results

Similarly to a temporal filtering method for cloud removal the algorithm effectively reduces the impact of cloud cover. Applying these cloud covered pixels with the potential to melt as an error bound increases the confidence in the fit of a depletion curve. As seen in Figure 2.3 the depletion curve agrees very well with the measured snow covered area and potential to melt error bounds. The depletion curve, on a given DOY, should ideally fall on top of the snow covered area point or in between the snow covered area and potential to melt error bound. The significant portion of the melt from 75% to 25% snow covered area is represented well with the depletion curve.

Like a temporal interpolation method applied to cloud covered imagery, this method effectively reduces cloud cover. MODIS daily imagery ranged in average cloud cover over the same time period as the depletion curve from 56.89% to 65.59% of the total watershed area. The algorithm reduced the average daily cloud cover to a range of 1.36% to 3.83% of the total area. The maximum cloud coverage was also reduced from 100% in the MODIS daily imagery to a range of 8.06% to 20.97% across all 17 years. Reducing the effects of cloud cover puts more confidence in the depletion curve. The fit of a depletion curve can be quantified using the root mean squared error (RMSE) across all the years. Using the same depletion, but altering only the timing of 50% melt (when snow covered area is equal to 50% of the area) and the rate of melt produces a small RMSE over all 17 years. The RMSE is lowest for 2013 with a value of 0.0084 and highest for 2010 at 0.0326.

The issue of cloud masking is decreased using this method. To illustrate the ability of this method to reduce the influence of cloud cover 2013 was analyzed. Since the melt season is the only important time period for this analysis only days 50 through 200 were compared. The MODIS daily imagery had a far larger spread in days with cloud masking. Only 41 days had less than 20% cloud covered area. Using the algorithm all 151 days had 20% or less cloud covered area with 136 of those days at 5% or less cloud covered area.

The depletion curve fit to 2013 (Fig. 2.3) is based on the dimensionless depletion curve by altering X_{50} and σ to minimize the sum of squares error. A dimensionless depletion

curve is shown in Figure 2.4. The curve crosses the y axis at the value of X_{50} , a z value of 0 as calculated by Equation 2.2. By only altering the two parameters this curve is easily fit to any year, even with the smaller error bounds. The application of the dimensionless curve is based on the repeatable pattern of melt.

Figures 2.5 and 2.6 both show the fitted depletion curves for the highest and lowest snow water equivalent years 2001 and 2011, respectively. The snow water equivalent for 2013 is slightly below the average peak SWE of 17.2 inches over the 17 years at 15.5 inches. The shape of the curve appears equal to the 2013 curve however the timing is shifted greatly in 2001 and 2011. The utility of dimensionless depletion curve (Fig. 2.4) is seen in these differing years.

While it does seem odd that the cloud covered area points exceeds 1 in each of the plots, the reasoning is sound. On the dates this occurs, the pixels contributing to this are cloud covered for an extended period of time, but were last seen as snow. In this particular watershed the cloud persistence is greatest early in the melt season. Since this method aims to track the influence of cloud cover rather than interpolate the cloud covered pixels a value greater than 100% of the area is permissible. The depletion curve does follow the melting trend the rest of each year indicating the error associated with this extended cloud coverage is limited. For years 2001 and 2011 the curve struggles to fit the beginning of melt, but closely tracks from 75% snow covered area to 0%. The nearly horizontal number of points for 2011 around day 170 may be attributed to cold snap or a snowfall event, however the date is late in the year. Speculation aside the curve tracks the melting of snow after the brief disagreement only slightly overestimating the day of complete melt. Comparison of the three plots shows the variation in the timing of the 50% melt data for each year and the difference in the rate of melt or decline. The timing of 50% melt was day 118, 162, and 128 for 2001, 2011, and 2013 respectively. The rate of melt in terms of days was 23, 24, and 21 for 2001, 2011, and 2013 respectively. The rate does not vary as much as the timing of 50% melt but a small change in rate extends the melting period substantially. For 2001 and 2011 with only a difference of 1 in rate the melting period between 75% snow covered area to 25% is extended by 2 days.

The dimensionless depletion curve (Fig. 2.4) relies on snowmelt patterns. For this particular watershed with a hypsometric curve is similar to a Gaussian decay curve making

it a good representation of melt. As seen in Table 2.2, the depletion curve fits each year independent of snowpack. As seen in Figure 2.2 the peak SWE at the Base Camp station varied greatly over the 17 years in this analysis. The smaller error bounds defined through the algorithm add confidence to the validity of this method for modeling snow covered area across multiple years.

Conclusion

Modeling snow melt is a vital component of water management. With the help of remote sensing daily coverage of snow covered area can be calculated although cloud coverage is a persistent issue. Although snow covered area does not explicitly define the snow water equivalent, the spatial resolution of MODIS imagery at 500 meters is applicable to complex terrain and the daily coverage is ideal as an input for models such as SRM. The method outlined in this research adds spatial and temporal error bounds to depletion curves. These error bounds establish more confidence in the use of depletion curves, namely dimensionless depletion curves. Over the 17 years of imagery a dimensionless depletion curve showed agreement with the measured snow covered area independent of the highly variable peak snow water equivalent. The modelled curve also handled extreme differences in the timing and duration of melting periods with 50% melt varying between day 162 and 118 for the most significantly different years, 2001 and 2011.

The overall benefit of this methodology for developing depletion curves is its applicability to all years independent of snowpack. With only one equation for a depletion curve and altering only two parameters, all seventeen years could be represented. In other methods cloud cover poses a substantial issue forcing data loss or interpolation, however this method decreases the effects of cloud cover while maintaining the original data. The dimensionless curve relies on the year to year spatial pattern of snow melt and simply shifts the timing and rate of the melt to describe each year. Using remote sensing is not a new method for calculating snow covered area for deterministic models such as SRM as applied by Baumgartner et al. [1987]. It has long been established as a method, but this new method increases available data, and may provide new opportunity with higher confidence for deriving depletion curves in gauged and ungauged catchments. This can be used in modeling scenarios real time and historic. It can be applied in climate change modeling since the curve is independent of snow pack and relies only on two parameters. In other watersheds a

different depletion curve may be necessary to describe the melt patterns, which may complicate the procedure but the year to year patterns associated with snowmelt in this watershed lend themselves to a simple approach.

References

- Adams, W. P. (1976). Areal differentiation of snow cover in east central Ontario, *Water Resources Research*, 12(6), 1226–1234, doi:10.1029/WR012i006p01226.
- Baumgartner, M., Seidel, K., & Martinec, J. (1987). Toward Snowmelt Runoff Forecast Based on Multisensor Remote-Sensing Information. *Geoscience and Remote Sensing, IEEE Transactions on, GE-25*(6), 746-750.
- Gafurov, A., & Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrology and Earth System Sciences*, 13(7), 1361-1373.
- Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., & Merz, B. (2016). MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environmental Earth Sciences*, 75(14), 1-15. doi:10.1007/s12665-016-5869-x.
- Gao, Y., Xie, H., Lu, N., Yao, T., & Liang, T. (2010). Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra–Aqua MODIS and Aqua AMSR-E measurements. *Journal of Hydrology*, 385(1), 23-35. doi:10.1016/j.jhydrol.2010.01.022.
- Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12), 1534-1547. doi:10.1002/hyp.6715.
- Hall, D. K. & Riggs, G. A. (2016). *MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6*. [NDSI snow cover]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: <https://doi.org/10.5067/MODIS/MOD10A1.006>. [January 2017 - March 2018].
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., & Megown, K., (2015), [Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information](#). *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354.
- König, M. & Sturm, M. (1998). Mapping snow distribution in the Alaskan Arctic using aerial photography and topographic relationships, *Water Resour. Res.*, 34(12), 3471–3483, doi:10.1029/98WR02514.
- Krajčí, P., Holko, L., Perdigão, R. A. P. & Parajka, J. (2014). Estimation of regional snowline elevation (RSLE) from MODIS images for seasonally snow covered mountain basins. *Journal of Hydrology*, 519(PB), 1769-1778. doi:10.1016/j.jhydrol.2014.08.064.

- Luce, C. H., & Tarboton, D. G. (2004). The application of depletion curves for parameterization of subgrid variability of snow. *Hydrological Processes*, 18(8), 1409-1422. doi:10.1002/hyp.1420.
- Martinec, J., & Rango, A. (1986). Parameter values for snowmelt runoff modelling. *Journal of Hydrology*, 84(3), pp.197–219. doi:10.1016/0022-1694(86)90123-X.
- Painter, T. H., Rittger, K., Mckenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4), 868-879. doi:10.1016/j.rse.2009.01.001.
- Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. *Water Resources Research*, 44(3), N/a. doi:10.1029/2007WR006204.
- Parajka, J., Pepe, M., Rampini, A., Rossi, S., & Blöschl, G. (2010). A regional snow-line method for estimating snow cover from MODIS during cloud cover. *Journal of Hydrology*, 381(3), 203-212. doi:10.1016/j.jhydrol.2009.11.042.
- Qualls, R. J., & Arogundade, A. B. (2012). Modeling snowmelt runoff under climate change scenarios using MODIS-based snow cover products. In N.B. Chang & Y. Hong (Eds.), *Multiscale hydrologic remote sensing: Perspectives and applications*. New York, NY: CRC Press, pp. 213–249.
- Qualls, R. J., Taylor, R. G., Hamilton, J., & Arogundade, A. B. (2013) Climate change opportunities for Idaho's irrigation supply and deliveries. *Journal of Natural Resources Policy Research*, 5(2-3), 91-105, DOI: 10.1080/19390459.2013.811856.
- Qiu, L., You, J., Qiao, F., & Peng, D. (2014). Simulation of snowmelt runoff in ungauged basins based on MODIS: A case study in the Lhasa River basin. *Stochastic Environmental Research and Risk Assessment*, 28(6), 1577-1585. doi:10.1007/s00477-013-0837-4.
- Rango, A., & Martinec, J. (1994). MODEL ACCURACY IN SNOWMELT-RUNOFF FORECASTS EXTENDING FROM 1 TO 20 DAYS 1. *JAWRA Journal of the American Water Resources Association*, 30(3), 463-470. doi:10.1111/j.1752-1688.1994.tb03305.x.
- Riggs, G.A., Hall, D. K., & Roman, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, 9(2), pp.765–777. doi:10.5194/essd-9-765-2017.
- Safari Shad, M., Habibnejad Roshan, M., & Ildoromi, A. (2014). Integration of the MODIS Snow Cover Produced Into Snowmelt Runoff Modeling. *Journal of the Indian Society of Remote Sensing*, 42(1), 107-117. doi:10.1007/s12524-013-0279-y.

- Serreze, M.C., Clark, M. P., Armstrong, R. L., McGuinness, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research* 35(7), 2145–2160. doi:10.1029/1999WR900090.
- Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A., & Bleiweiss, M. (2017). Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the Rio Grande headwaters. *International Journal of Applied Earth Observations and Geoinformation*, 63, 234-243. doi:10.1016/j.jag.2017.08.007.
- Stewart, I., Cayan, T., & Dettinger, D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a `Business as Usual' Climate Change Scenario. *Climatic Change*, 62(1), 217-232. doi:10.1023/B:CLIM.0000013702.22656.e8.
- Sturm, M., & Wagner, A. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*, 46(12), N/a. doi:10.1029/2010WR009434.

Tables

*Table 2.1: Dates of peak snow water equivalent at Base Camp SNOTEL station.
Start Day (Maximum*

<i>Year</i>	<i>SWE at Base Camp)</i>
2000	80
2001	79
2002	87
2003	87
2004	68
2005	95
2006	92
2007	69
2008	100
2009	93
2010	100
2011	113
2012	80
2013	111
2014	97
2015	85
2016	89

Table 2.2: Algorithm efficiency for reducing cloud cover and fitting a dimensionless depletion curve

<i>Year</i>	<i>Average Daily Imagery Cloud Cover</i>	<i>Average Algorithm Derived Cloud Cover</i>	<i>Algorithm Maximum Cloud Coverage</i>	<i>Depletion Curve RMSE</i>
2000	56.89	3.65	19.65	0.0182
2001	58.36	1.88	10.34	0.0122
2002	60.11	2.85	17.59	0.0152
2003	60.98	2.26	11.21	0.0183
2004	57.89	1.59	10.80	0.0273
2005	59.46	2.54	13.15	0.0231
2006	58.51	1.36	8.82	0.0172
2007	60.93	1.66	8.06	0.0235
2008	61.10	2.25	15.82	0.0158
2009	62.46	2.59	16.52	0.0157
2010	64.25	3.83	20.97	0.0326
2011	65.59	2.34	11.76	0.0145
2012	58.08	1.51	8.50	0.0219
2013	60.43	1.53	10.67	0.0084
2014	62.46	1.55	9.51	0.0116
2015	58.18	2.94	12.88	0.0254
2016	61.32	2.05	14.68	0.0146

Figures

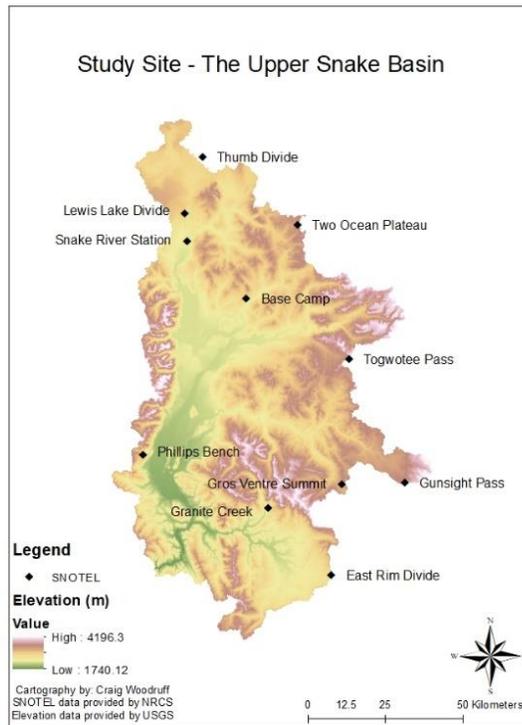


Figure 2.1: Study site, Upper Snake Basin and associated SNOTEL stations.

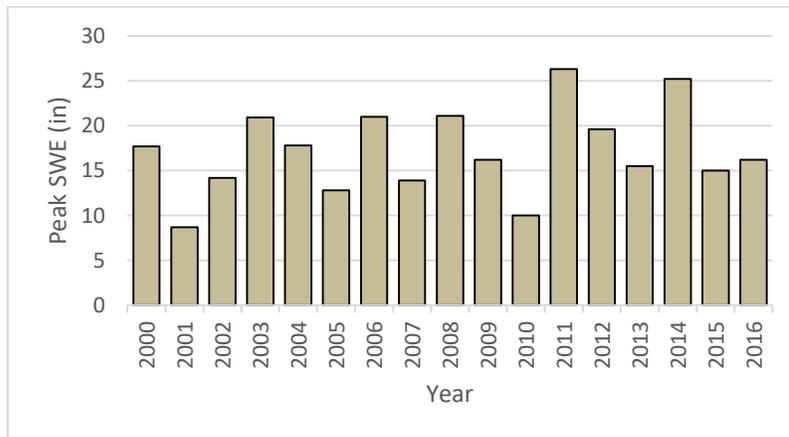


Figure 2.2: Peak snow water equivalent values for Base Camp SNOTEL station.

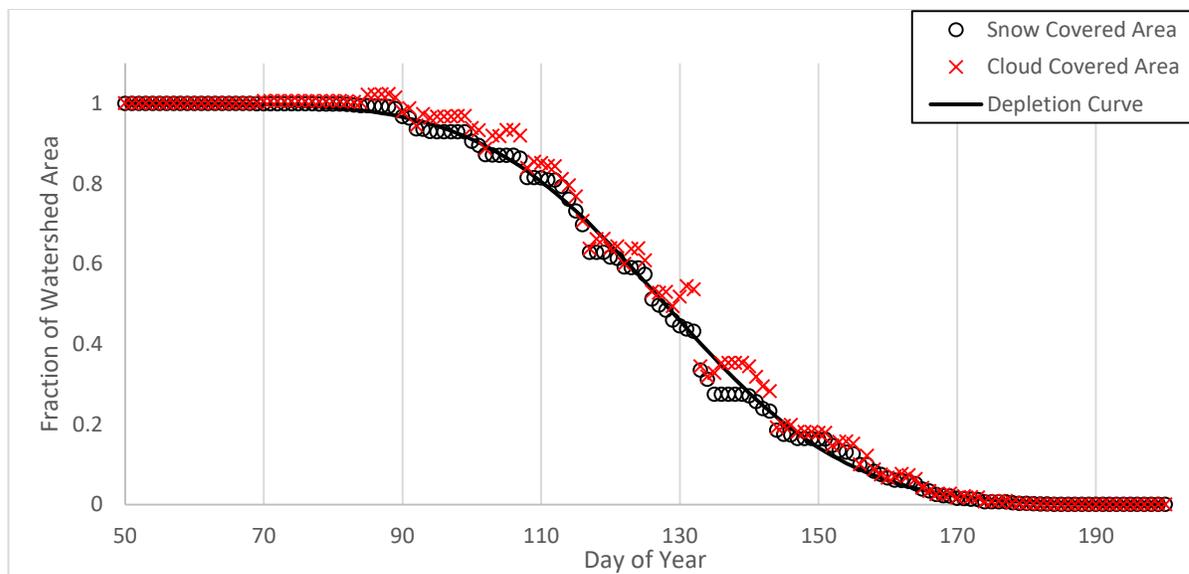


Figure 2.3: Fitted dimensionless depletion curve for the 2013 melt season. Snow covered area is indicated by the hollow circles, cloud covered area by the "X", and the dimensionless depletion curve by the solid line.

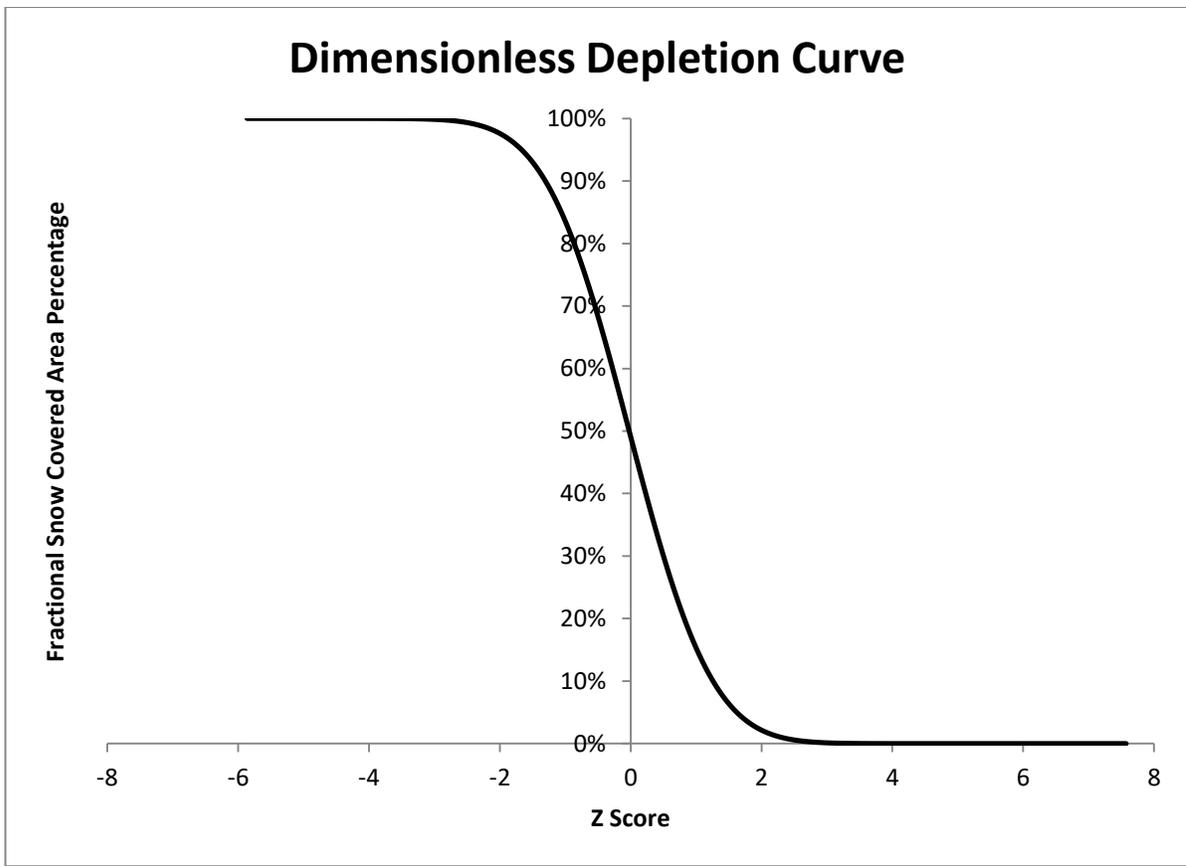


Figure 2.4: Dimensionless depletion curve which is applied to each year.

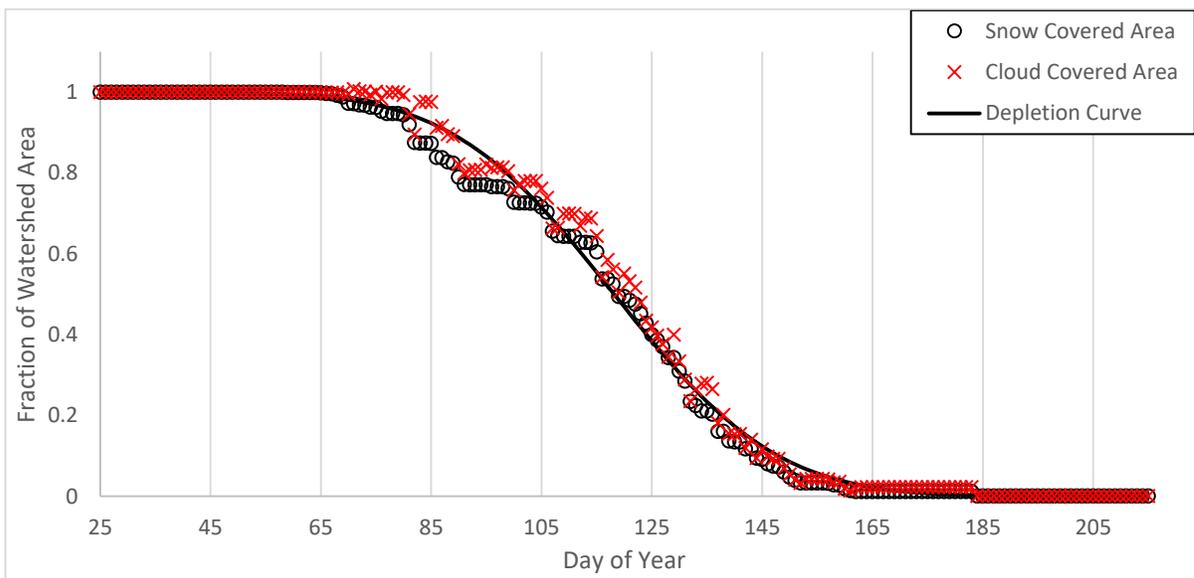


Figure 2.5: Fitted dimensionless depletion curve for the 2001 melt season. Snow covered area is indicated by the hollow circles, cloud covered area by the "X", and the dimensionless depletion curve by the solid line.

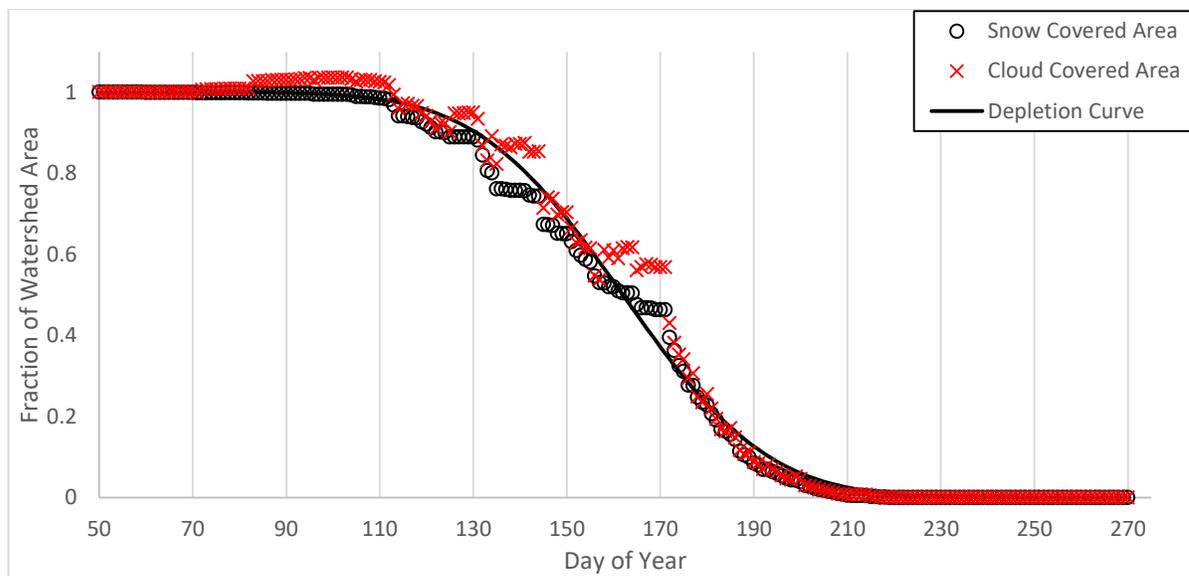


Figure 2.6: Fitted dimensionless depletion curve for the 2011 melt season. Snow covered area is indicated by the hollow circles, cloud covered area by the "X", and the dimensionless depletion curve by the solid line.

Chapter 3: Derivation and Application of a Spatial Model (PCA)

Introduction

Snow melt is a vital component of available fresh water in the Western United States. Between 39% and 67% of the annual precipitation in the North American West comes as snowfall, and between 50–80 % of the annual flow is contributed by snowmelt [Serreze et al., 1999; Stewart et al., 2004]. Remote sensing of snow coverage has been available since the 1970's. Daily imagery at a spatial resolution of 1.1 kilometers at nadir has been available since 1979 with the Advanced Very High Resolution Radiometer AVHRR [Jensen, 2016]. Although this sensor had exceptional temporal resolution, the spatial resolution of 1.1 kilometers may be too large for small basins [Rango, 1993]. The launch of the MODerate Resolution Imaging Spectoradiometer (MODIS) in 1999 on board the Terra satellite provides daily worldwide coverage at a spatial resolution of 500 meters. MODIS imagery snow detection is subject to errors and uncertainty at the pixel level, which is affected by viewing conditions, land cover, the amount of snow, and cloud masking [Riggs et al. 2017]. Of these errors and uncertainty cloud cover poses the most persistent problem. For the watershed in this research during the annual snowmelt period, lasting approximately 150 days, only 28 days on average had less than 10% cloud cover. In Austria an average cloud coverage of 63% over mountainous basins was observed during the snowmelt period [Parajka & Blöschl, 2006]. Cloud cover poses a data availability issue while other errors at the pixel level are an issue of accuracy. On cloud free days the MODIS product showed 93% agreement with ground measurements, therefore cloud cover is a persistent challenge [Hall & Riggs, 2007]. Remotely sensed imagery can be incorporated into deterministic models such as the Snowmelt Runoff Model (SRM) to generate depletion curves which describe how the area in a watershed changes from snow cover to exposed land with time [Martinec & Rango, 1986]. MODIS imagery has been used to increase the accuracy of SRM, however cloud cover is such a persistent issue that the imagery is interpolated or not used [Qui L. et al., 2014; Safari et al., 2014; Steele et al., 2017]. Interpolation of MODIS imagery has been an area of particular interest to produce cloud free imagery.

Interpolation methods can be divided into three main categories: spatial, temporal, and spatiotemporal methods. Spatial methods have been developed based on adjacent pixels and more complex methods. A regional snow line method has been applied by many to reduce cloud cover. The regional snow line uses visible pixels on a given day along with elevation to interpolate cloud covered pixels [Parajka et al., 2010; Krajčí et al., 2014]. This method assumes elevation to be the only influence of snowmelt. The relationship between elevation and snowmelt is only a portion of the topographic effects that drive snowmelt, where slope and aspect play an important role [Hock, 2003]. The regional snowline method acknowledges this by defining an elevation range considered to be partial snow [Parajka et al., 2010]. Krajčí et al. [2014] found a decrease in accuracy during the month of April using the regional snow line attributing it to patchy snow melt during the end of the melt season; this was also discussed in Parajka & Blöschl [2012]. Use of neighboring pixels has also been incorporated as a spatial interpolation method [Gafurov & Bárdossy, 2009]. This method relies on the visibility of pixels next to or near cloud obscured pixels.

Temporal filtering methods have also been developed. The MODIS sensor is flown aboard both the Aqua and Terra Satellites. The Terra satellite has a morning overpass and the Aqua an afternoon overpass time. This combination produces two images with different cloud cover location in one day. Cloud cover reduction by combining the two has achieved 39.9% cloud cover removal over Europe [Dietz et al., 2012]. Sensor combinations of MODIS and Aqua Advanced Microwave Scanning Radiometer (AMSR-E) have also been proposed to produce cloud free imagery [Gao et al., 2010]. Unlike the visible product from MODIS, AMSR-E uses microwave observations and measures snow water equivalent, which is not hindered by cloud cover, however it has a coarse spatial resolution of 25 x 25 kilometers [Jensen, 2016]. Combining the two MODIS sensors, one on Terra and the other on Aqua, was found to reduce average cloud cover by 12.2% and 7.2% for Aqua and Terra respectively. Other combinations use imagery from before and after the image on a given date to interpolate whether a cloud covered pixel is snow or land [Parajka & Blöschl, 2008; Wang & Xie, 2009]. Cloud cover can be greatly reduced using a combination of temporal methods, but for real time snowmelt forecasting these methods require imagery from before and after the date in question, delaying data availability for modeling scenarios.

Spatiotemporal methods use combinations of spatial and temporal interpolation, often in sequential steps to achieve substantial cloud reduction, and in many cases cloud free imagery. Use a chain of decisions to produce cloud free imagery. Gafurov & Bárdossy [2009] proposed a six step process implementing a temporal method in steps one and two, and spatial interpolation in steps three and four. The final step uses a threshold day indicating over a 365 day time series to determine whether the pixel is snow covered. A similar methodology was derived by Ronco & De Michele, [2014] implementing only 5 steps. In both of these a probability of snow is defined in the last step based on a “snow season”, the multi-year average time period of snow cover over a pixel. Probability of snow cover has been an area of particular interest in recent years. Li et al., [2017] used a two-step method where the second step is defined by the probability of snow cover based on spatial probability (a regional snowline style approach) and temporal probability incorporating a weighting scheme based on prior and subsequent images where a pixel is visible. More complex methods incorporate image analysis methods such as the Hidden Markov Random Field to create a statistical measure to remove cloud [Huang et al., 2018]. Simpler spatial and temporal combinations have also been developed, but do not achieve cloud free imagery on a daily time step [Parajka & Blöschl, 2008].

The incorporation of conditional probability of snow cover has been increasing in popularity among research groups. The idea proposed by Gafurov & Bárdossy, [2009] and expanded in the development of the MODSNOW-tool, a methodology for daily cloud free imagery, incorporates information from multiple water years [Gafurov et al., 2016]. Monthly probabilities of snow coverage are developed based on historic MODIS imagery as well as a conditional probability relating pixels. The conditional probability relates pixels throughout the area by a probability of snow cover. The probability of one pixel being snow covered can be determined by other pixels in the watershed. They use this concept to remove cloud cover. If a pixel is visible and another is cloud covered the probability can be used to identify whether the cloud covered pixel is snow or land. A final step in the MODSNOW-tool uses the snow onset and snowmelt days in previous years for each pixel to identify whether it is snow covered on a given day. This method has been expanded on to include meteorological data such as temperature, snow depth, and precipitation interpolated using a kriging interpolation to create spatially distributed grids [Dong & Menzel, 2016]. These are

complex multistep methodologies that incorporate multiple year data. The incorporation of multi-year data is weakened by outlier snowpack years especially in the use of monthly probability and snow onset and snowmelt days. The watershed used in this study saw a year with melt out dates a full month later than many other years and melt out dates occurring in three separate months over a 19 year study period, based on SNOTEL station data. The use of interpolated meteorological data can induce errors through the interpolation method and requires a number of gauging stations in the watershed. These methods produce cloud free imagery, but may be prone to error in early melt years and late melt years. Earlier snowmelt has been shown to relate to earlier discharge [Tan et al., 2011]. For modeling purposes discharge timing is critical for water management especially in states such as Idaho whose legal authority for water management relies on the timing of water availability [Qualls et al., 2013]. Modeling accuracy in these years is particularly important, yet available probability methods are likely to produce the largest errors in these years.

While the conditional probability approach is effective it overlooks one consistency in snowmelt. It has long been accepted that snow melts in the same pattern across years independent of snow depth [Adams, 1976; König & Sturm, 1998; Luce & Tarboton, 2004; Sturm & Wagner, 2010]. Although the timing of melt does not always fall on the same day, or last the same duration it follows a similar distribution. The spatial pattern of snow melt is consistent even though the timing may shift. It is precisely this shift that the probability approaches overlook. The pattern of snowmelt along with variable snow duration and onset was observed in the Tianshan Mountains over the relatively short period of 2001 through 2006 [Wang & Xie, 2009]. Thus far no research has applied this idea as the basis for developing a method for cloud removal. MODSCAG, a model used to characterize fraction of snow cover at the subpixel level, models the overall shape of snow with precision of 93.7% to 100% for correctly identifying snow [Painter et al., 2009]. This model relies on remotely sensed spectral information to identify snow, but does not use spatial information from multiple years to inform future scenarios. The model developed in this research uses spatial patterns observed over multiple years to inform cloud removal by deriving a single map describing snowmelt patterns. This map is applicable for cloud removal, and is equally valuable for simulating the impact of climate change on snowcover variability.

Study Area and Data Sources

Study area

Our study area, the Upper Snake River Basin, lies in the northwestern United States (Fig. 3.1), covers an area of 3,465 mi² (8,894 km²), and its elevation ranges between 5,799 - 13,760 feet (1,737 – 4,194 meters) above mean sea level (amsl). The area serves as an important water resource for Idaho agriculture, power generation, in stream fish requirements, as well as municipal, commercial, and industrial uses. The watershed is bordered to the West by the Teton Mountains and includes the US Geological Survey hydrologic units known as the Snake Headwaters, Gros Ventre, and Greys-Hobock. Snow seasonally covers the watershed making remote sensing a useful tool for collecting snowmelt information. The study site also has 11 ground based monitoring stations (SNOTEL sites) each of which hosts a 2 meter by 2 meter snow pillow that measures snow water equivalent (SWE). Land cover over the study site is variable. Forest accounts for 47.9 percent of the total area with 46.6 percent of the total area classified as evergreen forest. Shrub and grass land account for 45.4 percent of the total area with the remaining 6.7 percent belonging to agriculture, developed land, and wetlands. These percentages are calculated based on the 2011 National Land Cover Database [Homer et al. 2015].

Data Sources

The MODerate Resolution Imaging Spectoradiometer (MODIS) on board the Terra satellite was used. The version six data product, MOD10A1 (MODIS Terra snow cover daily L3 global 500 m Sin GRID V006) distributed by the National Snow and Ice Data Center (NSIDC) was used. This data product comes with the Normalized Difference Snow Index (NDSI) snow cover, NDSI snow cover basic QA, NDSI snow cover algorithm flags QA, NDSI, and snow albedo data. In this study only the NDSI snow cover was used. MOD10A1 was chosen as it has consistently shown less cloud obstruction [Parajka & Blöschl, 2008]. The calculation of NDSI is a combination of band 4 and band 6, which are the visible red and mid infrared wavelengths respectively. The MODIS sensor aboard Aqua had a failure in the mid infrared band forcing a slightly different NDSI calculation. Since no temporal combination of Terra and Aqua was applied, only the MOD10A1 product from the Terra satellite has been used in this study. To convert the data to fractional snow covered area of the total watershed area (SCA) an NDSI threshold of 0.4 was used, as documented in

the version six user guide [Hall & Riggs, 2015]. Two MODIS swaths were needed to encompass the study region, h09v04 and h10v04. All available MODIS imagery from years 2000 through 2018 was used in this study.

SNOTEL data distributed by the Natural Resource Conservation Service (NRCS) was also used as an input in model development. There are eleven sites in the watershed. Base Camp station ID 314 daily SWE measurements, were used as an input for model accuracy. This station is not the lowest elevation site in the study region at 7060 feet, but over the 19 year study period was one of the first stations to melt completely each year.

Methods

Model Development

To compare melt patterns across multiple years one must first define and identify the important days. The pattern of melt can be extracted for each year by identifying two days of interest for each pixel. The last day on which snow (LDS) is seen, and the first day land (FDL) is seen on a pixel by pixel basis can describe melt. The two output files are equal in spatial resolution and geolocation to MODIS imagery. It is also essential to capture the melt of the maximum snowpack, rather than secondary snowfall which can occur in the spring time throughout the melting season. Snow water equivalent (SWE) measurements from ground based SNOTEL stations can help identify secondary snowfall. The model starts each year on the day of year (DOY) of maximum SWE for the SNOTEL station that achieves a SWE of zero early in the melt season. The SNOTEL stations in this watershed all see seasonal snow cover for extended time periods, which may not be true for every watershed. The SNOTEL station Base Camp was used to identify the DOY of maximum SWE. In the model development stage only the years 2000 through 2016 were used.

The model extracts the two days, LDS and FDL, by iterating through all the available imagery one year at a time. A few scenarios can take place during processing. Given a pixel that is identified as snow on the DOY the model starts on, the model begins forward processing, known as the snow protocol. Forward processing means the next day of MODIS imagery is analyzed, and the next, for only the pixels identified as snow on the first day, saving the LDS in one grid file (raster) and the FDL in another raster until each pixel has a value for the LDS and FDL saved in each raster. For the pixels identified as land on the start day the model back processes, analyzing the previous day of MODIS imagery until

snow is seen for each pixel, known as the land protocol. For pixels which were never snow covered, or cloud cover never allowed for snow to be seen, the LDS is saved as DOY zero. The final scenario, which can take place are pixels identified as cloud on the DOY the models starts on. To handle this, the model begins back processing. As the model back processes if pixels are identified as snow first they are sent to the snow protocol beginning on the start DOY day plus one day. For pixels identified as land first they are sent to the land protocol. If a pixel is cloud covered for the entire period from the start day to January 1st the model begins forward processing from the start DOY plus one day. For these pixels if land is seen first, the DOY is saved to the raster containing the FDL. The raster containing LDS is saved as a zero for these pixels, to signify the pixels were never seen as snow before they were identified as land.

This method brings in ground based data (DOY of max SWE at a SNOTEL station) to efficiently reduce the large amount of imagery available to only two images per year. These images describe the melt patterns and the difference between the FDL raster and LDS raster for each pixel is the number of days cloud cover obscured the view of the pixel during the crucial melting time period. Similar to the temporal method applied by Wang & Xie [2009], two images are extracted that describe the dates each pixel transitions from snow to land. This method produces two cloud free images that describe the melt progression through the melting season using the DOY as the value to describe the melt. The DOY and year is used in the nomenclature for MODIS imagery distributed by NSIDC. Given the year 2001 and DOY 145 a MODIS image for tile h09v04 would include the following MOD10A1.A2001145.h09v04 followed by additional version information. The first nine characters identify the product, Terra daily 500 meter sin grid, followed by the four character year, three character DOY, and finally the tile identifier. This information was used in the model to accurately save the DOY for the correct year to the LDS and FDL rasters.

Combining the Data

This model has an output of two images for each year, producing a total of 34 images. The FDL raster from each year can be combined using a principal component analysis into one raster. The principal component analysis is a data reduction method commonly used in remote sensing. Orthogonal directions defined by the Eigen vectors are

used to explain directions of increasing variance. The first principal component captures the direction of maximum variance, followed by the second component describing orthogonal direction of second most variance. The number of data series used determines the number of principal components, 17 years and 17 principal components. This method is used to condense large data sets into only a few principal components which can describe up to 95% or more of the total variance. For an in depth discussion of this method refer to Everitt, B. & Hothorn, Torsten, [2011]. The FDL raster was used because this describes the first day there is confidence a pixel has changed from snow to land. The LDS raster is greatly affected by cloud obscuration over the period of melt. The principal component analysis over data of the same type may be viewed as an indexing method, however since a pattern exists it may be helpful in reducing the amount of data needed to describe the melting pattern across multiple years. The first principal component should describe the melt pattern if the pattern is repeatable across years, meaning the maximum variance is not contained in the year to year changes, but rather the variance in the FDL in each year. This is to say that the distribution of FDL is similar year to year. If the melting trend can be described by the same function year to year by shifting the curve and altering the slope, then the first principal component should capture a great deal of the variance because the maximum variance is attributed to the range in DOY in the FDL for each year rather than the range across multiple years. Equation 3.1 shows the matrix multiplication used to take all 17 years and the first principal component to produce what are referred to as the transformed values. The matrix multiplication in essence projects the 17 years of data into the direction of the first principal component, which is defined by the first Eigen vector. Equation 1 shows the dimensions of the input data. The far left matrix is 41,503 x 17 multiplied by the first principal component rotation matrix dimensions 17 x 1. The rotation matrix projects the data onto the direction defined by the first Eigen vector. Multiplying these matrices in this order gives the transformed values. The final matrix has dimensions of 41,503 x 1 where each row is associated with one pixel, which has the same geographic location.

$$\begin{bmatrix} 1 & \cdots & 17 \\ \vdots & \ddots & \vdots \\ 41,503 & \cdots & 41,503 \end{bmatrix} \times \begin{bmatrix} 1 \\ \vdots \\ 17 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 41,503 \end{bmatrix} \quad \text{Equation 3.1}$$

If this is indeed the case and a pattern exists, the principal component analysis will reduce the 17 FDL rasters, one for each year, to one raster that describes the melting pattern

across all years. We can use the first principal component to transform all 17 years into one raster. The principal component transformation can then be scaled to 0 to 100 on a pixel basis where 1 would be equivalent to 1% of the pixels and therefore 1% of the area. From here a contour map can be developed where contours describe the interface between land and snow at different SCA percentages. Due to the limited number of days with less than 5% cloud cover, contour lines of 5% were developed. The contour map can be used for cloud removal as well as climate change modeling.

Results

Climate Modeling Results

The model produces two rasters FDL and LDS containing only DOY information for each year. The output for 2013 is shown in Figure 3.2. Although the difference is slight between these two images the range for the first day of land (right) has larger minimum and maximum values at 25 and 188 compared to the last day of snow (left) at 21 and 186. The difference between the two images is best shown in a histogram. The distribution of pixel values is shown for 2013 and 2007 (Fig. 3.3). 2013 has a higher frequency of pixels with values less than 5, while 2007 which had more cloud influence has a higher frequency of values greater than 5. The basic trend of melt throughout the 2013 melt season can be seen in both of these images where melt begins with the dark colors and finishes with the light. One can also see there are some spurious errors indicated by the small dark areas completely surrounded by light areas. This suggests that the area melted much earlier than the surrounding pixels sometimes 100 days earlier. This issue is observable for each year, but is the result of MODIS sensor values. The model only extracts the days therefore these errors are a part of the raw MODIS image and illustrated in the FDL and LDS images.

The output for each year is then combined using a principal component analysis. The rotation matrix of the first principal component is used for matrix multiplication to produce a single matrix. Each row in the final matrix (transformed values from Equation 3.1) is associated with a single pixel and when this is plotted with the x, y coordinates it produces the final single map describing melt patterns. The first principal component captured 85% of the total variance across all 17 years used. As discussed earlier the majority of the variance is contained in the singular DOY values in one year, and because there is a pattern of melt the 85% variance captured does not describe yearly fluctuations in melt patterns. Each

subsequent component is far less useful with the second component describing only 2% of the total variance. The additional 16 components, principal components two through seventeen, can be considered noise because of the small amount of variance they describe. The contribution of each year in the first principal component is shown in Table 3.1. Most years have a weight between 0.2 and 0.27 with only two less than 0.2, 2013 and 2006 both at 0.19. The two that fall outside the upper bound are 2004 and 2010 at 0.31 and 0.29 respectively. The principal component analysis is designed to capture the maximum variance and in turn years that are different than the rest are weighted higher. The most similar years, which produce the smoothest output with fewer spurious errors, 2006 and 2013, are weighted the lowest. The similarity in the weights among all 17 years supports the idea of a repeatable pattern across multiple years.

The transformed values can be re-plotted using their geographic coordinates to create a raster describing melt using the first principal component (Fig. 3.4). The isolated errors noted earlier in the 2013 output can be seen in the principal component map. The black area at the northernmost tip of the watershed along with the specs of black scattered throughout the north half of the watershed all appear to be spurious errors. The contour map is developed from this in ArcGIS (Fig. 3.5).

Model Accuracy

Two years were not used in the principal component analysis, 2017 and 2018. To verify the fit of the model, days with less than 5% cloud cover were used. In 2017 there were 15 days available and for 2018 at the time of completing the analysis 8 days were available. The contour map is drawn with 5% SCA intervals. Total error was assessed as the sum of the commission and omission errors. Since the map is modeling snow covered area an omission error is considered a pixel modeled as land, but measured as snow using the raw MODIS image. Because of the framework of ArcGIS it is easier to represent snow cover in terms of exposed land percentage. The snow covered area is therefore 100% minus the exposed land. This choice was made to make processing of multiple images easier in ArcGIS even though the model is aimed at snow covered area. In 2017 the greatest error of 15% is assessed at 75% SCA (Table 3.2). Between the range of 75% and 55% SCA, the error is above 10% in 2017. For 2018 the errors exceed 10% between 85% and 65% SCA. Outside these SCA ranges, the errors drop significantly. The smallest errors for 2017 and

2018 are 2.41% and 3.4% at the smallest available SCAs, 5% and 10% for 2017 and 2018 respectively.

The years included in the analysis also offer information about the model's accuracy for SCA. Five additional years were used for accuracy assessment 2001, 2004, 2007, 2011, and 2013. These years were chosen based on the peak SWE and melt-out dates to represent outlier years in snow pack and average years. Each year represents a different snow condition meaning low snowpack (2001 and 2007), high snowpack (2011), and average snowpack (2004 and 2013). The model performed similarly in each of these years as it did in 2017 and 2018. The minimum error for each year was assessed at the smallest available SCA. The error at 5% SCA varied from 2.14 to 2.42 percent and at 10% SCA errors varied from 3.67 to 4.39 percent. The greatest error for 2001 was 12.49% at 65% SCA, 12.49 in 2004 at 65% SCA, 12.88% in 2007 at 55% SCA, 14.9% in 2011 at 60% SCA, and 13.72% at 65% SCA in 2013. The error is greatest in the years used in the map development and the unused years at a SCA of 50% to 75%. Taking the average of the total error at each SCA shows the error is greatest at 60% with a value of 14.73%. On average, the errors are above 10% between SCAs of 80-40%, and drop significantly on both ends. These errors are favorable in comparison to other research. Gafurov et al. [2016] retrieved accuracies of 77.3 -99.8% from the Karadarya Basin in the 84 days with $\leq 10\%$ cloud cover in 2008, with the lowest accuracy appearing in the transitional periods such as the melting season. These numbers were calculated under 100% cloud cover, but require additional days to produce cloud free imagery. The Adaptive Spatio-Temporal Weighted Model (ASTWM) also produces cloud free imagery on a daily time step. This method found greater overall accuracy ranging from 93.11 – 98.92% with a five day temporal filter [Li, et al., 2017]. The watershed used in their study had seasonal snow fraction of 25% compared with seasonal snow cover of 100%, changes in snow coverages are limited in that study region. Identifying snow in this watershed guarantees an accuracy of at least 75% at the height of snow coverage with the potential for error reduced to only 25% of the watershed. A conditional probability method which incorporates interpolated meteorological data also found good agreement of 92% during the snow season, but noted reduced accuracy during the latter half of the snow season with errors ranging from 22% to 13% [Dong & Menzel, 2016]. The model presented in our research is a relatively easy technique for identifying SCA, and

produces accuracies of 85-97.9% during the melting season. MODSCAG produces an average RMS error of 5% across multiple images modeling fractional snow cover on a daily basis with added complexity [Painter et al., 2009]. This varies the spectral response on a pixel by pixel basis to identify snow cover based on the preceding MEMSCAG model [Painter et al., 2003]. This method is useful for determining fractional snow covered area, but requires a cloud free day and cannot be applied for climate modeling as it requires a sensor response to produce an image.

The increase and decrease of errors at specific snow covered areas is explained by the location of the errors in the contour map, and the type of errors assessed. Comparison of the average omission error percentage and the average commission error percentage in a year shows them to be nearly equal (Table 3.3). In each year except 2018, the omission error is slightly larger than the commission error. Total average error across all SCA was smallest in 2011 at 5.8% and largest in 2007 at 9.67%. In 2013, the average omission error was 3.68% and the total average error was 6.82%, meaning that omission errors accounted for about half of the total error. The omission and commission errors are nearly equal on average.

Identifying the type of error shows the error is actually due to the distribution of SCA. The contour map (Fig. 3.5) models SCA and if an area that does not melt early is incorrectly modeled as early melt the area is distributed from the area of true melt to an area that is not actually melting. The equal commission and omission errors are due to this error in the distribution of snow covered area. Errors exceeding 10% begin to occur when SCA drops to 80% with the average greatest error occurring at 60% SCA (Table 3.2). An example of where the error is occurring is shown in Figure 3.6 at 75% modeled snow covered area using an image from April 26, 2018 with a measured SCA of 77.99%. The modeled area of exposed land is overlaid on the image from 4/26/2018. The image is shown in three colors. Brown represents measured land, blue is measured snow, and white is cloud cover since the image had 2.5% cloud cover. The model predicted exposed land is shown in transparent pink. Blue areas are where the model correctly predicted snow. Brown shows where the model predicted snow, but the image shows land (commission error). The fuchsia areas (transparent pink over brown) are the areas where the model correctly predicted land. The purple areas represent omission errors where the model indicates land, but the imagery on

that day shows snow. In this image it is clearly seen that the north and western portions of the watershed suffer from spurious errors in the model. The locations where the model indicates land but the image shows snow in Figure 3.6 coincide with the initial model output before the principal component analysis. The initial model output refers to the yearly FDL image shown in Figure 3.2. As shown in Figure 3.2 some locations are showing an early melt out date surrounded by later melt out dates (dark pixels surrounded by light pixels). These occur in the same areas the final map incorrectly models land as shown in Figure 3.6. This type of error was seen in all 17 FDL images. These islands of incorrect “early melt” are passed through the principal component analysis and lead to errors in the models spatial accuracy of snow extent. These errors are omission errors, but because the model is built on a percentage of area basis if the model shows an area incorrectly melting early the area which is actually melting has commission errors. A percentage area representing exposed land is shifted into an incorrect location of the watershed away from the correct location.

When the areas which are incorrectly predicting exposed land are compared to the land cover a significant source of error becomes apparent. Figure 3.7 shows the impacts of land cover on the model accuracy for the same image as in Figure 3.6 using the 2011 National Land Cover Database. Areas where the model correctly identified exposed land are blacked out, whereas areas with omission errors where the model incorrectly identified land when images showed snow are outlined in red. Green indicates coniferous forests. Nearly all of the red outlined areas are filled with green. Use of a fixed threshold for each pixel regardless of land cover to identify binary classifications of land versus snow across images proves to be problematic. The threshold for identifying snow cover on a pixel by pixel basis has been shown to reduce from 0.45 in open areas to 0.2 in forested areas [Rittger et al., 2013]. The purpose of this research is not to develop more accurate methods for identifying a pixel on a cloud free day. The error due to forest cover is persistent throughout all years and therefore a part of the multiyear combination. These thresholds are not consistent based on land cover alone. Viewing angle, the dropping of snow from the canopy to the forest floor, and topographic effect all influence the NDSI value. These factors can produce a wide range of NDSI values and any NDSI value greater than 0 indicates presences of some snow in the pixel, but quantifying how much requires a more intensive approach [Riggs et al., 2017]. Where the model does not take into account the effect of viewing angle, low

illumination, and a varying NDSI spectrum that can be seen in pixels of differing land cover the FDL can develop biases. The reason these biases result in errors when compared to a cloud free image could be the result of viewing angle and illumination alone. A cloud free image taken at nadir with good illumination may produce all around higher NDSI values compared to an image taken at a poor viewing angle and partially cloud covered. The FDL image does not differentiate but rather uses the threshold of NDSI greater than or equal to 0.4 to identify snow.

The effects of forest on model accuracy are quantified during the melt season for SCA values ranging between 80 and 40 percent in Table 3.4. The entry for 45% was omitted because no imagery was available. This was done by identifying the percentage of the areas that were incorrectly modeled as land (omission errors) whose land cover was coniferous forest. Only one image was used for each SCA value in Table 3.4. Preference was given to imagery in 2017 and 2018, but if a specific SCA was not available the image with the highest omission error was used from 2001, 2004, 2007, 2011, or 2013. Coniferous forest covered between 75.2 and 90.0% of the model misclassified areas. Remote sensing of snow in forested areas is challenging; initially when snow falls on forested areas, it clings to the top branches of trees and is detectable in satellite images, however, after some time the snow is shed from branches and accumulates on the ground below the tree canopy where it is difficult to detect by visible remote sensing and the pixel is prematurely classified as land. This movement of snow from canopy to forest floor is made even more complex by viewing angle of the sensor, illumination, and landscape features. Eventually the below-canopy snow melts, but it is not readily apparent when this transition occurs. The use of a fixed threshold for land/snow classification in this research induces premature classification of some coniferous forest pixels as land, and implementation of the non-reversible “First Day of Land” classification, intended to identify the melt pattern of the primary snowpack as opposed to secondary snowfall events on previously melted land, causes this omission error to persist in the model. Most of the coniferous land cover pixel omission errors occur in isolated “islands”, separate from the main contiguous body of “land” pixels. Once these misclassified islands become engulfed by the main area of melt, which occurs around 35% snow covered area, the errors drop significantly. Agreement between modeled and remotely sensed SCA for a given day increases from around 85% when the coniferous forest pixel

omission errors are largest at SCA values of about 60%, to 97.8%. This shows the impact that early misclassification of a pixel as “land” has on the model accuracy.

Cloud Removal Results

Cloud removal is another area of application for this model. A primary area of research for snow melt modeling purposes with MODIS is cloud cover removal as discussed earlier. For this watershed in the 150 day period in which melting generally occurs, DOY 60 – 209, the number of days with less than 10% cloud cover is limited. A closer look at when these minimal cloud-cover days are available reveals they generally occur during the tail end of the melt period. Within the overall 150 day melt interval, the most critical window is when rapid melting occurs, especially between SCA of 80-10%. During this narrowed window, only a small number of days remain with less than 10% cloud cover. This time period is most relevant for snowmelt modeling in real time because a large fraction of the total snowmelt volume flows out during this period, and because this period also encompasses the peak discharge. Between 2000 and 2017, 25 days or more between DOY 60 and 209 for each year have less than 10% cloud cover. Even with this relatively large number of low cloud cover images none of the 19 years has more than 13 days with less than 10% cloud cover between SCA values of 80-10% (Fig. 3.8). The average number of days with less than 10% cloud cover during the transition from 80-10% SCA is 9 days. The average length of this period across the 18 years is 66 days. 2018 was not included because the melting was not complete at the time analysis. This period is relatively short for the amount of change that occurs in the snowpack. The longest melting period (80-10% SCA) occurred in 2010 at 92 days and the shortest occurred in 2014 at 44 days. This spatial change in snow coverage does not explicitly mean small or large snowpack, but rather a fast spatial melt rate once it begins. With the short window of time critical to real time modeling using MODIS becomes difficult. While many methods produce cloud free imagery they all employ a temporal interpolation that can take days to complete. The delay added to availability because of the temporal interpolation in the short time period of changes in SCA can be troublesome. This delay to cloud free imagery adds on to the delay for imagery from the NSIDC. While these seem trivial, if the melt period from 80-10% SCA is only 44 days long a 5 day wait for imagery can influence real-time snowmelt modeling.

The principal component spatial model presented in this article (Fig. 3.5) can be applied to daily imagery to remove cloud cover. As long as some portion of the interface between snow and land is visible in a daily image, interpolation of the snow covered area can be accomplished. Each line shown in the model (Fig. 3.5) represents the interface between land and snow. To apply this model as a method for cloud removal the only step is choosing the best fit line in the model to the interface of snow and land in the cloudy image. The accuracy provided by cloud-free image days is not relinquished in this method, but rather serve as a refinement for the progression of snow melt. For example, if the watershed was measured at 80% snow covered area on a cloud free day, and the next day is cloud covered the value of 80% can be used to narrow the number of lines in the model under consideration for cloud removal. The cloud covered day can be compared against the 90%, 85%, 80%, 75%, and 70% lines in Figure 3.5. Since the time period we are most interested in removing cloud is during the melting period the progression of SCA should be declining. For this reason we consider 80%, 75%, and 70% SCA lines against the visible interface of snow and land on the cloudy day. It is unlikely in this watershed that more than 10% of the area melts in the span of one day. We also look at 90% and 85% SCA in case there was a secondary snowfall event which would increase SCA. The impacts of forest cannot be overlooked when applying this model. Certain areas have better agreement with the interface between land and snow than others. Cloud removal can be achieved quickly. As soon as data is available and converted to binary snow cover, this contour map can be applied.

A relatively clear day in 2006, DOY 152 with a total of 4.2% measured cloud cover, is used to illustrate the use of the model presented here for cloud cover removal and to assess the overall accuracy achieved. A “cloud” that obscured 85% of the watershed was superimposed on the image (Fig. 3.9). The areas that remain visible were used to apply our model and produce a cloud-free image. Even with the substantial cloud cover the snow covered area can be determined because portions of the interface between snow and land are visible. Contours of 20%, 25%, and 30% SCA from the model are displayed. As one would do in real-time modeling, information from the previous day was used to inform the best guess of snow covered area. DOY 151 was measured to have 27% SCA with 0.5% cloud cover. Because of this the contours of 20% through 30% were selected for comparison. Very small portions of the interface are visible on DOY 152, however as shown Fig. 3.9 the 30%

SCA line depicted in red clearly overestimates the snow coverage. The ideal line would pass through each pixel of snow that is adjacent to a pixel of land. The 20% line appears to underestimate the coverage however the 25% also is not a perfect fit. The best fit contour lies between 25% SCA and 20% SCA, but of the two the 25% contour appears to provide better agreement. Once the imposed cloud is removed the contour interpolation accuracy is revealed (Fig. 3.10). Using only the small segments of the visible interface the best fit is identified as an SCA percentage between 20% and 25%, leaning closer to the 25%. In some areas the 20% contour fits best, but generally underestimates snow covered area. The 25% contour is a better fit overall since it equally underestimates and overestimates the snow covered area. The actual measured snow covered area was 25% on DOY 152 in 2006 with 70.7% land cover.

Discussion

The final product of this analysis is a contour map (Fig. 3.5) describing the interface between land and snow at different SCA percentages. Applying this to produce daily cloud-free images is done quite simply. A daily NDSI snow cover image can be imported into ArcGIS, reclassified into fractional snow covered area, and the snow covered area can be extracted by identifying the interface between land and snow. For real time application this method can provide daily data of snowmelt as long as a portion of the interface is visible. Current methods that achieve complete cloud removal require time to process since they incorporate a temporal filter. Real time modeling is hindered by this delay of data availability and thus the methods cannot be used during a current melting season, especially during the middle of the melt season when melt occurs rapidly. The only model which outperforms this approach is the ASTWM with agreement of 93.11% or better [Li et al., 2017]. This was achieved in a watershed that sees maximum fractional snow coverage of 25%, a temporal filter of 15 days, and 26 meteorological stations. This method may face difficulty when applied in a watershed with complete seasonal snow coverage and no meteorological stations. Although a SNOTEL station was used in the model developed in this research, it is not a requirement and a start date appropriate to the watershed could be applied each year in the absence of ground data. Aside from these issues for real-time modeling none of the current methods can be used in climate models. This method produces a spatial map which characterized the melt patterns of this watershed.

A criticism of this technique may be the similarity of snow packs seen over the 19 year study period; however this is not the case. A wide variety of snowpack duration and peak SWE was observed across these 19 years (Fig. 3.11). Lewis Lake Divide, situated at 7,850 feet had the largest snow water equivalent measured between 2000 and 2018 at 56.9 inches. This value far outweighs the value recorded at Gunsight Pass of 27.3 inches even though Gunsight Pass is 1,970 feet higher in elevation. This difference shows the variability not only in the years but how elevation is not always an indication of higher snowpack. The lowest elevation site Granite Creek at 6,770 received a maximum of 30.1 inches of SWE. All of these maximum SWE values substantially exceed the minimum peak SWE values of 19.1, 10.5, and 10.0 inches at Lewis Lake Divide, Gunsight Pass, and Granite Creek respectively. For Lewis Lake Divide the range of maximum SWE values recorded across all 19 years is 37.8 inches. Median SWE values at these latter stations were 31.7, 13.4, and 15.7 inches respectively. It is important to note the maximum SWE values across all eleven SNOTEL Stations were not all recorded in the same year. The largest maximum SWE values on a station by station basis were recorded in 2011 (7 stations), 2017 (3 stations), and 2014 (one station). No single years had all of the largest maximum SWE measurements at all stations, but some years such as 2011 and 2017 had overall larger SWE values at all eleven stations. The smallest maximum SWE values recorded across all 19 years on a station by station basis were also spread across multiple years. Six of the smallest maximum values were recorded in 2001, five in 2010, and one in 2004. Like the large snowpack years of 2011 and 2017 these low snowpack years had overall smaller maximum SWE values at all the SNOTEL stations. In the 17 years used for model development two years could be considered large snow fall years 2011 and 2014 and three years could be considered drought years, 2001, 2004, and 2010. In the years used for model verification 2017 had considerably larger SWE measurements than most other years. The DOY complete melt was achieved at each station further documents the differences among the years (Table 3.5). At all stations 2011 had the latest melt out date (first day of SWE equal to zero). The earliest melt out dates were shared between 2000, 2001, 2007, 2012, and 2015. Large peak SWE was recorded in 2011 with a late melt, whereas 2001 saw small peak SWE and early melt. This ground based information shows the variability in the snowpack in all 19 years and the timing melt. The principal component model presented here applies independent of snowpack and timing.

In fact this great variation is handled well by the model. Since the pattern is consistent, the model performs well regardless of snow depth or timing. The methods that use a probability of snow cover based on snow cover season, add this step at the end of their cloud removal methodology since it is error prone [Gafurov et al., 2016]. Other research groups have used elevation zones to identify areas with similar snow cycles. A snow cycle was determined based on accumulation and ablation. Areas were grouped together if the “cycle” was similar in multiple years; however in application the group only used a single year cycle to remove cloud cover. Patterns were noted in this method, but cloud removal relied on the current year snow cycle, namely the day of first accumulation, maximum snow extent, and day of minimum snow extent. Increased accuracy using this method was seen. The snow cycle step was added at the end of the 5-step cloud removal protocol, but takes advantage of the shift in snow melt timing [e.g., Paudel & Andersen, 2011]. A set back to this method is the entire snow season has to pass before the cloud can be removed. In our watershed alone it is easy to see why using a probability of snow cover based on other pixels is error prone given a highly variable snow pack. Of the eleven SNOTEL stations five (Thumb Divide, Lewis Lake, Snake River Station, Phillips Bench, and Gunsight Pass) each melted out completely in 3 different months over the 19 year study period (Table 3.5). This variability in melt timing shows the issues with using a bulk statistic to estimate melt timing. In the model presented in this article the melt timing is adjusted for a particular time in the pattern which is consistent year to year. Using elevation zones and snow cycles as proposed by Paudel & Andersen [2011] may help deal with this problem, but is still limited by the single-year-at-a-time approach, overlooking the pattern. This method uses an aggregated approach rather than a spatial model which limits its applicability in the middle of a melting period since the first day of accumulation, day of maximum extent, and day of minimum extent all have to be known before cloud removal can be completed. Extreme years may pose problems for these methodologies since they lie in the tails of statistical distributions. In snowmelt modeling, extremely high snowpack years are of utmost importance for flood control especially when snowmelt occurs late in the season (e.g., 2011), and small snow pack years with early melt (2001) are important for drought simulation. The point that is overlooked by these previous models is that the melt pattern is the same, and only the time and rate of melt are shifted. With a contour map derived from a collection of highly variable

snow pack years this issue can be dealt with effectively. This is shown with the model performance in 2017 a large snow pack year and 2018, as well as the performance in the significantly different years such as 2011 and 2001. The model performs adequately independent of snowpack, a difficult achievement using other approaches.

The principal component analysis achieves a good fit at varied SCA percentages across multiple years. The model achieves this with 84.8% of the total variance captured in the first principal component. The other 15% of variance is thrown away. A simple analysis can show where this percentage which is thrown out comes from. A principal component analysis was conducted on the difference between the FDL and LDS imagery. This is the influence of cloud in the model accuracy. Significantly less information is explained by the first principal component showing the variability in the cloud persistence during the most crucial time as a pixel melts. Relating the principal components of the difference PCA and the FDL PCA may show the source 15% of variance that was not used. A correlation matrix was created to compare the two analyses. Table 3.6 shows the matrix with the highest correlations bolded and boxed. Also shown along the bottom is the contribution of each principal component in describing the variance. The first principal component describes 84.8% of the total variance and the second describes only 2%. In fact, after the first principal component no subsequent principal component describes more than 2% of the total variance. An addition of the next eight components, 2-9, only describes 10 additional percent of the total variance. Any component which describes less than 1% of the total variance can be considered noise. For this section of analysis principal components 2-9 of the FDL PCA are analyzed for their relationship to the influence of cloud.

Each high correlation relates the FDL PCA and the difference PCA. By tracing the principal component which is most highly correlated to the FDL PCA for each principal component the year related to the high correlation from the difference PCA can be extracted. This is meaningful because if the year with the highest weight in the difference PCA also has a high occurrence of cloud cover then the relationship can show that the additional variance described in the FDL PCA is actually only capturing variance due to cloud influence. The difference image once again shows the duration of cloud cover between the last day of snow and first day of land on a pixel by pixel basis. From the correlation matrix we can identify the principal components in the difference PCA related to PC 2-9 in the

FDL PCA. This is best shown through the relation Table 3.7. For example, for PC 2 of the FDL PCA the highest correlation calculated as PC 10 of the difference PCA with a correlation of 0.73. PC 10 of the difference PCA has the highest weights for the years 2000 and 2004. The correlation is positive so therefore the year should be positively weighted in the principal component. With this logic we can say the year most related to PC 2 of the FDL PCA is 2000. Both of the years with the highest weights, positive and negative, are shown for transparency in Table 3.7.

Once the years are identified the relationship between the principal components of the FDL PCA and cloud can be determined. The average number of days a pixel was obscured on a yearly basis was calculated. This was accomplished by summing up the difference image values and dividing by the total number of pixels. Figure 3.12 shows the average number of cloud obscured days for each year. The average of all the years was calculated to be 5.37 days which is shown in the figure as a horizontal line. Comparing the years from Table 3.7, one can see 2000, 2002, and 2011 all have greater average cloud influence than the overall average. 2004 does fall below average, but is only related to PC4 of the FDL PCA which may represent an outlier not the overall relationship. Across all the years shown in Table 3.7 only 2004 and 2001 fall below the average cloud cover line.

With this information it can be shown that in fact the principal components discarded in the derivation of the contour map (Fig. 3.5) add no additional information about the overall pattern. The information left out is a product of the influence of cloud and does not, with confidence, describe any additional information about the pattern.

Using the pattern as a method for developing a map of characteristic melt pattern makes this model superior and applicable for climate models. Years of highly variable peak snow water equivalent and melt out dates were used in the derivation of the map. The model shows good agreement during extremely high or low snowpack years as well as during years of average snowpack. It has also shown good agreement with validation years not used in development of the snowmelt pattern, making it applicable for use in climate models. None of the other methods focus on this untapped area of research. Current methods for climate modeling of snowpack are downscaled approaches taking advantage of meteorological data [e.g., Abatzoglou & Brown, 2012]. These methods may be useful over large areas; however snowpack accumulation and ablation is not driven by elevation alone, as it is merely a

portion of the topographic effect [Hock, 2003]. Smaller-scale, spatially distributed models developed from observable spatial distribution of snow are lacking. Incorporation of these types of models into climate change models may better depict spatially what happens with rising temperatures.

The simple application of this model for an entire snow season can be seen in Figure 3.13. The only missing information with the map derived in this research is a dimension of time. To add a dimension of time all that is required is fitted curve to the melt progression with time, or a depletion curve. By using cloud free days (cloud cover $\leq 5\%$) a line can be fit. Using a least squares optimization the curve can be fit to the cloud free days, which is what was done in Figure 3.13 for two snow seasons, 2007 and 2011. Each horizontal line represents a 5% decrease in snow covered area, which is what the contour map (Fig. 3.5) models. The raw data showed a similar distribution to a Gaussian decay function when plotted with time. Theoretically if a pattern is repeatable the same type of curve could be applied to multiple snow years. This is shown here as a Gaussian decay curve is fit to two very different snow years 2007 and 2011. The difference in the rate of melt or steepness of the curve and timing of 50% melt differ greatly. By incorporating a time component into this, once again using only raw data to inform the decision and curve fitting, the contour map shown in Figure 3.5 has the power to accurately represent a melting season with time. This simple application is of great worth for snowmelt modeling is all still rooted in the fundamental idea of consistent melting patterns across multiple years. While the Gaussian decay curve fits this data it may not fit other watersheds. It was chosen and applied here because the data appeared to follow that curve in two very different years. Each watershed may have its own curve which is characteristic to that area and melting pattern.

A significantly valuable component of this method for modeling snow distribution is its data collection methodology. Only free data collected through remote sensing is required to develop the model. The final model is not a function of elevation alone, but rather a combination of all the drivers of snowmelt on a pixel by pixel basis taking full advantage of the intra-annual pattern of snowmelt. Consequently, this model is ideal for gauged and ungauged watersheds alike. For gauged watersheds use of ground based data to inform the starting day of the model can increase the accuracy of the model and in ungauged stations a standard starting day could be used.

Limitations

This model was applied to a watershed that experiences complete snow coverage each year. A change in this may result in a change in the model accuracy due to differencing snow extents. A study conducted over an area of 950,000 square kilometers in Eurasia encompassing Kazakhstan which receives a maximum snow coverage of 32% in some areas, and up to 80% in other portions, noted a pattern [Zhou et al., 2013]. While the derivation of the map may change, the principal would still apply. The characteristic map is also limited by the available years, the use of a consistent NDSI threshold rather than a variable threshold for different land cover types [Painter et al., 2009], and the watershed size. For smaller watersheds a clear repeatable pattern may be more difficult to derive due to the impact cloud cover may have. It can also be noted that the confidence in the pattern is variable throughout the melt period.

Accuracy may vary at SCA between 80-40% due to early misclassification of pixels as land. The model extraction of FDL and LDS is set up in such a way that if the threshold is passed the pixel is deemed as melted. To combat this issue and produce more accurate results, a variable threshold, or spectral mixing could be employed [Rittger et al., 2013]. The early detection errors are propagated through the model displacing area of actual melt to areas of incorrect melt since the model works on the basis of percent area. A pixel which is incorrectly deemed melted in a forested area early in the melting period commits a portion of the total percentage of melted area incorrectly. The use of a principal component analysis to combine the data favors the most distinctly different patterns as well. The goal of the principal component analysis is to reduce the data into orthogonal independent planes described by the Eigen vector. These planes capture the directions of decreasing variance where the first component indicates the direction of maximum variance. Because of this the two most dissimilar years from the majority are favored in the first principal component as well as the areas of early detection of melt due to land cover. Proof of this favoritism can be seen in the errors associated with these years. One of these years 2011, showed some of the best average agreement. This year had a very late melt out date for the SNOTEL sites and had the latest starting date for the model.

Conclusion

The goal of this research was to develop a spatially distributed model of snowmelt processes. The model was based on 17 years of data and combined to produce a contour map which has applications in cloud removal during melt, and climate change modeling. The model performed with 85.0-97.6% agreement in 2017 and 2018. Compared with other models which also produce cloud free imagery such as MODSNOW-tool and ASTWM the model performs adequately, but outperforms all in terms of timing of data availability [Gafurov et al., 2016; Li et al., 2017]. No current method produces daily cloud free imagery without a temporal filter which requires imagery on subsequent days of the cloudy day. This method reduces the time required for cloud free image production from multiple days, a requirement for the temporal interpolation to create cloud free imagery, to near real-time as long as some small portion of interface between land and snow is visible. For real time snowmelt modeling this method pioneers a new approach for the production of cloud-free snow covered area images on a daily time step which is simpler to use on an ongoing basis compared to existing, more complex methods such as MODSCAG [Painter et al., 2009].

The model is not restricted to cloud removal and real-time modeling. The map describes a long lasting intra-annual pattern of snowmelt over a variety of differing snow packs and durations with 85.0- 97.9% agreement. It serves as a spatially distributed model, characteristic of a watershed which may increase modeling accuracy in climate change scenarios. Current methods rely on downscaled data and modeling of snow packs. While this approach is accepted spatially distributed data is far more valuable than interpolated data. This model is applicable to watersheds with or without *in situ* data, intuitive to use for cloud removal, and easily incorporated as spatial snow cover extent into modeling scenarios.

While the focus of remote sensing of snow has been restricted to cloud removal and increasing accuracy a significant application of long term trends has been overlooked. Many researchers have noted the existence of consistent patterns of snow melt across multiple years [Adams, 1976; König & Sturm, 1998; Luce & Tarboton, 2004; Sturm & Wagner, 2010; Wang & Xie, 2009; Zhou et al., 2013]. To our knowledge, this research is the first to develop a methodology to assimilate multi-year data for the purpose of defining the seasonal spatial melt pattern for a watershed. This article uses Principal Component Analysis to identify the melt pattern of a test watershed, and applies the pattern for cloud removal. Use

of the method to simulate the impact of climate change on snowmelt is suggested as an application for future research.

References

- Abatzoglou, J. T., & Brown, T. J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, 32(5), 772-780. doi:10.1002/joc.2312.
- Adams, W. P. (1976). Areal differentiation of snow cover in east central Ontario, *Water Resources Research*, 12(6), 1226–1234, doi:10.1029/WR012i006p01226.
- Dietz, A., Wohner, C., & Kuenzer, C. (2012). European Snow Cover Characteristics between 2000 and 2011 Derived from Improved MODIS Daily Snow Cover Products. *Remote Sensing*, 4(8), 2432-2454. doi:10.3390/rs4082432.
- Dong, C., & Menzel, L. (2016). Producing cloud-free MODIS snow cover products with conditional probability interpolation and meteorological data. *Remote Sensing of Environment*, 186, 439.
- Everitt, B., & Hothorn, Torsten. (2011). *An introduction to applied multivariate analysis with R (Use R)*. New York: Springer.
- Gafurov, A., & Bárdossy, A. (2009). Cloud removal methodology from MODIS snow cover product. *Hydrology and Earth System Sciences*, 13(7), 1361-1373.
- Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., & Merz, B. (2016). MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environmental Earth Sciences*, 75(14), 1-15. doi:10.1007/s12665-016-5869-x.
- Gao, Y., Xie, H., Lu, N., Yao, T., & Liang, T. (2010). Toward advanced daily cloud-free snow cover and snow water equivalent products from Terra–Aqua MODIS and Aqua AMSR-E measurements. *Journal of Hydrology*, 385(1), 23-35. doi:10.1016/j.jhydrol.2010.01.022.
- Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12), 1534-1547. doi:10.1002/hyp.6715.
- Hall, D.K. & Riggs, G. A. (2015). *MODIS Snow Products Collection 6 User Guide*. doi: <http://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf>. [July 12, 2018].
- Hall, D. K. & Riggs, G. A. 2016. *MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6*. [NDSI snow cover]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: <https://doi.org/10.5067/MODIS/MOD10A1.006>. [January 2017 - March 2018].
- Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, 282(1), 104-115. doi:10.1016/S0022-1694(03)00257-9.

- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., (2015). [Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information](#). *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354
- Huang, Y., Liu, H., Yu, B., Wu, J., Kang, E. L., Xu, M., et al. (2018). Improving MODIS snow products with a HMRF-based spatio-temporal modeling technique in the Upper Rio Grande Basin. *Remote Sensing of Environment*, 204, 568-582. doi:10.1016/j.rse.2017.10.001.
- Jensen, J. (2016). *Introductory Digital Image Processing: A Remote Sensing Perspective* (4th ed.).
- König, M. & Sturm, M. (1998). Mapping snow distribution in the Alaskan Arctic using aerial photography and topographic relationships, *Water Resour. Res.*, 34(12), 3471–3483, doi:10.1029/98WR02514.
- Krajčí, P., Holko, L., Perdigão, R. A. P., & Parajka, J. (2014). Estimation of regional snowline elevation (RSLE) from MODIS images for seasonally snow covered mountain basins. *Journal of Hydrology*, 519(PB), 1769-1778. doi:10.1016/j.jhydrol.2014.08.064.
- Li, X., Fu, W., Shen, H., Huang, C., & Zhang, L. (2017). Monitoring snow cover variability (2000–2014) in the Hengduan Mountains based on cloud-removed MODIS products with an adaptive spatio-temporal weighted method. *Journal of Hydrology*, 551, 314-327. doi:10.1016/j.jhydrol.2017.05.049.
- Luce, C. H., & Tarboton, D. G. (2004). The application of depletion curves for parameterization of subgrid variability of snow. *Hydrological Processes*, 18(8), 1409-1422. doi:10.1002/hyp.1420.
- Martinec, J., & Rango, A. (1986). Parameter values for snowmelt runoff modelling. *Journal of Hydrology*, 84(3), pp.197–219. doi:10.1016/0022-1694(86)90123-X.
- Painter, T. H., Dozier, J., Roberts, D. A., Davis, R. E., & Green, R. O. (2003). Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, 85(1), 64-77. doi:10.1016/S0034-4257(02)00187-6.
- Painter, T. H., Rittger, K., Mckenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4), 868-879. doi:10.1016/j.rse.2009.01.001.
- Parajka, J., & Blöschl, G. (2006). Validation of MODIS snow cover images over Austria. *Hydrology and Earth System Sciences*, 10(5), 679-689.

- Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. *Water Resources Research*, 44(3), N/a. doi:10.1029/2007WR006204.
- Parajka, J., Pepe, M., Rampini, A., Rossi, S., & Blöschl, G. (2010). A regional snow-line method for estimating snow cover from MODIS during cloud cover. *Journal of Hydrology*, 381(3), 203-212. doi:10.1016/j.jhydrol.2009.11.042.
- Parajka, J., & Blöschl, G. (2012). MODIS-based snow cover products. validation and hydrologic applications. In: Chang, Y., Ni-Bin, H. (Eds.), *Multiscale Hydrologic Remote Sensing Perspectives and Applications*. Springer, 185–212.
- Paudel, K. P., & Andersen, P. (2011). Monitoring snow cover variability in an agropastoral area in the Trans Himalayan region of Nepal using MODIS data with improved cloud removal methodology. *Remote Sensing of Environment*, 115(5), 1234-1246. doi:10.1016/j.rse.2011.01.006.
- Qiu, L., You, J., Qiao, F., & Peng, D. (2014). Simulation of snowmelt runoff in ungauged basins based on MODIS: A case study in the Lhasa River basin. *Stochastic Environmental Research and Risk Assessment*, 28(6), 1577-1585. doi:10.1007/s00477-013-0837-4.
- Qualls, R. J., Taylor, R. G., Hamilton, J., & Arogundade, A. B. (2013) Climate change opportunities for Idaho's irrigation supply and deliveries. *Journal of Natural Resources Policy Research*, 5(2-3), 91-105, DOI: 10.1080/19390459.2013.811856.
- Rango, A. (1993). II. Snow hydrology processes and remote sensing. *Hydrological Processes*, 7(2), 121-138. doi:10.1002/hyp.3360070204.
- Riggs, G.A., Hall, D. K., & Roman, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data Records. *Earth System Science Data*, 9(2), pp.765–777. doi:10.5194/essd-9-765-2017.
- Rittger, K., Painter, T. H., & Dozier, J. (2013). Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, 51(C), 367-380. doi:10.1016/j.advwatres.2012.03.002.
- Ronco, P., & De Michele, C. (2014). Cloud obstruction and snow cover in Alpine areas from MODIS products. *Hydrology and Earth System Sciences*, 18(11), 4579-4600. doi:10.5194/hess-18-4579-2014.
- Safari Shad, M., Habibnejad Roshan, M., & Ildoromi, A. (2014). Integration of the MODIS Snow Cover Produced Into Snowmelt Runoff Modeling. *Journal of the Indian Society of Remote Sensing*, 42(1), 107-117. doi:10.1007/s12524-013-0279-y.

- Serreze, M.C., Clark, M. P., Armstrong, R. L., McGuinness, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research* 35(7), 2145–2160. doi:10.1029/1999WR900090.
- Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A., & Bleiweiss, M. (2017). Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the Rio Grande headwaters. *International Journal of Applied Earth Observations and Geoinformation*, 63, 234-243. doi:10.1016/j.jag.2017.08.007.
- Stewart, I., Cayan, T., & Dettinger, D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a 'Business as Usual' Climate Change Scenario. *Climatic Change*, 62(1), 217-232. doi:10.1023/B:CLIM.0000013702.22656.e8.
- Sturm, M., & Wagner, A. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*, 46(12), N/a. doi:10.1029/2010WR009434.
- Tan, A., Adam, J. C., & Lettenmaier, D. P. (2011). Change in spring snowmelt timing in Eurasian Arctic rivers. *Journal of Geophysical Research: Atmospheres*, 116(D3), N/a. doi:10.1029/2010JD014337.
- Wang, X., & Xie, H. (2009). New methods for studying the spatiotemporal variation of snow cover based on combination products of MODIS Terra and Aqua. *Journal of Hydrology*, 371(1), 192-200. doi:10.1016/j.jhydrol.2009.03.028.
- Zhou, H., Aizen, E., & Aizen, V. (2013). Deriving long term snow cover extent dataset from AVHRR and MODIS data: Central Asia case study. *Remote Sensing of Environment*, 136(C), 146-162. doi:10.1016/j.rse.2013.04.015.

Tables

Table 3.1: Principal component analysis weights for the first principal component. Larger values describe more influence in the first principal component.

<i>Year</i>	<i>Principal Component Weight</i>
2000	0.2250318
2001	0.2114065
2002	0.2385
2003	0.2503311
2004	0.3099047
2005	0.254732
2006	0.1991916
2007	0.2612577
2008	0.2362055
2009	0.2289876
2010	0.2864082
2011	0.2309159
2012	0.2680333
2013	0.1940405
2014	0.2220782
2015	0.2474614
2016	0.2294085

Table 3.2: Total error by snow covered area percentage and year. 45% snow covered area is not shown since no imagery with less than 5% cloud cover was available. A "-" signifies there was no data for that year (column) at that particular SCA (row).

<i>Total Error by Year (Omission and Commission)</i>								
<i>Snow Covered</i>								<i>Average</i>
<i>Area</i>	<i>2001</i>	<i>2004</i>	<i>2007</i>	<i>2011</i>	<i>2013</i>	<i>2017</i>	<i>2018</i>	<i>Total</i>
<i>Percentage</i>								<i>Error</i>
90%	-	-	-	5.37	4.10	-	-	4.73
85%	-	6.97	7.48	-	5.77	-	12.30	8.13
80%	-	-	-	-	-	8.27	14.44	11.35
75%	-	-	8.48	-	13.09	14.99	-	12.19
70%	8.35	-	11.24	-	-	13.78	-	11.12
65%	-	12.49	-	-	13.72	-	11.43	12.55
60%	-	-	-	14.90	-	14.56	-	14.73
55%	7.75	12.02	12.88	-	-	10.25	-	10.73
50%	-	10.30	-	-	-	-	-	10.30
40%	11.52	8.68	11.57	-	-	-	-	10.59
35%	-	-	-	-	-	7.35	-	7.35
30%	-	-	-	9.40	-	8.08	6.71	8.06
25%	5.87	-	-	-	-	5.97	6.63	6.16
20%	-	-	6.38	5.86	8.15	-	5.44	6.46
15%	-	-	-	-	-	4.82	-	4.82
10%	4.22	4.00	-	4.39	3.67	3.30	3.40	3.83
5%	2.42	-	-	2.14	2.39	2.41	-	2.34

Table 3.3: Average omission, commission, and total errors by year. Average error is an average of error at all available SCA for each year.

<i>Year</i>	<i>Average Omission</i>	<i>Average Commission</i>	<i>Total Average Error</i>
2001	4.10%	2.23%	6.34%
2004	5.29%	3.69%	8.98%
2007	5.81%	3.86%	9.67%
2011	3.26%	2.55%	5.80%
2013	3.68%	3.14%	6.82%
2017	4.06%	3.11%	7.17%
2018	3.88%	4.09%	7.97%

Table 3.4: Fraction of omission errors associated with coniferous forest land cover for SCAs between 40 and 80 percent. SCA of 45% was omitted owing to lack of images with sufficient cloud-free area.

<i>Snow Covered Area Percentage</i>	<i>Omission of Snow by Model</i>	<i>Percentage of Pixels with Omission Errors Classified as Coniferous Forest</i>
80	6.61%	86.19%
75	7.40%	89.50%
70	8.26%	86.08%
65	7.84%	75.19%
60	8.92%	89.24%
55	6.42%	79.82%
50	8.38%	89.99%
40	7.40%	88.92%

Table 3.5: Earliest and latest melt-out dates for SNOTEL stations among the 19 years of images included in this study.

<i>SNOTEL Station ID</i>	<i>Latest Melt Out Date</i>	<i>Earliest Melt Out Date</i>
Base Camp	6/10/2011	5/1/2001 and 5/1/2007
Thumb Divide	6/16/2011	4/23/2015
Lewis Lake Divide	7/3/2011	5/13/2015
Snake River Station	6/6/2011	4/29/2015 and 4/29/2007
Two Ocean Plateau	7/18/2011	6/6/2007
Togwotee Pass	7/22/2011	6/6/2007
Phillips Bench	7/6/2011	5/19/2007
Gros Ventre Summit	6/21/2011	5/11/2012
Gunsight Pass	7/6/2011	5/18/2001 and 5/18/2007
Granite Creek	6/14/2011	5/5/2000 and 5/5/2007
East Rim Divide	5/28/2011	4/23/2012

Table 3.6: Correlation matrix comparing the principal components of the First Day of Land analysis (columns) and the difference principal component analysis (rows). The highest correlations are bolded and boxed.

		Principal Components of First Day of Land PCA																
		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17
Principal Components from PCA on Difference Images	PC1	0.43	0.00	-0.38	0.18	-0.12	-0.74	0.03	-0.46	-0.13	0.04	0.11	0.05	0.01	-0.01	0.07	0.10	-0.03
	PC2	0.08	0.42	0.61	0.29	-0.14	0.08	-0.07	-0.30	-0.02	0.12	-0.25	-0.13	-0.20	-0.25	-0.13	-0.14	0.09
	PC3	-0.20	0.17	-0.39	-0.19	0.47	0.38	0.00	0.16	-0.19	0.09	0.06	0.17	0.02	0.32	0.35	0.22	0.18
	PC4	0.02	-0.03	0.47	0.03	0.00	-0.06	0.14	-0.16	-0.12	0.05	0.16	0.34	0.65	0.29	0.22	-0.04	0.06
	PC5	0.11	0.20	-0.18	0.27	-0.27	0.13	0.22	0.17	-0.53	-0.05	-0.30	0.51	-0.09	-0.05	-0.16	0.12	0.07
	PC6	0.06	-0.17	0.05	-0.35	0.36	0.00	0.36	-0.28	0.03	-0.37	-0.03	0.35	-0.09	-0.40	-0.17	-0.21	-0.03
	PC7	0.13	0.16	0.03	0.22	-0.21	0.03	0.42	0.13	0.06	-0.68	0.27	-0.28	-0.07	0.18	0.16	0.00	0.09
	PC8	-0.13	0.16	-0.13	0.12	0.15	0.26	0.60	-0.41	0.17	0.45	0.03	-0.15	-0.15	0.20	-0.01	0.04	0.02
	PC9	-0.02	0.04	0.21	-0.12	0.33	-0.34	0.28	0.37	-0.52	0.19	0.10	-0.36	-0.02	0.01	-0.23	-0.05	-0.09
	PC10	-0.49	0.73	0.03	-0.45	-0.01	-0.29	0.01	0.14	0.29	-0.01	-0.06	0.15	0.00	-0.02	0.06	0.19	0.07
	PC11	0.43	0.15	0.07	0.42	0.67	-0.11	-0.28	-0.11	-0.07	-0.25	-0.23	-0.01	-0.14	0.07	0.29	0.02	0.11
	PC12	0.17	-0.19	0.30	-0.19	-0.06	0.08	0.00	-0.20	-0.17	-0.08	0.02	-0.06	-0.23	-0.06	0.18	0.79	-0.21
	PC13	0.13	0.07	-0.03	-0.42	-0.21	0.12	-0.17	-0.31	-0.47	-0.03	0.08	-0.19	-0.18	0.04	0.31	-0.34	0.36
	PC14	0.33	-0.20	0.21	0.11	-0.05	-0.19	0.13	0.30	0.15	0.27	0.27	0.34	-0.49	-0.10	0.42	-0.13	0.17
	PC15	0.39	-0.31	0.04	-0.12	-0.07	-0.20	0.23	0.10	0.15	0.00	-0.67	-0.17	0.09	0.11	0.09	0.10	0.49
	PC16	0.22	0.08	-0.15	0.15	0.03	0.11	0.10	0.09	-0.03	0.12	0.12	-0.19	0.39	-0.74	0.29	0.15	0.21
	PC17	0.05	0.04	-0.05	-0.06	-0.10	-0.01	0.15	0.06	-0.04	0.01	-0.39	-0.07	0.04	-0.04	0.52	-0.26	-0.68
Proportion of Variance		0.848	0.020	0.015	0.014	0.012	0.011	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.006	0.005	0.004

Table 3.7: Relation table. For each principal component adding 2% additional variance described the related difference principal component is shown with the related year.

Look Up Table

<i>First Day of Land Principal Component (PC)</i>	<i>Sign Direction of Weight</i>	<i>Related Difference PC</i>	<i>Year Related to Difference PC (Same Sign)</i>	<i>Year Related to Difference PC (Different Sign)</i>
PC2	Positive	PC10	2000	2004
PC3	Positive	PC2	2011	2015
PC4	Negative	PC10	2004	2000
PC5	Positive	PC11	2002	2001
PC6	Negative	PC1	2011	2010
PC7	Positive	PC8	2011	2005
PC8	Negative	PC1	2011	2010
PC9	Negative	PC5	2000	2011

Figures

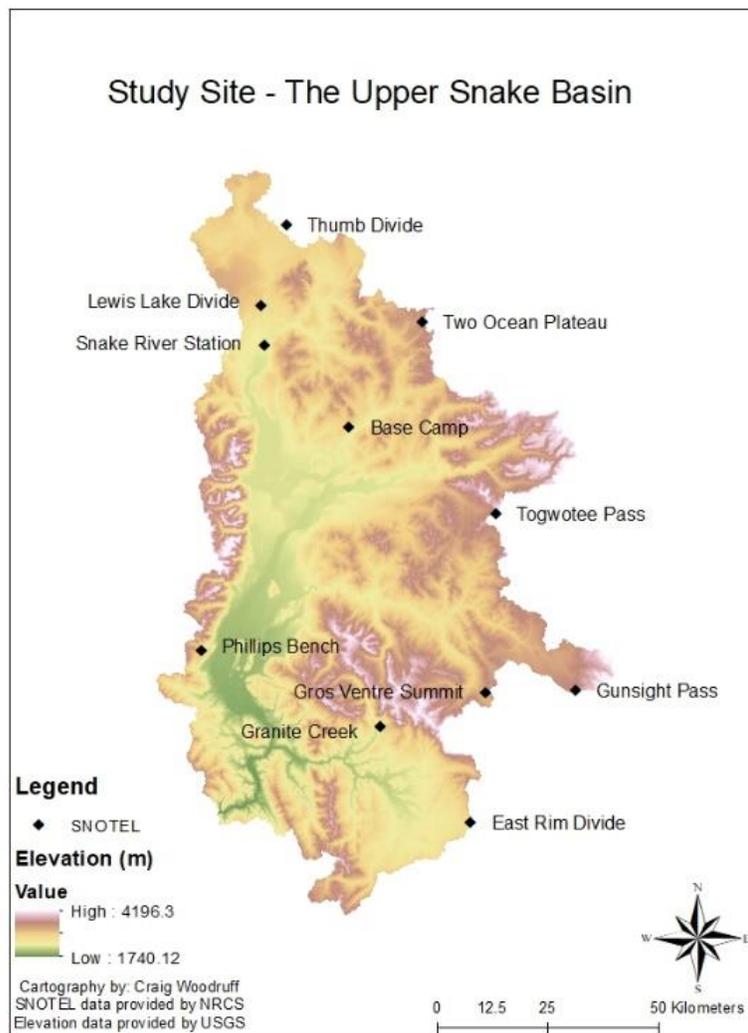


Figure 3.1: Upper Snake Basin with associated SNOTEL stations.

Model Output 2013
Last Day of Snow and First Day of Land
Upper Snake Basin

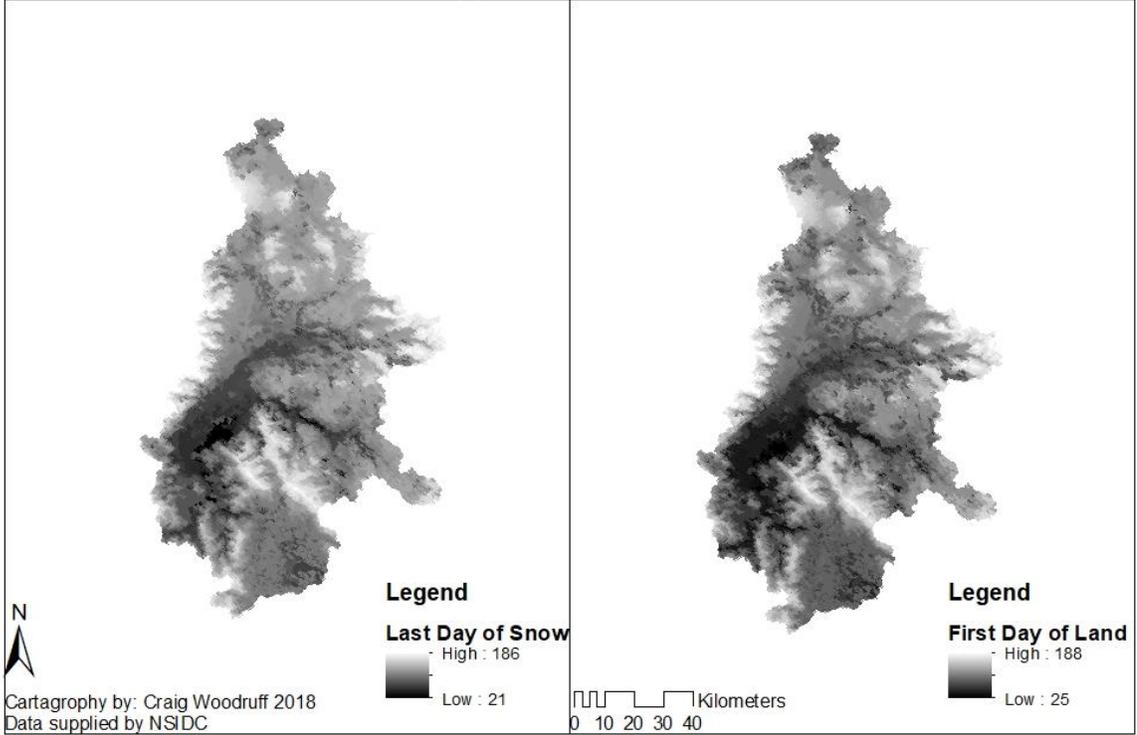


Figure 3.2: Model output for 2013. First day of land right, and last day of snow on the left. Larger DOY values are depicted by white and smaller values by black.

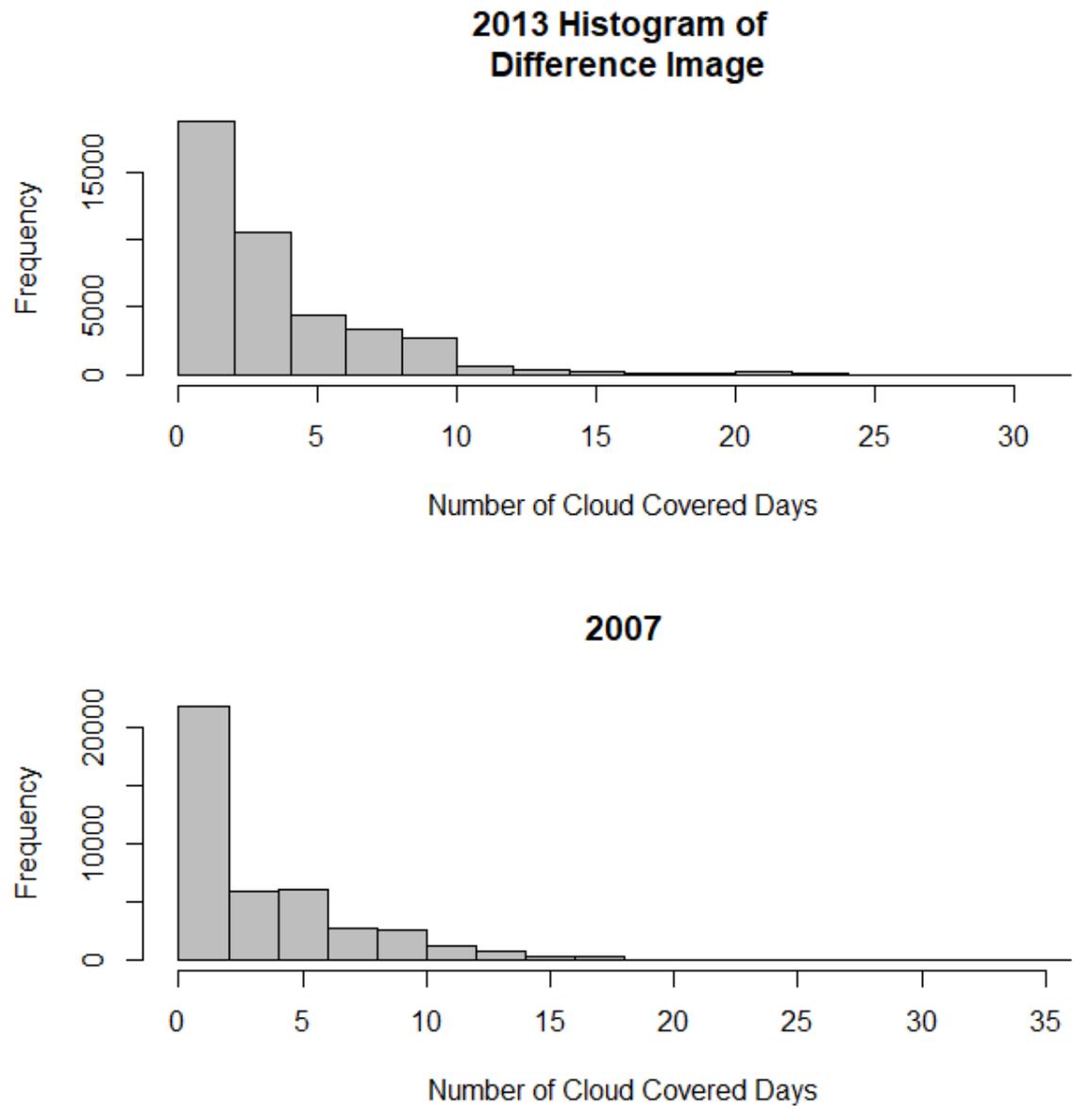


Figure 3.3: Histograms showing distribution of the pixel values for the difference between FDL and LDS images for 2013 and 2007.

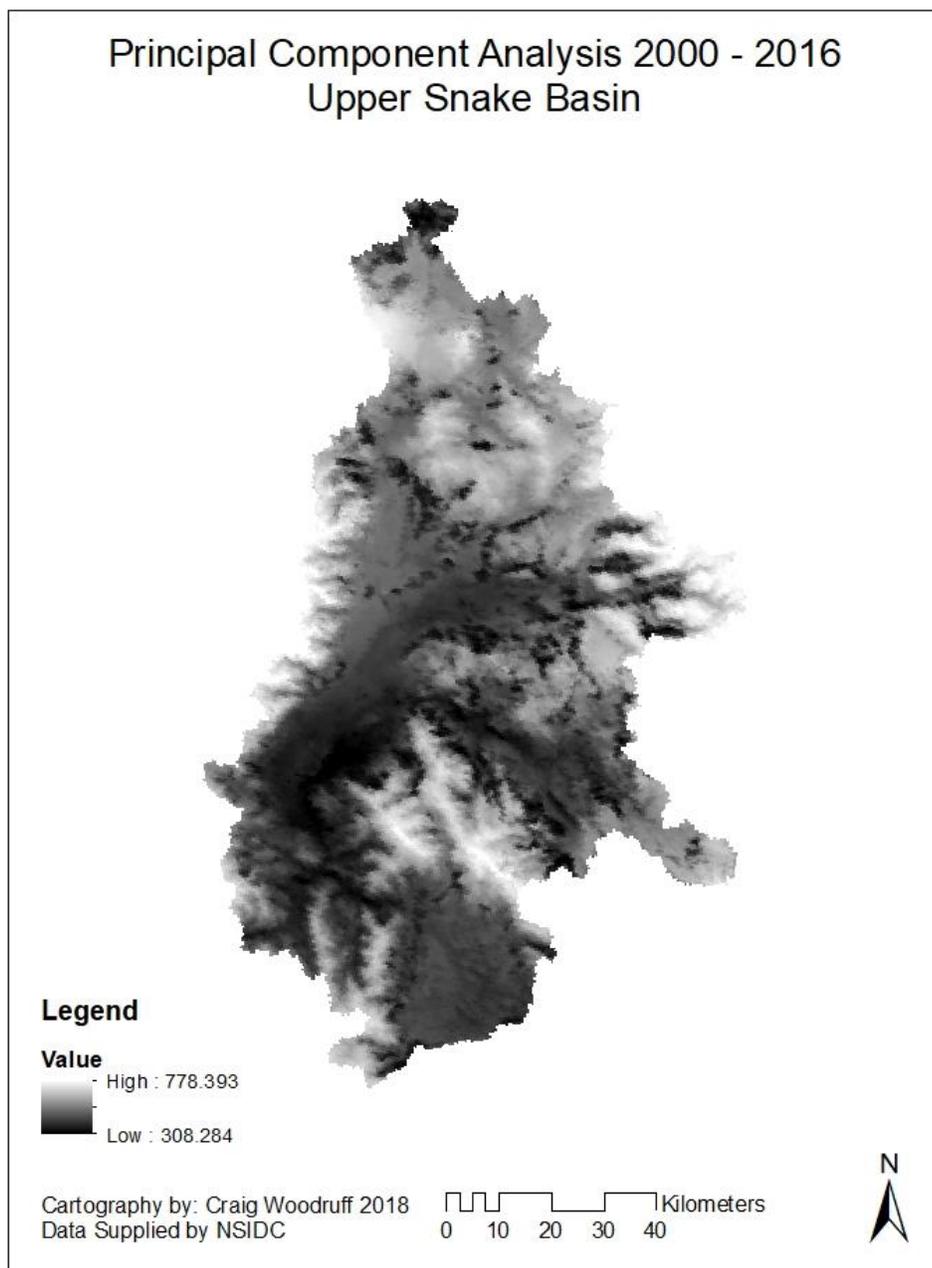


Figure 3.4: Principal component raster output. Values are not significant because they are the product of matrix multiplication. The year to year melting pattern is seen with black indicating early melt and white the latest melt.

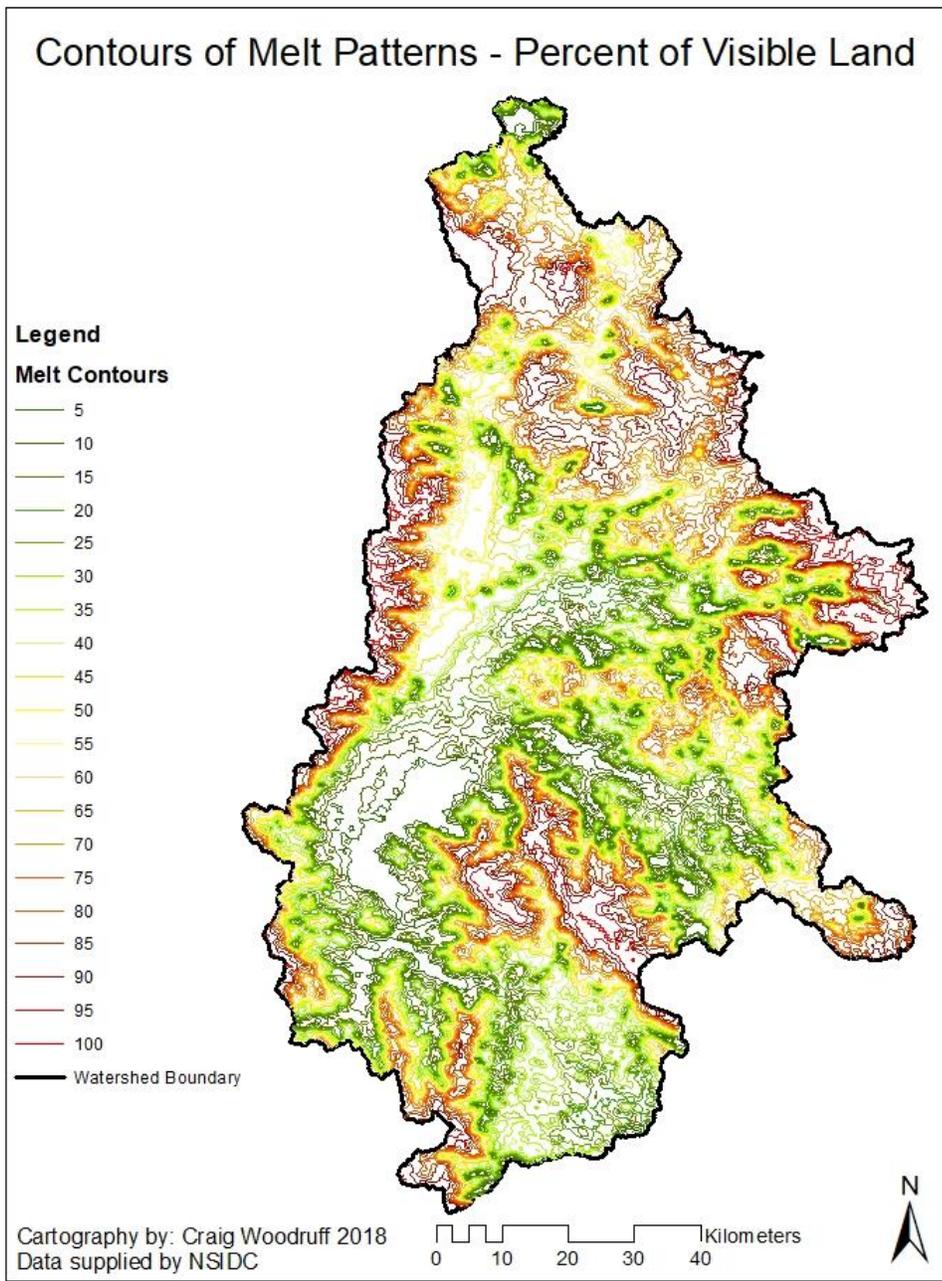


Figure 3.5: Contour map describes intra-annual melting trends 2000 through 2016. Each contour describes the interface between snow and land at different exposed land covered area percentages. SCA is the 100 minus the line value in the legend. Large snow covered areas are depicted in dark green and small areas in dark red.

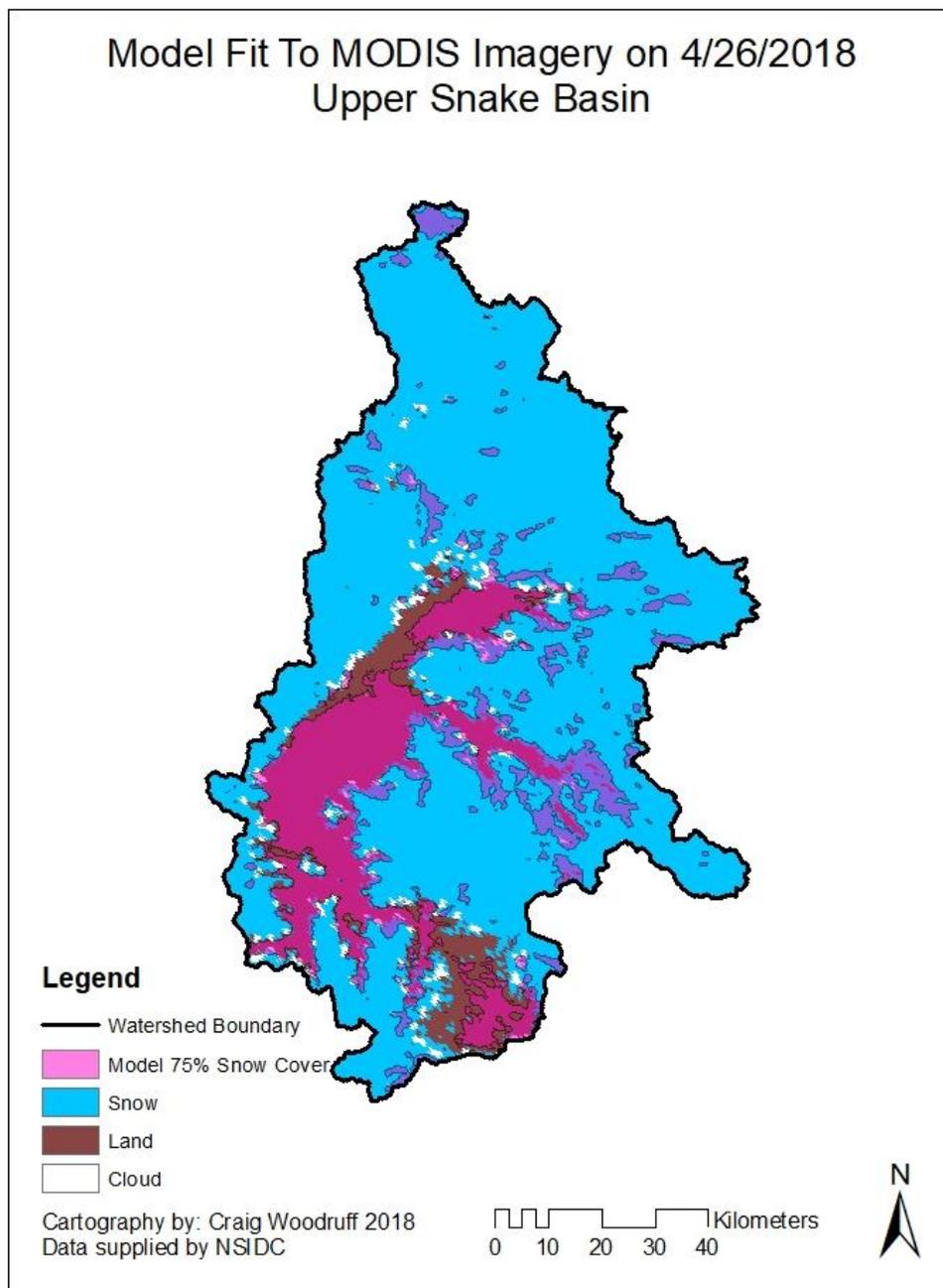


Figure 3.6: Location of the model error at 75% SCA. Imagery from 4/26/2018 was used depicted in blue (snow), brown (land), and white (cloud). The modeled 75% SCA contour is overlaid in pink.

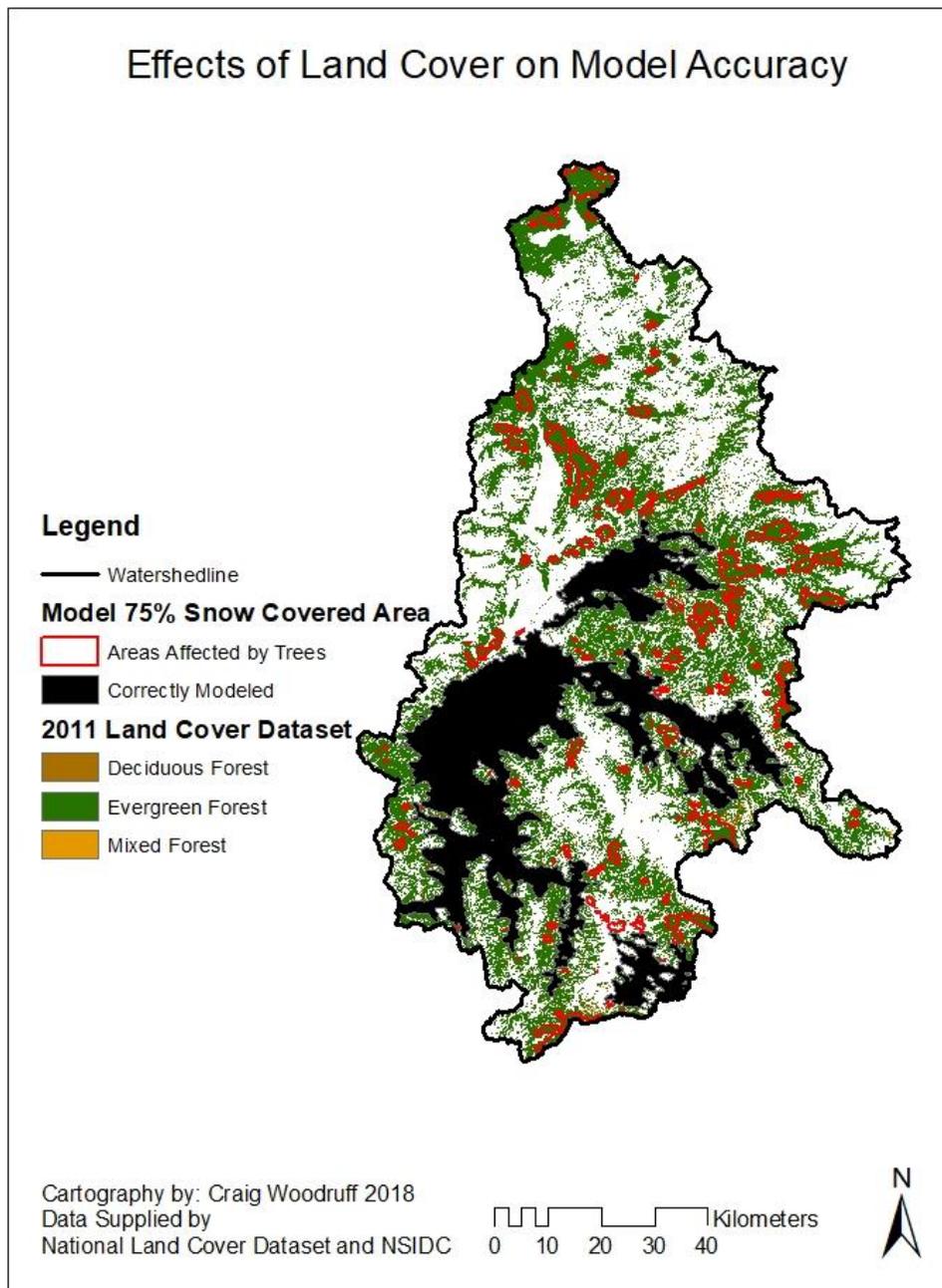


Figure 3.7: Effects of land cover on model accuracy. Forest is depicted in green, brown, and orange (2011 NLCD). Correctly modeled locations from 4/26/2018 are blacked out while incorrectly modeled areas are outlined in red.

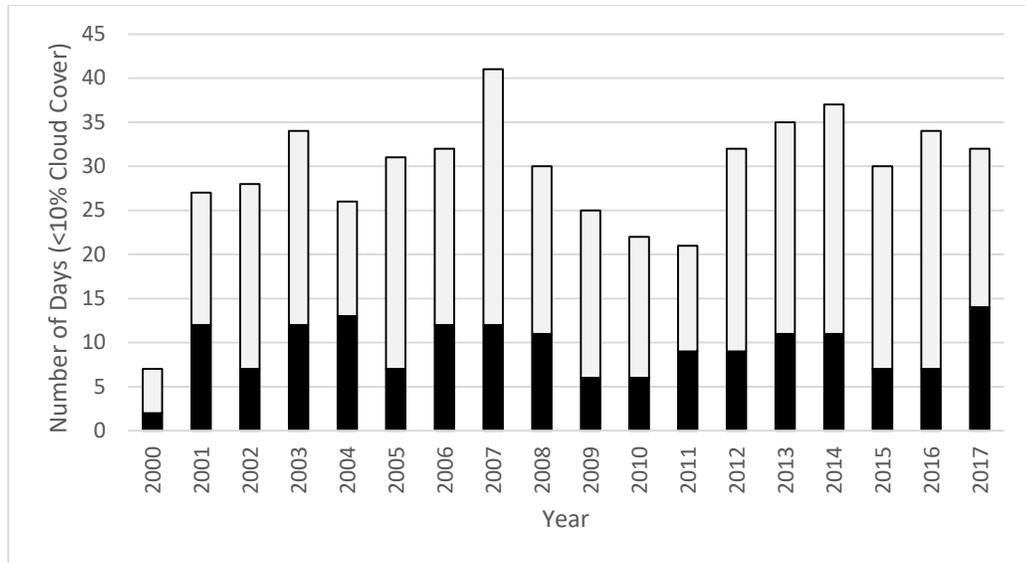


Figure 3.8: Number of days with less than 10% cloud cover. Black indicates days with less than 10% cloud cover during the transition of 80-10% SCA. Black plus white indicates total less than 10% cloud cover days in 150 day period.

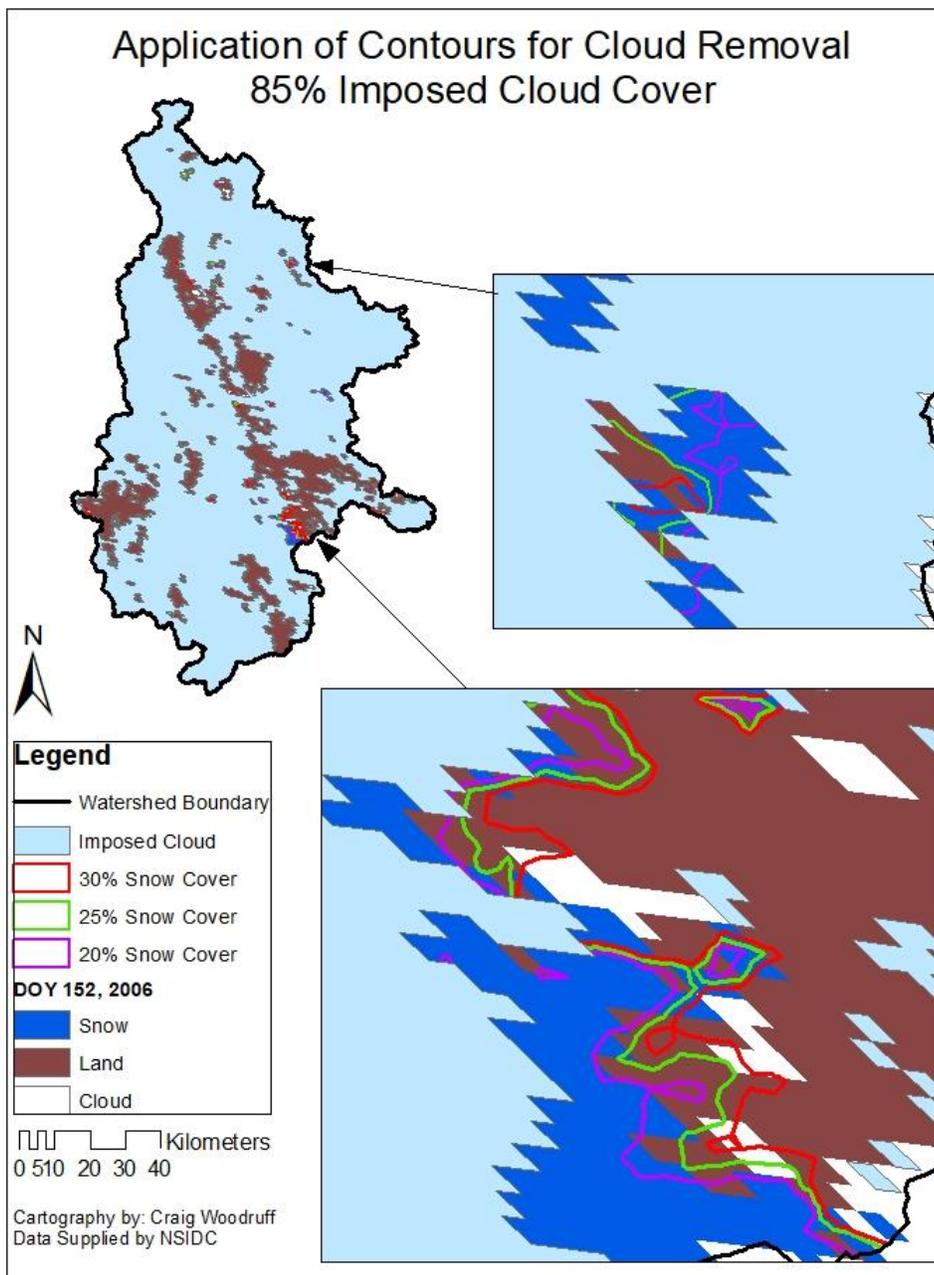


Figure 3.9: Cloud cover proof of concept application. 85% cloud cover imposed over 2006, DOY 152. Dark blue indicates measured snow, brown measured land, white measured cloud, and light blue imposed cloud. Model contours of 20, 25 and 30% SCA are depicted.

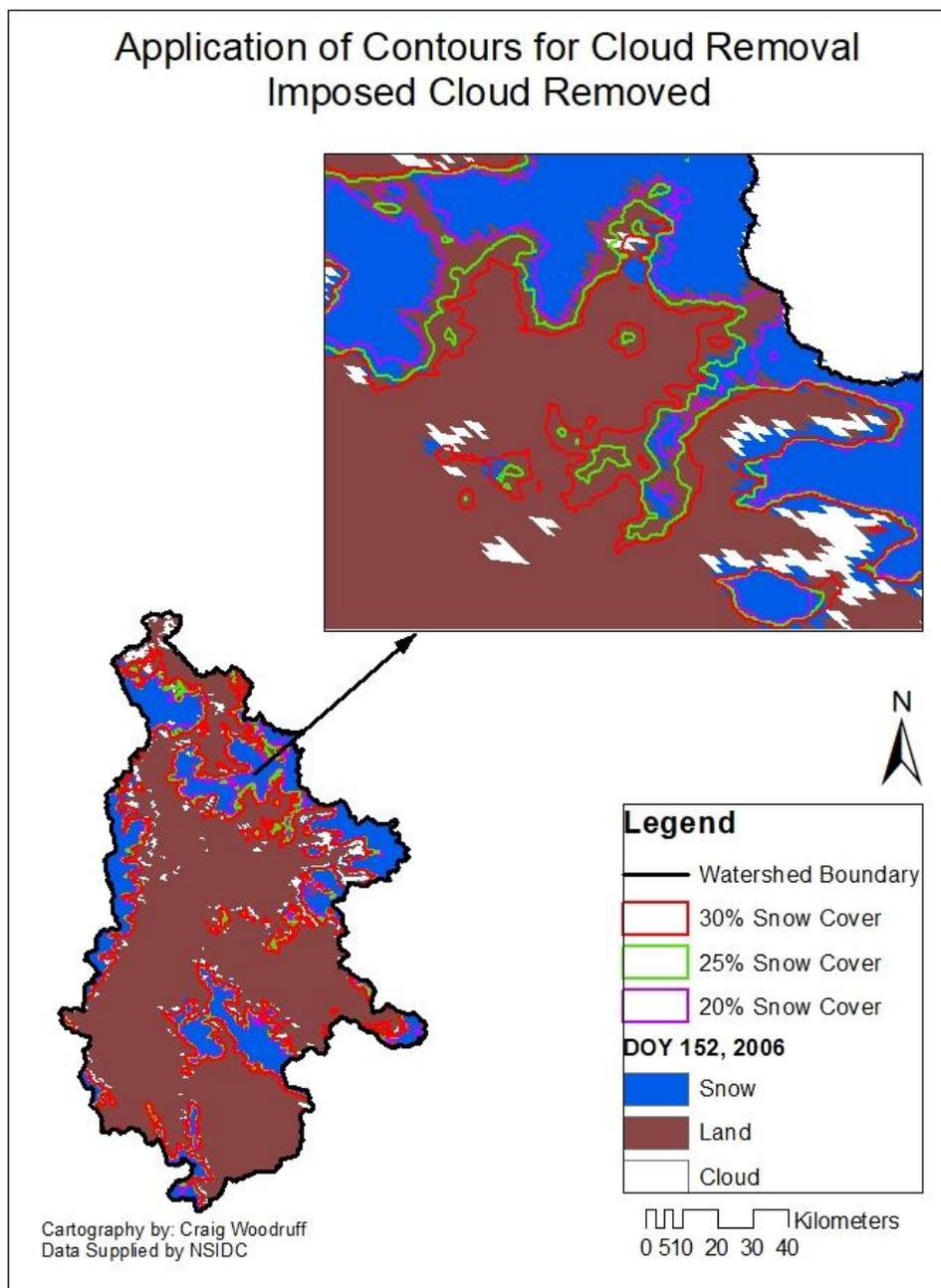


Figure 3.10: Cloud removal proof of concept. Fit of the contours to "cloud" free day. Blue indicates measured snow, brown measured land, and white measured cloud.

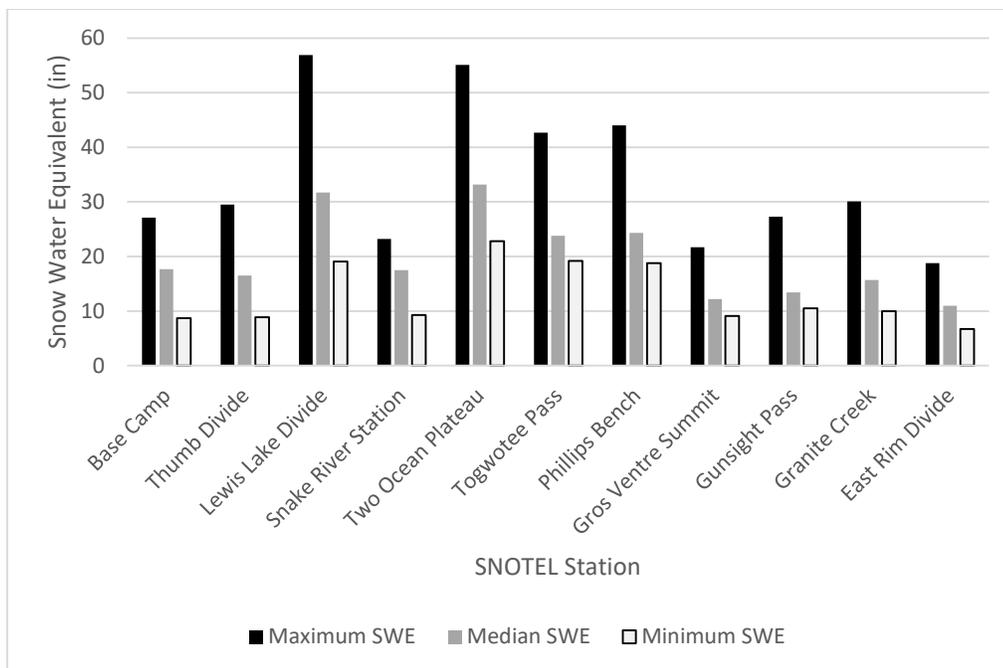


Figure 3.11: Peak SWE measurements by SNOTEL station. Maximum peak SWE measured in a year is depicted in black, median peak SWE over all 19 years in grey, and minimum peak SWE in white.

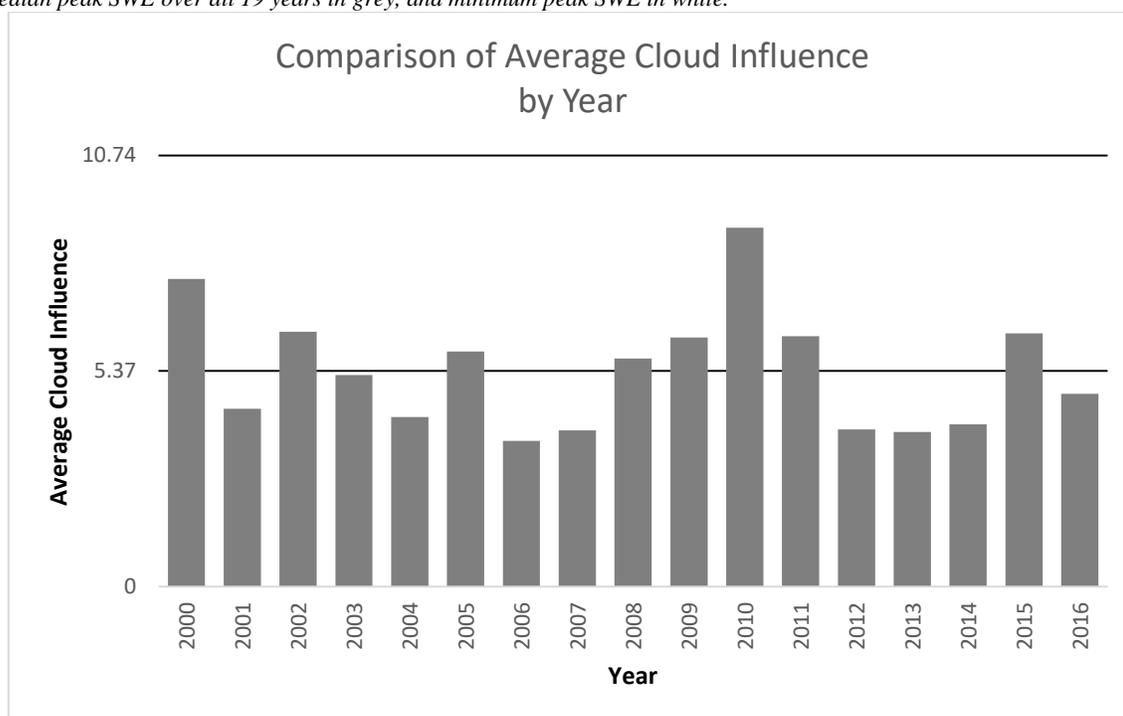


Figure 3.12: Average number of days of cloud influence by year. Cloud influence is defined as the difference between the FDL and LDS images. The average number of days across all years is 5.37 indicated by the horizontal line.

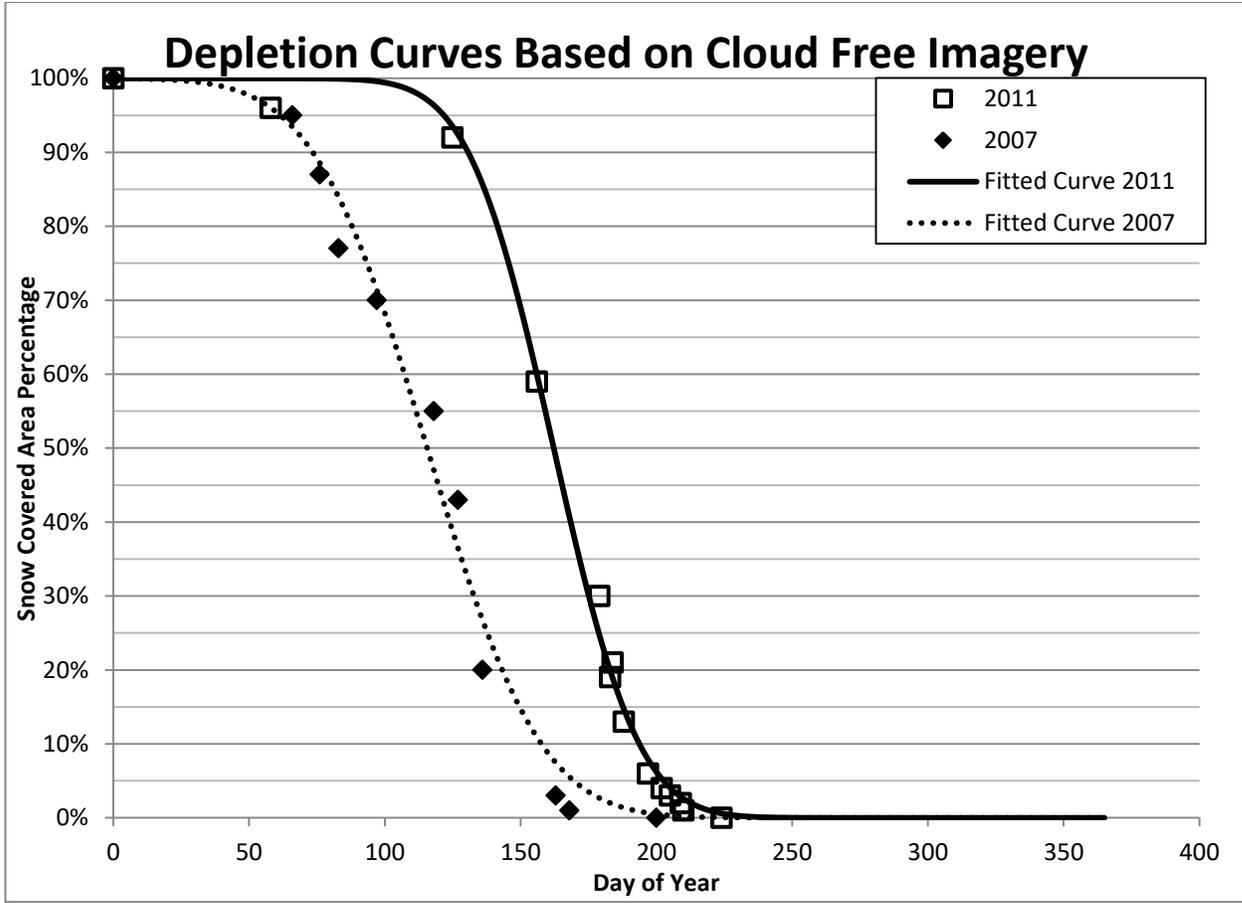


Figure 3.13: Application of contour map throughout two melt seasons. Fitted curves for 2007 and 2011 are shown, fit to the <5% cloud cover days. The curves describe the change in snow covered area with time. Each horizontal gridline of 5% is associated with the 5% melt contours of the spatial model developed.

Chapter 4: Spatial Model Development (Clustering)

Introduction

Snow is an important resource in the Western United States with 50-80% of the annual streamflow coming from snowmelt [Stewart et al., 2004]. With a changing climate the snowmelt dominated watersheds are sensitive to changes in temperature and precipitation which may affect accumulation of snow and therefore storage of water as snow [Barnett et al., 2005]. These changes in temperature and precipitation are vital to understand and prepare for, especially where legal systems such as water rights are built around timing and amount of available water. Remote sensing is a useful tool for mapping the extent of snow as melt progresses, but cloud cover often poses an issue. On cloud free days the Moderate Resolution Imaging Spectroradiometer (MODIS) has shown 93% agreement with in-situ measurements [Hall & Riggs, 2007]. Remote sensing has been used to verify climate model predictions of snow coverage, and is used in current snowmelt models [Curtis et al., 2014; Steele et al., 2017].

Climate modeling is subject to many uncertainties which can be summarized into four components. Greenhouse gas emission scenario, global climate model (GCM), downscaling method, and impact or catchment model [Ficklin et al., 2016]. Modeling at the watershed scale, is of the utmost importance in Southern Idaho where agriculture, power generation, municipalities, and in-stream fish requirements are all reliant on a legal system built around the timing of available water [Qualls et al., 2013]. Preparing for extreme years is likely as important as a long term changes where these legal constructs exist. Minimizing the error in the hydrologic model may lead to more confidence in climate modeling, especially in snowmelt driven watersheds. Poulin et al., [2011] applied two snowmelt models to quantify the uncertainty in hydrologic models using 22 emission scenarios for 2060. A lumped temperature index model and a spatially distributed combination of an energy budget and temperature index model were compared. Both models displayed higher uncertainty during spring runoff and summer months with less uncertainty during the accumulation season. Variability in the GCM can often be biased with precipitation and temperature leading to up to 2°C changes in mean projected temperature and 30% of points when bias correcting precipitation [Pierce et al., 2015]. With the variability and bias that is

introduced in climate modeling by greenhouse gas scenarios and the GCM, it is fundamental to reduce the uncertainty of what we know: the hydrologic models.

The current use for remote sensing in climate change modeling is verification. A number of different sensors have been used to verify surface albedo, snow water equivalent (SWE), and snow extent. Verification spatial resolutions, whether remotely sensed or *in-situ* measurements, often differ greatly from the modeled output [Curtis et al., 2014; Brown et al., 2014; Dickerson-Lange & Mitchell, 2014]. Because the climate models simulate accumulation of snow, the existence of measureable SWE is often used to calculate snow cover fraction [Thiemeßl et al., 2012]. The relationship between SWE and snow cover is non-linear in nature [Luce & Tarboton, 2004]. This complicates the accurate spatial extent of snow if modeling is not correct. This is made even more difficult due to the varied spatial resolution currently available for input data (GCM) and verification methods (remote sensing). Microwave remote sensing is promising because it can detect snow depth and snow water equivalent. The spatial resolution of microwave remote sensing is 25 kilometers and larger [Jensen, 2016]. Typical GCMs do not have sufficient resolution for the mountainous Western United States and watershed scale modeling requires resolutions of ~1-10 kilometers [Lanzante et al., 2018]. If input data at resolutions greater than 10 kilometers is not applicable then verifying models using remotely sensed imagery at spatial resolutions greater than 10 km are not applicable. This discrepancy in spatial resolution in complex terrain and the non-linear relationship between SWE and snow cover fraction hide errors in model output. Downscaling methods have been employed to remove the well-known biases in GCMs due to spatial resolution [Ficklin et al. 2016; Lanzante et al., 2018; Wood et al., 2004]. Even with these corrections many have noted there is less uncertainty in the accumulation period of snow than the melting period [Poulin et al., 2011; Thrasher et al., 2012; Poyck et al., 2011]. As the spatial resolution of the input precipitation and temperature data is refined the hydrologic models will still be a source of error.

Remote sensing of snow offers an untapped resource in climate modelling. While snowmelt timing and snow depth may vary each year, the pattern is the same [Adams, 1976; König & Sturm, 1998]. This pattern is a result of the watershed characteristics. In the arctic it was found that repeated patterns could be used to model snow distribution year to year [Sturm & Wagner, 2010]. Even at a smaller scale intra-annual patterns in snow distribution

were noted in the Reynolds Creek Experimental Watershed in Southern Idaho [Luce & Tarboton, 2004]. The pattern has also been used to develop a dimensionless depletion curve which describes the change in snow cover fraction as melt progress. The curve is dimensionless in time making it applicable to any snow year [Luce et al., 1999]. The key idea that has been overlooked, applied in by Luce et al., [1999], is the shift in the timing and duration of melt. A characteristic map describing snow melt patterns independent of snow depth and timing can be developed using remotely sensed imagery. MODIS offers 19 years of daily imagery at a spatial resolution of 500 meters making derivation of such a map accomplishable. If a map was created it could be used in deterministic models such as the Snowmelt Runoff Model (SRM) [Martinec & Rango, 1986; Rango, 1993]. In this research a characteristic melt map was developed, describing the spatial changes in snow cover in any given year. An integral part to the utility of this characteristic map for climate modeling is an accompanying dimensionless depletion curve. These two components give climate modelers snow cover fraction, spatially represented and the ability to forecast the timing of melt. The dimensionless depletion curve can be given dimensions of time with two parameters. To our knowledge no at this time no research uses the large amount of remotely sensed imagery to predict future scenarios.

Data Sources and Study Site

Study area

Our study area, the Upper Snake Basin Fig. 4.1, covers an area of 3,465 mi² (8,894 km²) and elevation ranges between 5,799 -13,760 feet (1,737 – 4,194 meters). The area serves as an important water resource for Idaho agriculture, power generation, in stream fish requirements, as well as municipal, commercial, and industrial uses. The watershed is bordered to the West by the Teton Mountains and includes the hydrologic units known as the Snake Headwaters, Gros Ventre, and Greys-Hobock. Snow seasonally covers the watershed making remote sensing a useful tool for collecting snowmelt information. The study site also has 11 ground based monitoring stations (SNOTEL sites) which measures a 2 meter by 2 meter snow pillow and collect snow water equivalent (SWE) measurements. The study site land cover is variable. Forest accounts for 47.9 percent of the total area with 46.6 percent classified as evergreen forest. Shrub and grass land account for 45.4 percent of the total area with the remaining 6.7 percent belonging to agriculture, developed land, and

wetlands. Percentages calculated based on the 2011 National Land Cover Database [Homer et al. 2015].

Data Sources

The MODerate Resolution Imaging Spectoradiometer (MODIS) on board the Terra satellite was used. The version six data product, MOD10A1 (MODIS Terra snow cover daily L3 global 500 m Sin GRID V006) distributed by the National Snow and Ice Data Center (NSIDC) was used. This data product comes with the NDSI snow cover, NDSI snow cover basic QA, NDSI snow cover algorithm flags QA, NDSI, and snow albedo data. In this study only the NDSI snow cover was used. MOD10A1 was chosen as it has consistently shown less cloud obstruction [Parajka & Blöschl, 2008]. The calculation of NDSI is a combination of the red band and a middle infrared band. The MODIS sensor aboard Aqua had a failure in the mid infrared band forcing a slightly different NDSI calculation. Since no temporal combination of Terra and Aqua was applied, only MOD10A1 product from the Terra satellite was used. To convert the data to fractional snow covered area (SCA) a threshold of 0.4 was used, as documented in the version six user guide [Hall & Riggs, 2015]. Two MODIS swaths were needed to encompass the study region, h09v04 and h10v04. All available MODIS imagery from 2000 through 2018 was used in this study.

SNOTEL data distributed by the Natural Resource Conservation Service (NRCS) was also used as an input in model development. There are eleven sites in the watershed. Base Camp station ID 314 daily SWE measurements, were used as an input for model accuracy. This station is not the lowest elevation site in the study region at 7060 feet, but over the 19 year study period melt was one of the first stations to melt completely.

Methods

Model Development

To compare melt patterns across multiple years one must first identify the important days. The pattern of melt can be extracted for each year by identifying two days of interest for each pixel. The last day snow (LDS) is seen and the first day land (FDL) is seen on a pixel by pixel basis can describe melt. The two output files are equal in spatial resolution and geolocation to MODIS imagery. It is also essential to capture the melt of the maximum snowpack, rather than secondary snowfall. Secondary snowfall is common in spring time after the main snowpack has begun to melt. SWE measurements from ground based

SNOTEL stations were used to inform the extraction of the LDS and FDL to avoid secondary snowfall. The algorithm processes one year at a time starting each on the day of year (DOY) of maximum SWE. Choosing the SNOTEL station to use is important. A station with high elevation may achieve max SWE at a late date making the model susceptible to missing the peak snow pack at lower elevations. To combat this, a station of lower elevation which achieves max SWE early is ideal. The SNOTEL stations in this watershed all see seasonal snow cover for extended time periods, which may not be true for every watershed. The SNOTEL station Base Camp elevation 7060 feet was used to identify the DOY of maximum SWE. In the model development stage only the years 2000 through 2016 were used.

The model extracts the two days for each pixel, LDS and FDL, by iterating through all the available imagery. This is done one year at a time. A few scenarios can take place as the algorithm works through a year. A pixel identified as snow on the DOY the model starts on, the model begins forward processing, known as the snow protocol. Forward processing means the next day of MODIS imagery is analyzed and the next for only the pixels identified as snow on the first day. As the snow protocol works it saves the LDS in one grid file (raster) and the FDL in another raster until each pixel has a value for the LDS and FDL saved in each raster. For the pixels identified as land on the start day the model back processes, analyzing the previous day of MODIS imagery until snow is seen for each pixel, known as the land protocol. For pixels which were never snow covered, or cloud cover never allowed for snow to be seen, the LDS is saved as DOY zero. The final scenario is pixels identified as cloud on the model start DOY. To handle this, the model begins back processing. As the model back processes if pixels are identified as snow first they are sent to the snow protocol beginning on the start DOY day plus one day since all the images in between have already been analyzed. For pixels identified as land first they are sent to the land protocol. If a pixel is cloud covered for the entire period from the start day to January 1st the model begins forward processing from the start DOY plus one day. For these pixels if land is seen first, the DOY is saved to the raster containing the FDL. The raster containing LDS is saved as a zero for these pixels; to signify the pixels were never seen as snow before they were identified as land.

This method brings in ground based data (DOY of max SWE at a SNOTEL station) to efficiently reduce the large amount of imagery available to only two rasters per year. These raster grids describe the melt patterns and the difference between the FDL raster and LDS raster is the number of days cloud cover obscured the view of the pixel during the time period in which we know melt occurred. These rasters contain melt DOY and confidence in dates on a pixel by pixel basis signified by the difference between the FDL and LDS. This is similar to the temporal method applied by Wang & Xei, [2009] to images. Images are extracted from all the daily imagery available, which describe the dates each pixel transitioned from snow to land. This method produces two cloud free images that describe the melt progression through the melting season using the DOY as the value to describe the melt. The influence of cloud is contained in the difference between the rasters.

The method for iterating through the 366 and less images available for each year takes advantage of the nomenclature to identify MODIS images. The DOY and year are contained in each MODIS image name as distributed by NSIDC. Given the year 2001 and DOY 145 a MODIS image for tile h09v04 would include the following MOD10A1.A20001145.h09v04 followed by addition version information. The first nine characters identify the product, Terra daily 500 meter sin grid, followed by the four character year, three character DOY, and finally the tile identifier. This information was used in the model to accurately save the DOY for the correct year to the LDS and FDL rasters. This allows the algorithm to quickly iterate through imagery. Processing is also sped up by sub-setting pixels in the snow, land, and cloud protocol. Entire MODIS tiles 2400 pixels by 2400 pixels can be run in 4 hours depending on the cloud cover on the start day.

Combining the Data

Each year's melt pattern can be described by the combination of the FDL and LDS rasters. The first day there is faith a pixel has melted is contained in the FDL and it was used to develop the spatial model. The FDL raster for the years 2000 – 2016 were combined using a statistical method to classify pixels into groups. The DOY and coordinates of each pixel for all 17 years were brought into excel and analyzed in the statistical language R. A K-means clustering analysis was used to identify DOY clusters. Clusters were based on the DOY only. The clustering analysis is a multivariate approach which identifies separation of data points in multiple dimensions. Because 17 years were used there were 17 dimensions

the analysis took into account to define each cluster. Much like an x, y, z coordinate scheme puts a data point into 3 dimensions the DOY for each pixel (17 for each pixel) plots the point into 17 dimensional space. For a singular pixel the DOY acts as coordinates in the multidimensional space. It is in this multidimensional space that the K-means clustering analysis partitions pixels into groups.

K-means clustering is a step by step partitioning method which minimizes the Within Group Sum of Squares Error (WGSS). The WGSS is the sum of squares error based on Euclidean distance for each point in a cluster to the centroid. The number of clusters is first identified by the user. The algorithm then selects centroids for each cluster, which is a guess first. The data points are then partitioned into clusters based on distance from the centroid, and the WGSS is calculated. The algorithm then begins optimizing the WGSS by moving the centroids based on the last WGSS. Optimization occurs in a step by step method, moving the centroid, partitioning the data, and calculating the Within Group Sum of Squares Error. The partition with the lowest WGSS is saved. For a complete discussion of the methods behind K-means clustering refer to Everitt, B. & Hothorn, Torsten, [2011]. In essence this clustering is identifying the pattern of snow melt across all the input years. K-means clustering is generally applied to data that shows distinct clustering. At first glance while analyzing this data clusters are not apparent due to the continuous melting of snow rather than a discrete form of data. Only when the clusters are plotted with the pixel coordinates does the pattern emerge. The application of this statistical method to a continuous dataset like this led to the serendipitous finding of the intra-annual snow melt pattern. The results make sense when the idea of a consistent melting pattern independent of snow depth and timing is accepted. This reinforces this concept.

The K-means clustering analysis used to develop the final model used 10 clusters. A difficulty with the K-means approach is it is impossible to select the number of data points the algorithm assigns to each cluster. Clusters are assigned through the optimization of the WGSS and each cluster varies in number of data points belonging to it. Twenty-five iterations were used in the optimization to find the best partitioning with the lowest WGSS. Ideally every cluster would be tested but computers are not capable of this, which means the most optimal clustering may not have been chosen using only 25 iterations. With so many pixels 41,503 a second run may produce slightly different results. The ten groups did not

partition into equal percentages of the total number of pixels, however membership to each cluster was not wildly different. The clusters were ordered so cluster 1 would be considered the first “melt”, or group of pixels which melt first, and cluster 10 are the last pixels to melt. Percentages of area were calculated based on number of pixels in each cluster divided by the total number of pixels. Each cluster could be grouped together and assigned an area percentage. This percentage describes the land visible at that particular stage of melt.

Ideally the spatial model would describe between 1% and 5% snow covered area changes. Since the clustering analysis does not allow the user to set the number of data points included in each cluster (pixels in this case), describing changes in snow covered area with such fine precision is impossible with this dataset. As the number of clusters go up the partitioning does not result in equal membership for each cluster. With 10 clusters the membership was nearly equally distributed so each cluster represented about 10% of the pixels and therefore area. Twenty clusters was attempted, which would hopefully produce 5% by area clusters, however the results were not as promising. One cluster ended up with 10.2% of the area while another ended up with 0.66%. In fact using 20 clusters only produced three clusters between 4-6% by area. For this reason 10 clusters were used. The inability of this method to produce equal percentages of area makes it less applicable for real-time cloud removal during melt. The model is very useful for climate modeling purposes due to the simplicity and quick results when combining the data. A typical k-means clustering analysis for a data set of this size runs in under 10 minutes.

The K-means clustering output was manipulated in ArcGIS to create polygons associated with each percentage. For ease of analysis due to the tools available in ArcGIS, the polygons describe the area that would be considered land at each percentage of snow covered area. The outermost edge of each polygon represents the interface between snow and land at the given snow covered area. Ten polygons were created, each equal to a cluster. These could be compared against cloud free imagery from 2017 and 2018. Since these two years were not used in analysis comparing the model to them shows the model’s ability to forecast spatial coverage of snow in future scenarios. Two types of error were assessed, omission and commission errors. An omission error is a pixel measured as snow in the raw image, but modeled as land. Commission error is a pixel measured as land, but modeled as snow. Total error can then be calculated as the addition of commission and omission errors.

With the limited number of cloud free images, finding an image with the exact snow covered area represented by the model is difficult. The errors can be adjusted to account for the difference between the raw data image snow covered area and the model represented snow covered area. For the adjustment to be applicable the difference in the model snow covered area percentage and the raw data snow covered area percentage must be less than either the omission error or the commission error not the total error.

Results

The final product of this climate model is a map describing melting patterns throughout any given year. The map can be seen as a progression of snowmelt where each percentage describes a later date in a melting season. This map was verified at multiple snow covered areas with the unused years of 2017 and 2018. The snow covered area percentages represented by the map are shown in the final map (Fig. 4.2). Percentages are calculated based on number of pixels in each cluster divided by the total number of pixels. The spatial fit of three snow covered area percentages are shown in Figures 4.3-4.5. Each figure shows a snow covered area percentage with two images one from 2017 and the other from 2018 side by side. Overall agreement with the modeled percentage and the measured snow cover can be seen in two snow years. Areas where the model is correct are shown in dark green. Black indicates measured land modeled as snow (commission error) and light green shows the areas modeled as land, but measured as snow (omission error). To separate out the impact of cloud cover, cloud is shown in red (orange when it is within the modeled land area).

The model fit for 2017 is shown in Table 4.1. The model performs best at 4.73% snow covered area with 1.42% adjusted total error. Adjusted error accounts for the difference between the measured snow covered area for the image and the model snow covered area. Since it is nearly impossible to find an exact match in snow covered area with cloud cover less than 5% an adjustment is needed. To calculate the adjusted error (AE) the difference in snow covered area is calculated. This will either be a commission adjusted error (CAE) or omission adjusted error (OAE). An OAE means measured snow covered area is greater than the model snow covered area. An example of the OAE would be an image where the measure snow covered area is 70.07% but the model can only predict 69.07%, and the OAE would be 1%. The overall error of the fit is calculate and adjusted by the OAE or

CAE. This can only be done if the measured omission error or commission error is greater than the OAE or CAE. In all the images the measured error was greater than the OAE or CAE. The worst fit for 2017 was measured at 69.07% snow covered area with an AE of 12.31%. Agreement between the model and the imagery from 2017 was greater than 90% for all snow covered areas except for 69.07% and 55.65% with agreements between 87.69% and 89.11% respectively. In 2018 agreement was even better (Table 4.2). For all snow covered areas the AE was less than 10% with the best fit at 11.92% snow covered area, an error of 3.43%. At 75.78% snow covered area the error was less in 2018 than 2017 by 1.66%, with an AE of 6.85%. Agreement in 2018 ranged from 93.15% to 96.57%. This is a smaller range than 87.69% to 98.58% measured in 2017. While the model appears to fit better in 2018 it is important to note that certain snow covered areas had less error in 2017 than in 2018 such as 22.05% snow covered area with an AE of 3.82% in 2017 and 5.50% in 2018. This shows the model's ability to accurately predict snow cover independent of year and snow depth.

The fit of the model at different snow covered area percentages was better in certain locations of the watershed than others. At 75.8% the fit is overall good in the southern portion of the watershed, but underestimates and overestimates snow covered area in the northern half of the watershed. Overestimation of snow is seen in the Northwestern quadrant and underestimation in the Northeastern quadrant. As the snow covered area shrinks the error is lessened and more evenly distributed. At 32.8% snow covered area the fit is quite good. The Northern portion of the watershed still experiences some overestimation of snow cover, but the fit along the west and south follows the interface between snow and land. A similar result is achieved in the 22.1% snow covered area. The fit is nearly perfect with some overestimation of snow in the Northern area. The error at all of these snow covered areas is generally in a similar area. Even though the location is the same the extent of the error varies. This may be a result of including 17 years in model development. This is advantageous though since it uses information from large snow years and small snow years. The model accurately represents 2017 a very large snow year as seen in Figure 4.6. Changes in maximum SWE recorded at the SNOTEL station Two Oceans Plateau for 2000 through 2018 are shown. This site often records the largest SWE measurements of all SNOTEL sites in the watershed even though the elevation is not the highest of the sites at 9,240 feet. The

smallest maximum SWE was recorded in 2001 at 22.8 inches and the largest maximum SWE in 2017 at 55.1 inches, and the average over all the years is 35.7 inches. The years used in model development 2000-2016 include below and above and average snow packs, while the verification years are an extreme year (2017) and an average year (2018).

Although this model produces differing groupings of snow covered area, analyzing a snow covered area which falls in between two percentages shown in the model communicates the model's ability to accurately map snow extent. Using the same method for adjusting the error, an image 51.08% measured snow covered area was compared to the two closest model percentages 55.65 and 43.87 (Table 4.3). For the 55.65% modeled extent a high commission error would be expected, and a large omission error for 43.87%. The model performed with good agreement when adjusted. For 55.65 snow cover the commission error was 8.07%, far larger than the omission error of 5.69% and when adjusted the AE was found to be 10.25%. For 43.87% modeled extent the omission error was 10.77% with a commission error of 1.46%. When adjusted the AE was reduced to 8.67%. Since no imagery was available for modeled snow covered area of 55.65% and 43.87% this adjustment shows how the model may perform at those snow cover percentages. This resulted in similar agreement values of 89.75% and 91.33%. This proves the model can accurately depict snow cover percentages in between the model percentages. The model's ability to interpolate is valuable in climate change and snowmelt modeling.

No spatial model for snow covered extent as melt progresses based on remote sensing exists. A common issue in climate modeling is accuracy in snow cover extent as melt progresses. To combat this type of error Curtis et al., [2014] introduced cold air pooling into their model using MODIS imagery to verify the snow extent. This approach used downscaled data and complex methods for introducing cold air pooling to represent the snow cover fraction. They found representing SWE with this method was more accurate, but was often underestimated. They noted lower elevations were more susceptible to changes in April 1st SWE measurements. As shown in Table 4.4 even under current conditions in this watershed the lower elevations are variable in the timing of maximum SWE and melt out while upper elevations are more stable. Using a single date to analyze SWE overlooks the variability in snow packs for each year and the shift in accumulation and melting of watersheds. Depicting the snow cover changes may give a more accurate representation of

SWE. Rather than using SWE to estimate snow cover, a non-linear relationship, using snow cover which progresses in the same way each year to inform SWE may be a more precise way of approaching this rather than complex methods such as introduction of cold air pooling. This issue has been difficult to handle and many approaches to accurately depict snow cover and subsequently streamflow have been attempted.

Poyck et al., [2011] incorporated a highest melt factor day into a degree day/energy balance snowmelt model and 100 meter elevation bands modeled separately to predict streamflow. They found overestimation of winter flows and underestimation of summer flows. While they captured long term averages it is clear that a year with large snowfall late in the year could be difficult to model. Shepherd et al., [2010] experienced similar issues in estimating streamflow with earlier modeled peak streamflow and an R^2 of 0.64 for modeled and historic flows. This in part could be a misrepresentation of albedo due to snow cover an important component for modeling energy exchange [Roesch, A., 2006].

The concept overlooked in many of these approaches is the shift in snow accumulation and melting periods. This shift is especially sensitive to lower elevations in this watershed as seen in Table 4.4. Not only are the values of maximum SWE recorded at each SNOTEL site larger in 2017 than 2018, but the timing of melt differs greatly. The melt period defined here is the number of days between the first day the maximum SWE is measured and the first day the station reaches a value of 0 inches of SWE. While the higher elevation sites such as Two Ocean Plateau and Gunsight Pass are relatively stable with the melt period in 2018 4 days shorter than 2017 and 3 days earlier in 2017 than 2018 for Two Ocean Plateau and Gunsight pass respectively, the lower elevations are far less stable. The melt period was a whole 49 days shorter in 2018 than 2017 at Base Camp. Every station below an elevation of 8000 feet except for Lewis Lake Divide had a melt period shorter by 23 or more days in 2018 than 2017. These shifts in melt period are measured between 6,770 and 9,820 feet while the watershed elevation ranges from 5,709 to 13,768 feet. The beginning of melt is therefore sensitive, and where the higher elevation, later melt, is far more stable. With this large variability in the timing and duration of melt it is clear to see why modeling is difficult. Using the single date for each year may incorrectly underestimate lower elevation maximum SWE and struggle with extreme years. Extreme here meaning in snowpack and/or timing of snowpack accumulation and melt. A spatial model of snow

covered area throughout a melt season helps to accurately geo-locate snow extent and timing of melt.

While the model performs well at smaller snow covered areas it has the greatest errors between 75-50% snow covered areas. A closer look at the land cover reveals the source of the error. To investigate this, the image in 2017 at 69.07% snow covered area (the largest error) was used. Every location the model was incorrect was identified. The 2011 National Land Cover Database image was then clipped to the extent of where the model was incorrect (Fig. 4.7). The areas are nearly all covered by evergreen forest with a small percentage belonging to mixed and deciduous forests and the rest to a mixture of other classifications. Of the area modeled incorrectly 85.7% belongs to evergreen forest. It is well documented that the spectral response in forest areas is different than in open areas, a shortcoming of using a constant threshold in the model development. Varying the spectral response of each pixel to identify snow is the basis for the model known as MODSCAG [Painter et al., 2009]. This combats the issues associated with forest cover by using a number of sensor band combinations (electromagnetic responses) to identify the spectral response on a gradient of snow cover in each pixel. The NDSI is affected by the viewing conditions and landscape features and lower NDSI values in boreal forests are associated with snow cover [Riggs et al., 2017]. While this issue poses a challenge in remote sensing the focus of this research is to derive a spatial model of snow cover rather than the changes in NDSI due to land cover.

The spatial model alone offers a wealth of information, but when combined with a depletion curve the true power of such a model is revealed. The widely used deterministic model SRM uses elevation zones and a depletion curve to inform melt timing in a watershed. The depletion curve describes the disappearance of snow cover with time. Remote sensing has been historically used to develop these curves using the advanced very high resolution radiometer (AVHRR) in as early as 1982 [Baumgartner et al., 1987]. With the new methods to produce cloud free MODIS imagery, it has also been used to develop depletion curves [Steele et al., 2017]. A spatial model that fits any snow year suggests a curve could be developed to fit any year, or a dimensionless depletion curve. This was proposed by Luce et al., [1999] and applied here. Fig 4.8 shows a dimensionless depletion curve for this watershed. The MODIS imagery from 2013 was used to derive this curve as

done in chapter 2. The y axis describes the fractional coverage of snow while the x axis describes the dimensionless z score. The z score is a function of time, X_{50} (the DOY 50% melt is achieved) and σ (rate of melt). By adjusting only X_{50} and σ this depletion curve can be fit to any year in the same way the spatial model can be fit to any year. When the dimensionless depletion curve and spatial model are combined a time component can be derived. Once x_{50} and σ are identified for a given year DOY can be found by Equation 4.1.

$$z = \frac{X_{50} - DOY}{\sigma} \quad \text{Equation 4.1}$$

This watershed can be represented by a Gaussian decay function each year. This may not be applicable to every watershed, but Luce et al., [1999] describes a methodology for deriving a dimensionless depletion curve. This map and dimensionless depletion curve together can be direct input into SRM. While remote sensing has remained a verification technique in climate modeling, this combination gives a method for using spatial information of snow derived at the watershed scale to model future scenarios in some of the most uncertain areas in the United States. Remote sensing also offers information about ungauged watersheds. It has been used in current methods to develop depletion curves in ungauged catchments [Safari et al., 2014; Qui L. et al., 2014]. This gives spatial information in watersheds which have historically relied on interpolation of the area since downscaling of input data cannot be verified.

Discussion

This research presents a spatially distributed model describing snowmelt patterns. The model was applied to verify the fit for 2017 and 2018. Spatial distribution of snow relies on the consistent melting pattern independent of snow depth and timing. Remote sensing of snow has been available since the 1970's with the AVHRR sensor launched in 1979 [Jensen, 2016]. Even with this large amount of data, and more recently daily data with MODIS, no one has used this to develop spatial patterns to model the future. Brown et al., [2014] used a host of remotely sensed information including MODIS imagery in conjunction with statistically downscaled GCM data to model climate scenario predictions for 2003 – 2010. Microwave remote sensing has also been used to verify SWE measurements, but as discussed earlier the spatial resolution of 25 km is far too coarse for the Western United States [Lanzante et al., 2018]. It has been said there is good certainty in the accumulation of snow, but high uncertainty during melt [Poulin et al., 2011; Thrasher et al., 2012; Poyck et

al., 2011]. A model such as the one presented here offers spatially distributed data and a time series in the depletion curve to help reduce this uncertainty during the melting season. The focus of research in snow extent monitoring, MODIS in particular, has been focused on cloud removal and accuracy of snow detection in varying landscapes [Painter et al., 2009; Gafurov et al., 2016; Dong & Menzel, 2016; Li, et al., 2017]. This is an important area of research for long term snow extent monitoring; however none of these techniques are applicable outside of a single year at a time. Many use information from a single year to interpolate missing days, but cannot use the information for the next snow year. This model uses long term patterns in snowmelt to create a spatial model of the stages of snowmelt.

Downscaling of GCM data has been applied to remove biases and increase the spatial resolution of the gridded temperature and precipitation data. While the biases are reduced using these methods non-trivial biases in resolution still exist [Ficklin et al., 2016; Wood et al., 2004]. These biases in spatial resolution during the melt season have been attributed to the incorrect representation of snow albedo, and correctly representing the snow cover fraction is crucial [Roesch, A., 2006; Wood et al., 2004]. Downscaling has been successful at reducing much of the bias in the GCM raw data and refining the spatial resolution of the data. Spatial resolutions have been downscaled to 1/16 of a degree, about 6 kilometers, and as low as 90 meters [Pierce et al., 2014; Brown et al., 2014]. This data is then used in snowmelt models which has often relied on the VIC model at 12 kilometer resolution, and in some cases physically based models at 150 meter resolution [Maurer et al., 2010; Dickerson-Lange & Mitchell, 2014]. Two major issues exist in this approach. Albedo is a source of error, and spatial resolution is too coarse in many cases. The model presented here may be a solution to both issues as it is a spatial representation of snow cover as melt occurs with a resolution of 500 meters which is much finer than most climate models. As the spatial resolution of the input data is refined the correct representation of snow cover may persist especially in complex terrain. A simple approach of using the concept of consistent snowmelt patterns as a means to develop spatial models may be easier and possibly more accurate than adding complexity to the models.

Modeling long term averages may not reveal the true impact of inaccuracies in snow cover. In areas which rely on accurate modeling of runoff timing, such as Southern Idaho, capturing the average is just as important as the extremes for water management and

preparation in a changing climate. It was shown by Gergel et al., [2017] the range in maximum and minimum SWE compared to the mean change in SWE in the Western United States is uncertain. In fact the Upper Snake Basin showed some the highest uncertainty when comparing 10 GCMs. The ability of a climate model to project the extremes, given an emissions scenario, is important for the agricultural needs of Idaho. Extremes in climate scenarios are difficult to handle with the current methods. Watershed characteristics have a large influence that is not well understood in a changing climate on streamflow [Teutschbein et al., 2018]. The model developed in this research is a product of the watershed characteristics. What drives a pixel to melt is the combination of elevation, slope, aspect, accumulation, and other factors which are what determine when a pixel melts in relation to the rest of the watershed. These are built into the model even though they are not specifically identified making it superior to models which rely on elevation alone.

Conclusions

Hydrologic modeling of snow has been aided by remote sensing since the late 1970's. While remote sensing has remained a verification technique in climate change modeling, there is opportunity to use this spatial information to inform future scenarios. In this research a spatially distributed model describing snow melt was developed. The model's ability to accurately represent snow covered area in future scenarios was assessed. One concept, consistent melt patterns, was used as the basis of the model development. It has been shown melting patterns are consistent across multiple years [Adams, 1976; König & Sturm, 1998; Luce & Tarboton, 2004; Sturm & Wagner, 2010]. The melt patterns can also be represented by a dimensionless depletion curve [Luce et al., 1999]. With the large amount of daily imagery available which measure snow covered area the model could be derived. The statistical approach allows the data from all the years included to organize into clusters independent of location. The model performed with 87.7-98.6% agreement in the two years. The combination of the dimensionless depletion curve and the spatial information is a new source of data that may help to accurately model snow melt.

This model also can provide information in ungauged watersheds. In current climate change modeling scenarios these ungauged areas rely on interpolation often at large spatial resolutions. With no information to help in predicting the climate change impacts on these watersheds errors may go unnoticed on the local level but show up in the streamflow as

gauged watersheds contribute. This model offers information which is currently not available at a spatial resolution adequate for representing complex terrain. The spatial and temporal information this model offers may fundamentally change the way modeling climate change scenarios is conducted. The model is accurate independent of snow year and timing giving climate modelers the ability to capture extremes and average years, a difficult task with the current methods. Water management relies on accurate modeling of all scenarios and this information may help areas such as Southern Idaho, which have high uncertainty in the future changes in SWE in the Upper Snake Basin and rely on a legal system for water management, prepare for the future.

References

- Adams, W. P. (1976). Areal differentiation of snow cover in east central Ontario, *Water Resources Research*, 12(6), 1226–1234, doi:10.1029/WR012i006p01226.
- Barnett, T. P., Adams, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438, 303-309.
- Baumgartner, M., Seidel, K., & Martinec, J. (1987). Toward Snowmelt Runoff Forecast Based on Multisensor Remote-Sensing Information. *Geoscience and Remote Sensing, IEEE Transactions on, GE-25*(6), 746-750.
- Brown, M. E., Racoviteanu, A. E., Tarboton, D. G., Gupta, A., Nigro, J., Policelli, F., et al., (2014). An integrated modeling system for estimating glacier and snow melt driven streamflow from remote sensing and earth system data products in the Himalayas. *Journal of Hydrology*, 519(PB), 1859-1869. doi:10.1016/j.jhydrol.2014.09.050.
- Curtis, J. A., Flint, L.E., Flint, A. L., Lundquist, J. D., Hudgens, B., Boydston, E. E., & Young, J. K. (2014). Incorporating Cold-Air Pooling into Downscaled Climate Models Increases Potential Refugia for Snow-Dependent Species within the Sierra Nevada Ecoregion, CA. *PLoS ONE*, 9(9), E106984.
- Dickerson-Lange, S. E., & Mitchell, R. (2014). Modeling the effects of climate change projections on streamflow in the Nooksack River basin, Northwest Washington. *Hydrological Processes*, 28(20), 5236-5250. doi:10.1002/hyp.10012.
- Dong, C., & Menzel, L. (2016). Producing cloud-free MODIS snow cover products with conditional probability interpolation and meteorological data. *Remote Sensing of Environment*, 186, 439.
- Everitt, B. & Hothorn, Torsten, (2011). *An introduction to applied multivariate analysis with R*, New York: Springer.
- Ficklin, D., Abatzoglou, J., Robeson, S., & Dufficy, A. (2016). The Influence of Climate Model Biases on Projections of Aridity and Drought. *Journal Of Climate*, 29(4), 1269-1285. doi:10.1175/JCLI-D-15-0439.1.
- Gafurov, A., Lüdtke, S., Unger-Shayesteh, K., Vorogushyn, S., Schöne, T., Schmidt, S., Kalashnikova, O., & Merz, B. (2016). MODSNOW-Tool: An operational tool for daily snow cover monitoring using MODIS data. *Environmental Earth Sciences*, 75(14), 1-15. doi:10.1007/s12665-016-5869-x.
- Gergel, D., Nijssen, B., Abatzoglou, J., Lettenmaier, D., & Stumbaugh, M. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, 141(2), 287-299. doi:10.1007/s10584-017-1899-y.

- Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12), 1534-1547. doi:10.1002/hyp.6715.
- Hall, D.K. & Riggs, G. A. (2015). *MODIS Snow Products Collection 6 User Guide*. doi: <http://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf>. [July 12, 2018].
- Hall, D. K. & Riggs, G. A. 2016. *MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6*. [NDSI snow cover]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: <https://doi.org/10.5067/MODIS/MOD10A1.006>. [January 2017 - March 2018].
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., (2015), [Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information](#). *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354.
- Jensen, J. (2016). *Introductory Digital Image Processing: A Remote Sensing Perspective* (4th ed.).
- König, M. & Sturm, M. (1998). Mapping snow distribution in the Alaskan Arctic using aerial photography and topographic relationships, *Water Resour. Res.*, 34(12), 3471–3483, doi:10.1029/98WR02514.
- Lanzante, J. R., Dixon, K. W., Nath, M. J., Whitlock, C. E., & Adams-Smith, D. (2018). SOME PITFALLS IN STATISTICAL DOWNSCALING OF FUTURE CLIMATE: Statistical downscaling is generally successful and used widely to refine projections of future climate, though in some circumstances it can lead to highly erroneous results. *Bulletin of the American Meteorological Society*, 99(4), 791-803. doi:10.1175/BAMS-D-17-0046.1.
- Li, X., Fu, W., Shen, H., Huang, C., & Zhang, L. (2017). Monitoring snow cover variability (2000–2014) in the Hengduan Mountains based on cloud-removed MODIS products with an adaptive spatio-temporal weighted method. *Journal of Hydrology*, 551, 314-327. doi:10.1016/j.jhydrol.2017.05.049.
- Luce, C. H., Tarboton, D. G., Cooley, K. R., Hardy, J., Albert, M., & Marsh, P. (1999). Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model. *Hydrological Processes*, 13(12-13), 1921-1933. doi:10.1002/(SICI)1099-1085(199909)13:12/13<1921::AID-HYP867>3.0.CO2.-S.
- Luce, C. H., & Tarboton, D. G. (2004). The application of depletion curves for parameterization of subgrid variability of snow. *Hydrological Processes*, 18(8), 1409-1422. doi:10.1002/hyp.1420.

- Martinec, J., & Rango, A. (1986). Parameter values for snowmelt runoff modelling. *Journal of Hydrology*, 84(3), pp.197–219. doi:10.1016/0022-1694(86)90123-X.
- Maurer, E., Hidalgo, H., Das, T., Dettinger, M., & Cayan, D. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, 14(6), 1125-1138.
- Painter, T. H., Rittger, K., Mckenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4), 868-879. doi:10.1016/j.rse.2009.01.001.
- Parajka, J., & Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. *Water Resources Research*, 44(3), N/a. doi:10.1029/2007WR006204.
- Pierce, D., Cayan, D., & Thrasher, B. (2014). Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. *Journal of Hydrometeorology*, 15(6), 2558-2585. doi:10.1175/JHM-D-14-0082.1.
- Pierce, D., Cayan, D., Maurer, E., Abatzoglou, J., & Hegewisch, K. (2015). Improved Bias Correction Techniques for Hydrological Simulations of Climate Change. *Journal of Hydrometeorology*, 16(6), 2421-2442. doi:10.1175/JHM-D-14-0236.1.
- Poulin, A., Brissette, F., Leconte, R., Arsenault, R., & Malo, J. (2011). Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of Hydrology*, 409(3), 626-636. doi:10.1016/j.jhydrol.2011.08.057.
- Poyck, S., Hendrikx, J., McMillan, H., Hreinsson, E., & Woods, R. (2011). Combined snow and streamflow modelling to estimate impacts of climate change on water resources in the Clutha River, New Zealand. *Journal of Hydrology (New Zealand)*, 50(2), 293.
- Qiu, L., You, J., Qiao, F., & Peng, D. (2014). Simulation of snowmelt runoff in ungauged basins based on MODIS: A case study in the Lhasa River basin. *Stochastic Environmental Research and Risk Assessment*, 28(6), 1577-1585. doi:10.1007/s00477-013-0837-4.
- Qualls, R. J., Taylor, R. G., Hamilton, J., & Arogundade, A. B. (2013) Climate change opportunities for Idaho's irrigation supply and deliveries. *Journal of Natural Resources Policy Research*, 5(2-3), 91-105, DOI: 10.1080/19390459.2013.811856.
- Rango, A. (1993). II. Snow hydrology processes and remote sensing. *Hydrological Processes*, 7(2), 121-138. doi:10.1002/hyp.3360070204.
- Riggs, G.A., Hall, D. K., & Roman, M. O. (2017). Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover Earth System Data

- Records. *Earth System Science Data*, 9(2), pp.765–777. doi:10.5194/essd-9-765-2017.
- Roesch, A. (2006). Evaluation of surface albedo and snow cover in AR4 coupled climate models. *Journal of Geophysical Research: Atmospheres*, 111(D15), N/a. doi:10.1029/2005JD006473.
- Safari Shad, M., Habibnejad Roshan, M., & Ildoromi, A. (2014). Integration of the MODIS Snow Cover Produced Into Snowmelt Runoff Modeling. *Journal of the Indian Society of Remote Sensing*, 42(1), 107-117. doi:10.1007/s12524-013-0279-y.
- Shepherd, A., Gill, K. M., & Rood, S. B. (2010). Climate change and future flows of Rocky Mountain rivers: Converging forecasts from empirical trend projection and down-scaled global circulation modelling. *Hydrological Processes*, 24(26), 3864-3877. doi:10.1002/hyp.7818.
- Steele, C., Dialesandro, J., James, D., Elias, E., Rango, A., & Bleiweiss, M. (2017). Evaluating MODIS snow products for modelling snowmelt runoff: Case study of the Rio Grande headwaters. *International Journal of Applied Earth Observations and Geoinformation*, 63, 234-243. doi:10.1016/j.jag.2017.08.007.
- Stewart, I., Cayan, T., & Dettinger, D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a 'Business as Usual' Climate Change Scenario. *Climatic Change*, 62(1), 217-232. doi:10.1023/B:CLIM.0000013702.22656.e8.
- Sturm, M., & Wagner, A. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*, 46(12), N/a. doi:10.1029/2010WR009434.
- Teutschbein, C., Grabs, T., Laudon, H., Karlsen, R. H., & Bishop, K. (2018). Simulating streamflow in ungauged basins under a changing climate: The importance of landscape characteristics. *Journal of Hydrology*, 561, 160-178. doi:10.1016/j.jhydrol.2018.03.060.
- Themeßl, M., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, 112(2), 449-468. doi:10.1007/s10584-011-0224-4.
- Thrasher, B., Maurer, E., McKellar, C., & Duffy, P. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, 16(9), 3309-3314. doi:10.5194/hess-16-3309-2012.
- Wang, X., & Xie, H. (2009). New methods for studying the spatiotemporal variation of snow cover based on combination products of MODIS Terra and Aqua. *Journal of Hydrology*, 371(1), 192-200. doi:10.1016/j.jhydrol.2009.03.028.

Wood, A., Leung, L., Sridhar, V., & Lettenmaier, D. (2004). Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Climatic Change*, 62(1), 189-216. doi:10.1023/B:CLIM.0000013685.99609.9e.

Tables

Table 4.1: Model fit to the 2017 imagery. Error type is identified as well as the snow covered area represented by the model and the measured percentage in the raw imagery. Expected errors are the error type due to the difference in modelled and measured snow covered area.

2017						
<i>Modelled</i> Snow Covered Area (%)	<i>Omission</i> Error (OE)	<i>Commission</i> Error (CE)	<i>Total</i> Error	<i>Measured</i> Snow covered Area (%)	<i>Expected</i> Error Type	<i>Adjusted</i> Error (AE)
75.78	5.49	5.70	11.20	73.09	CE	8.51
69.07	9.05	5.07	14.12	70.88	OE	12.31
55.65	6.49	4.41	10.90	55.63	CE	10.89
32.78	5.78	2.12	7.90	34.90	OE	5.78
22.05	4.36	1.84	6.21	24.44	OE	3.82
11.92	2.37	1.77	4.15	12.53	OE	3.53
4.73	2.19	0.28	2.48	5.78	OE	1.42

Table 4.2: Model fit to the 2017 imagery. Error type is identified as well as the snow covered area represented by the model and the measured percentage in the raw imagery. Expected errors are the error type due to the difference in modelled and measured snow covered area.

2018						
<i>Modelled</i> Snow Covered Area (%)	<i>Omission</i> Error (OE)	<i>Commission</i> Error (CE)	<i>Total</i> Error	<i>Measured</i> Snow covered Area (%)	<i>Expected</i> Error Type	<i>Adjusted</i> Error (AE)
89.31	3.92	5.55	9.47	85.89	CE	6.05
75.78	6.57	2.50	9.07	78.00	OE	6.85
32.78	2.58	5.68	8.26	28.34	CE	3.82
22.05	3.82	2.61	6.42	22.97	OE	5.50
11.92	1.69	2.75	4.45	10.90	CE	3.43

Table 4.3: Errors associated with an image with snow covered area in between two modelled snow covered areas.

<i>Modelled Snow Covered Area (%)</i>	<i>Omission Error (OE)</i>	<i>Commission Error (CE)</i>	<i>Total Error</i>	<i>Measured</i>			
				<i>Snow covered Area (%)</i>	<i>Expected Error</i>	<i>Adjusted CE and OE</i>	<i>Adjusted Total Error</i>
55.65	5.69	8.07	13.76	51.08	CE	3.51	10.25
43.87	10.77	1.46	12.22	51.08	OE	3.56	8.67

Table 4.4: Changes in SWE and melting periods for all eleven SNOTEL stations in the watershed for 2017 and 2018.

<i>SNOTEL Station ID</i>	<i>Elevation (ft)</i>	<i>Max</i>	<i>Max</i>	<i>Melt</i>	<i>Melt</i>	<i>Melt</i>	<i>Max SWE Difference (in)</i>
		<i>SWE 2017 (in)</i>	<i>SWE 2018 (in)</i>	<i>Period 2017 (days)</i>	<i>Period 2018 (days)</i>	<i>Period Difference (days)</i>	
Base Camp	7060	27.1	19.8	81	32	49	7.3
Thumb Divide	7980	22.9	21.7	56	33	23	1.2
Lewis Lake Divide	7850	48.7	45.6	48	51	-3	3.1
Snake River Station	6920	21.8	19.4	64	37	27	2.4
Two Ocean Plateau	9240	55.1	43.6	66	62	4	11.5
Togwotee Pass	9580	37	36.1	58	67	-9	0.9
Phillips Bench	8200	42	33.2	48	50	-2	8.8
Gros Ventre Summit	8750	21.7	17.5	38	45	-7	4.2
Gunsight Pass	9820	26.3	19.3	49	52	-3	7
Granite Creek	6770	30.1	18.8	75	40	35	11.3
East Rim Divide	7930	18.8	13	58	28	30	5.8

Figures

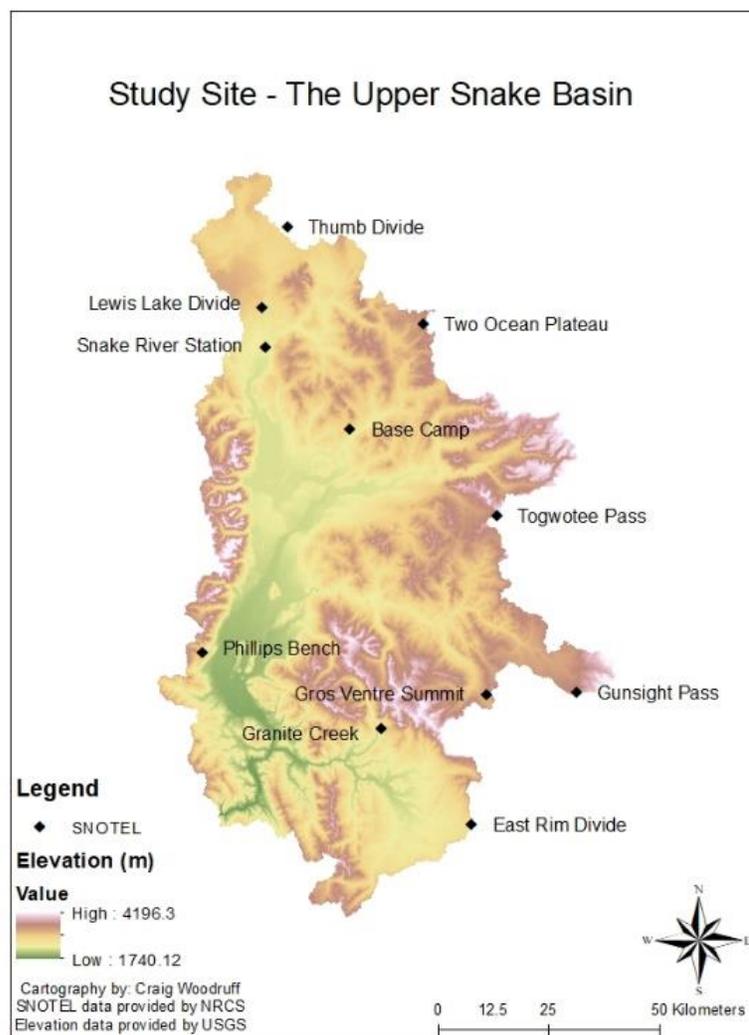


Figure 4.1: Upper Snake Basin with associated SNOTEL stations.

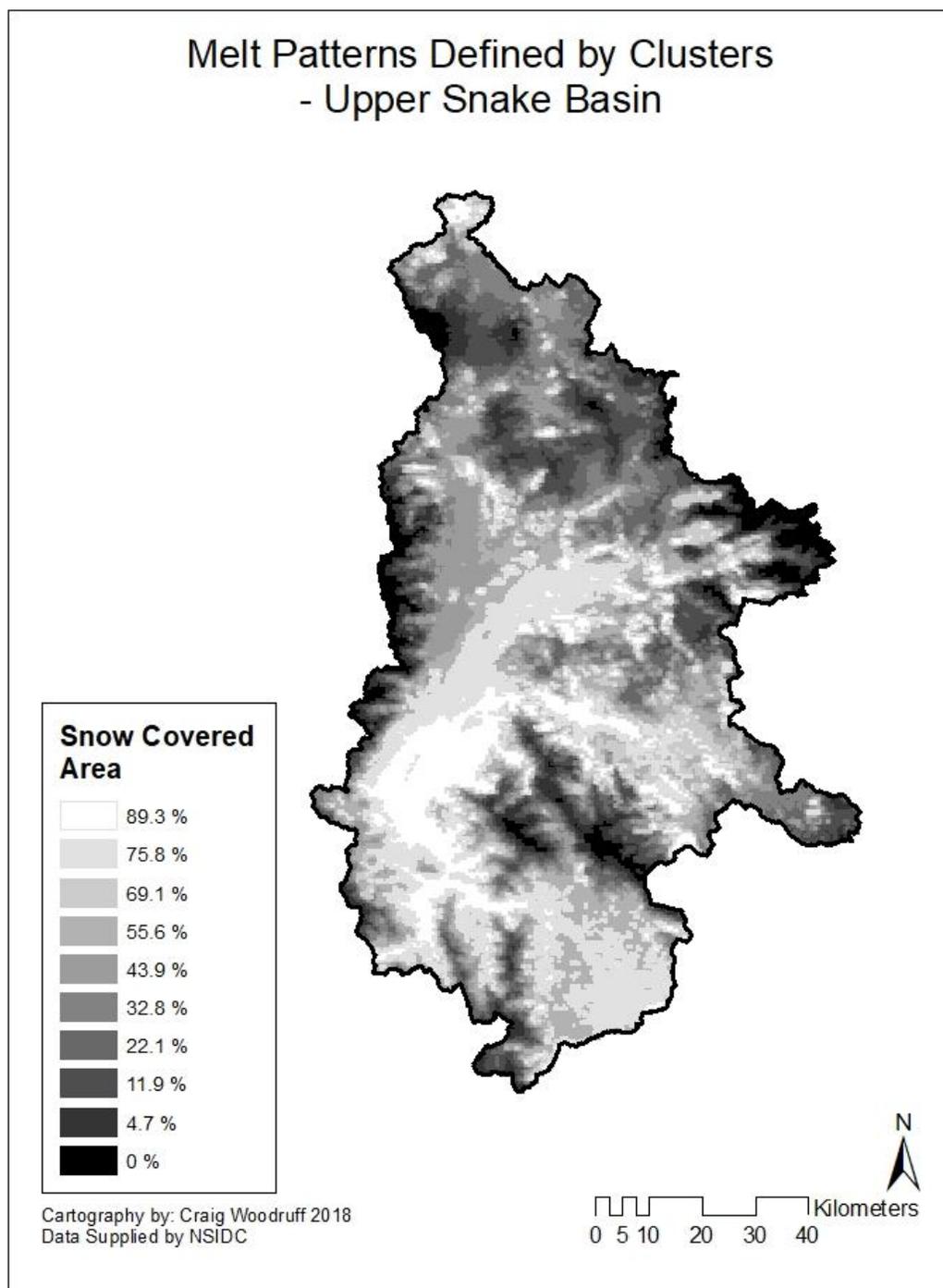


Figure 4.2: Final model of snow covered areas. Percentages of snow covered area are shown in a gray scale.

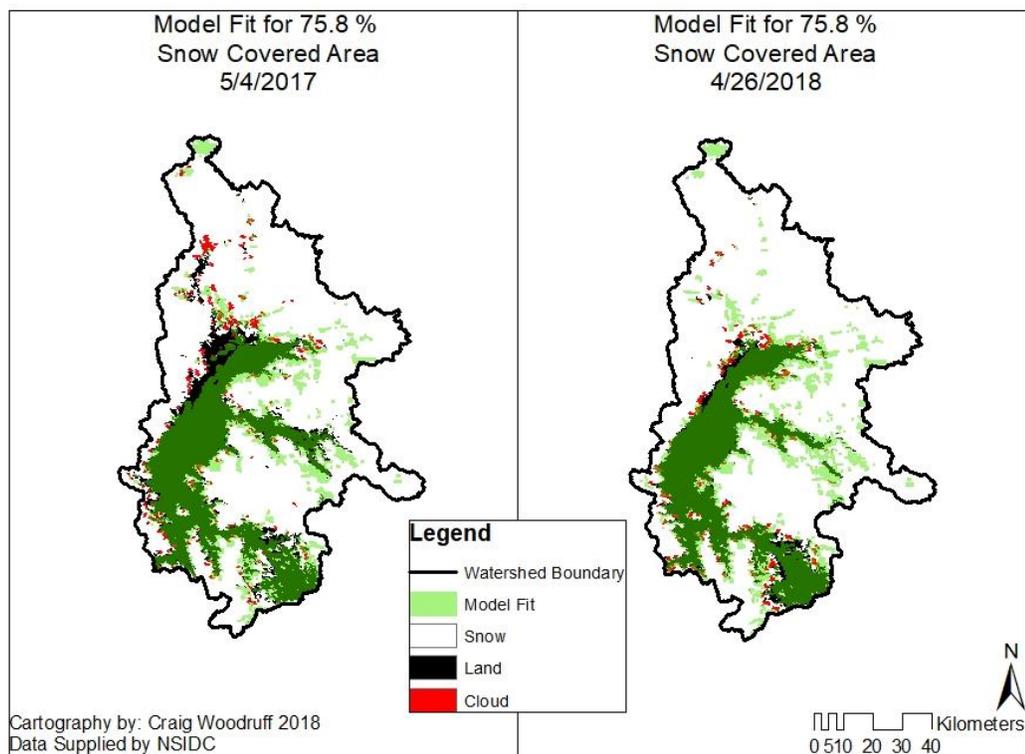


Figure 4.3: Side by side comparisons of model fit for 2017 (left) and 2018 (right) for 75.8% snow covered area. Correctly modelled areas are shown in dark green. Black indicates areas where the model indicated snow, but was measured as land. Light green shows the areas the model predicted land but snow was measured. Cloud cover is shown in red.

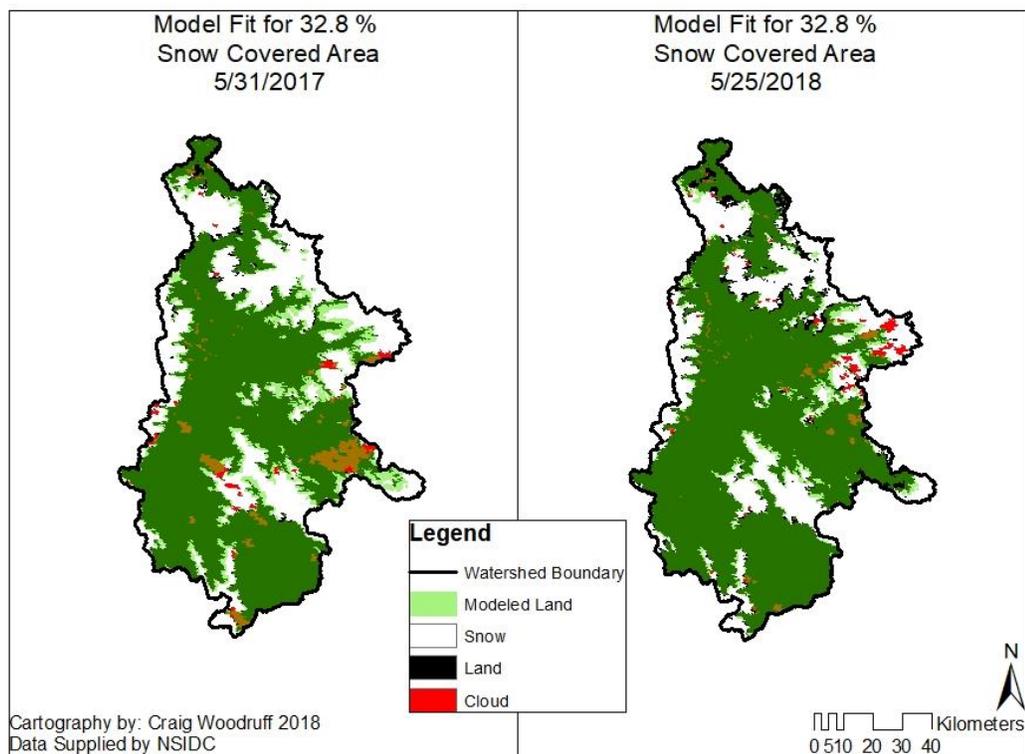


Figure 4.4: Side by side comparisons of model fit for 2017 (left) and 2018 (right) for 32.8% snow covered area. Correctly modelled areas are shown in dark green. Black indicates areas where the model indicated snow, but was measured as land. Light green shows the areas the model predicted land but snow was measured. Cloud cover is shown in red.

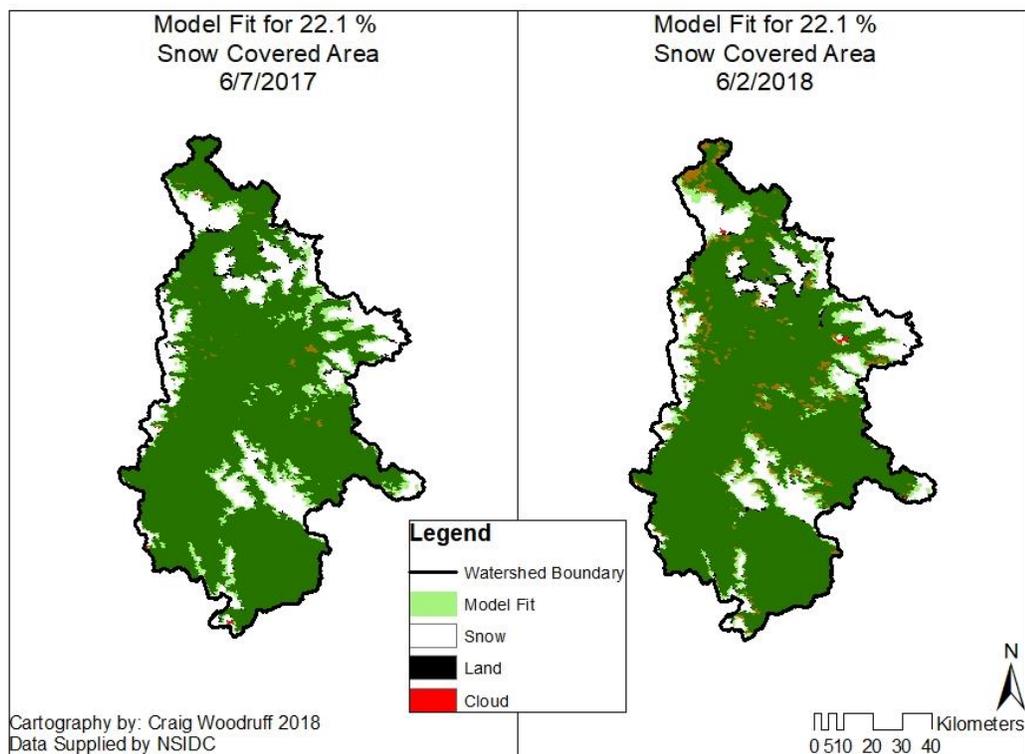


Figure 4.5: Side by side comparisons of model fit for 2017 (left) and 2018 (right) for 22.1% snow covered area. Correctly modelled areas are shown in dark green. Black indicates areas where the model indicated snow, but was measured as land. Light green shows the areas the model predicted land but snow was measured. Cloud cover is shown in red.

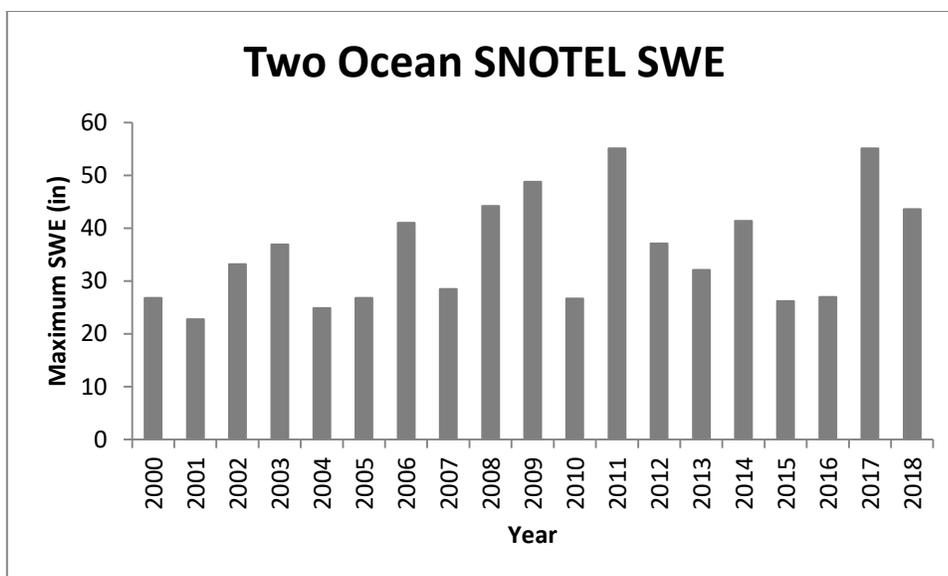


Figure 4.6: Variations in maximum SWE measured at Two Oceans Plateau from 2000-2018. SWE is shown in inches.

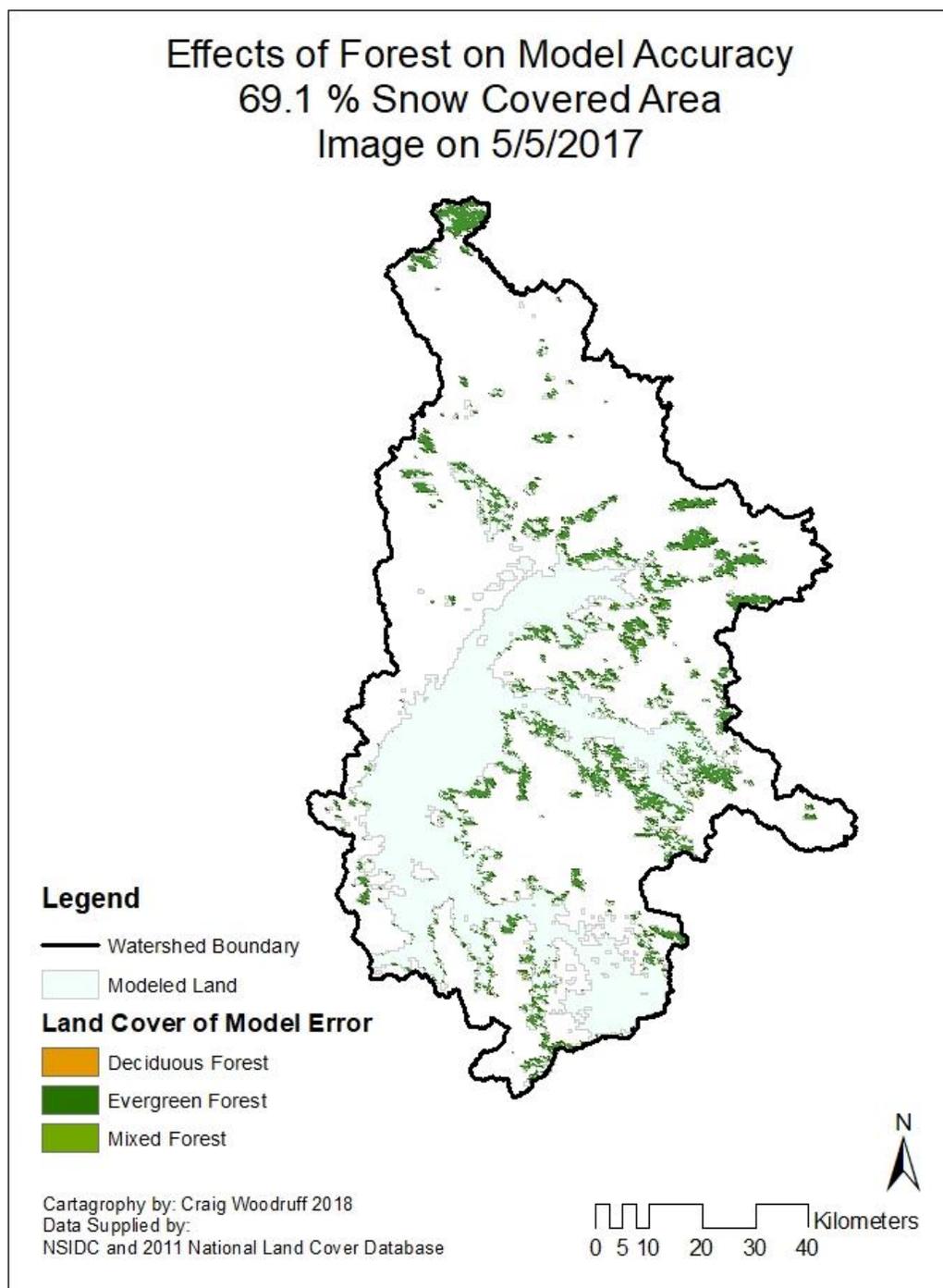


Figure 4.7: Effects of land cover on model accuracy due to NDSI threshold.

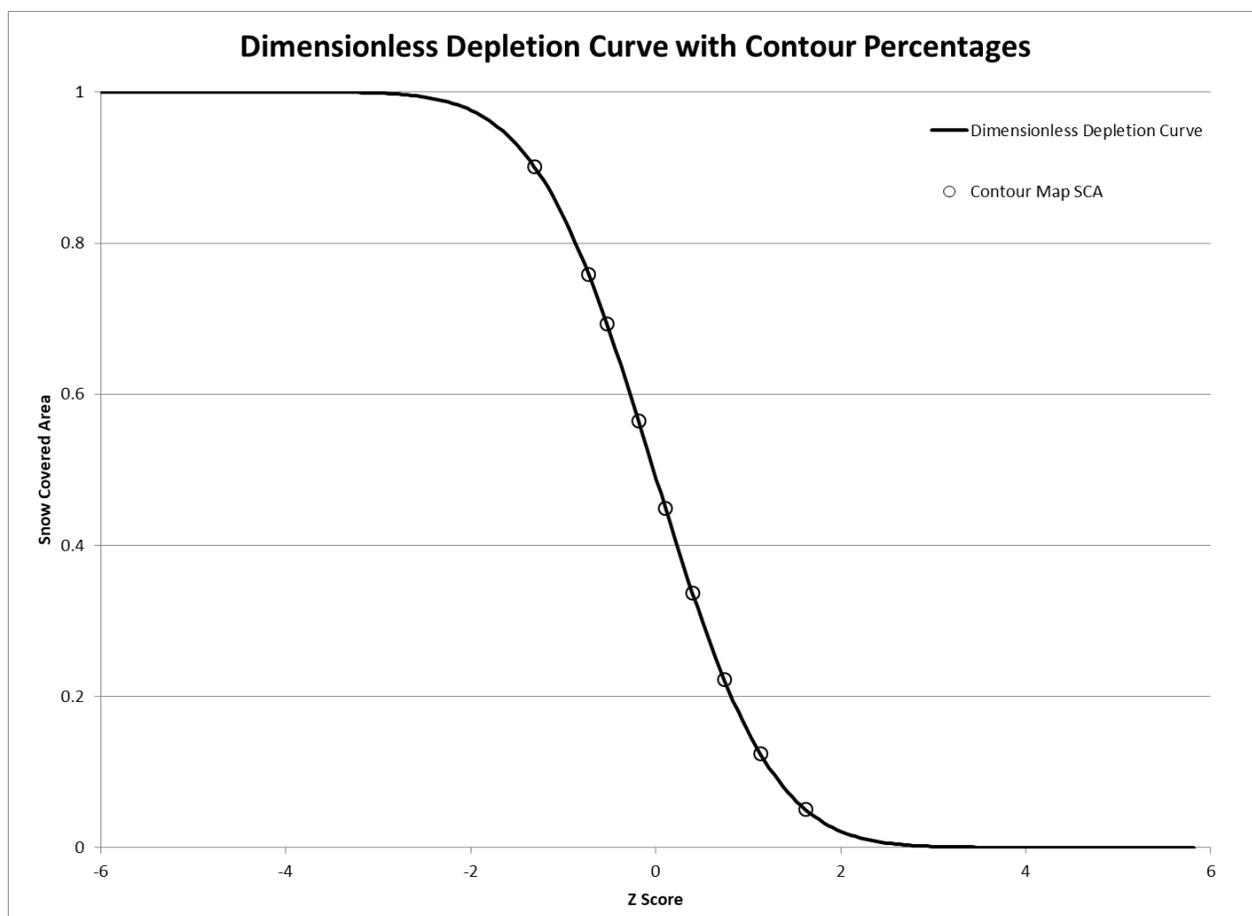


Figure 4.8: Dimensionless depletion curve for the watershed. Snow cover fraction on is shown on the y-axis and z score on the x-axis. The snow covered areas shown on the contour map are represented by circles along the curve.

Chapter 5: Concluding Remarks

The purpose of this thesis research was to develop a spatial model to accurately represent snow covered area changes during a melt season in multiple years. This was accomplished through three steps. We first proved the existence of a pattern, which is essential to this research. A method for accurately modeling snow covered area through a melting season with time was developed. Lastly, we developed two models which uses the repeatable pattern of snow covered area to represent snow covered area in any melting season. The truly fascinating finding in this research was the likeness of the two spatial models developed. In Chapter 3 a model was developed using a principal component analysis based on Eigen vectors and Eigen values which a mathematically derived, which was manipulated in ArcGIS to represent changes in snow covered area of 5%. The second method in Chapter 4 used a method based on an optimization technique which moves closer and closer to the true value like a Newtonian approximation. This model represented about 10% spatial changes in snow covered area. These methods differ fundamentally in their approach to finding the best way to describe the data. The fact that both methods derive a map which fits the actual data, with similar accuracy, and both appear visually similar is astounding. This once again proves the applicability of a repeatable intra-annual pattern in snow cover depletion. This long accepted idea has been excluded from remote sensing, but shows an exciting new area of research.