

THESIS

SNOW DEPTH MEASUREMENT VIA AUTOMATED IMAGE RECOGNITION

Submitted by

Kevin S. J. Brown

Department of Ecosystem Science and Sustainability

In partial fulfillment of requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2019

Masters' Committee:

Advisor: Steven Fassnacht

Jay Ham
Dan McGrath
Matt Ross

Copyright by Kevin Sidney Jordan Brown 2019

All Rights Reserved

ABSTRACT

SNOW DEPTH MEASUREMENT VIA AUTOMATED IMAGE RECOGNITION

Seasonal snow is a significant contributor to the water supply of nearly 2 billion people in semi-arid regions around the world. Quantification of this resource is critical to planning sustainable water and food supplies in these regions. While Snow Water Equivalent (SWE) is the most common parameter used to estimate snow water storage, snow depth has often been used as a proxy since it is much simpler to measure and can be converted to SWE if density can be estimated. Depth of snow varies greatly at the regional, watershed, and plot scales and better quantification of this variability can improve water storage estimates.

Installation and maintenance of new snow measurement sites is typically expensive and time consuming, so a technology that could produce high temporal resolution snow depth data for a low cost would be useful. Manual reading of snow depth from graduated staffs driven into the ground has been used by the Natural Resources Conservation Service (NRCS) for operational and research purposes. The amount of data available from this method has traditionally been limited by the time-consuming step of manually reading snow depths in images. The central objective of this research was to automate this process in order to reduce the time requirement and allow this technology to be deployed more widely.

Five sites were established with time lapse cameras and a set of snow depth staffs around the state of Colorado. Several image recognition methods were considered, and the Aggregate Channel Features technique was used to detect snow depths based on images of the depth staffs. At the most successful sites, absolute error was close to 20 cm, while at less successful sites consistent errors as high as 100 cm made the data unusable. The variety of site configurations examined allowed factors

which increased error such as forested backgrounds, close staff placement, and poor camera mounting, to be identified. Additional studies could take advantage of new, cloud-based image recognition technologies in order to allow anyone with a camera and an internet connection to measure snow depth automatically from pictures taken at specific locations.

ACKNOWLEDGEMENTS

I would like to thank my advisor Steven Fassnacht, the Colorado Water Conservation Board, my committee members, Jay Ham, Dan McGrath, and Matt Ross, as well as the Hill Memorial Fellowship, the College of Natural Resources, Jeff Derry and the Center for Snow and Avalanche Studies.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES	vi
LIST OF FIGURES.....	vii
CHAPTER 1 – INTRODUCTION	1
1.1 Background.....	1
1.2 Project Goals.....	5
CHAPTER 2 – METHODS.....	6
2.1 Site Selection.....	6
2.2 Equipment.....	7
2.3 Site Configuration	8
2.4 Site Maintenance.....	10
2.4 Image Recognition	10
2.5 Data Analyses.....	16
2.6 Snow Depth Data Filtering.....	21
2.7 Joe Wright Snow Depth Data Further Evaluation.....	22
CHAPTER 3 – RESULTS.....	23
3.1 Snow Depth Data.....	23
3.2 Site Accuracies	30
3.3 Optimal Calibration and Detection Process	31
3.4 Historical Joe Wright Snow Depth Distribution.....	37
Chapter 4 – DISCUSSION.....	41
4.1 Snow Depth Measurement Technologies Comparison.....	41
4.2 Snow Depth Variability	45
4.3 Snow Measurement via Terrestrial Photography	46
4.4 Recommendations.....	48
Chapter 5 – CONCLUSION.....	54
REFERENCES.....	56
APPENDICES	62
Appendix A: Miscellaneous Figures.....	62
Appendix B: Site Configurations.....	68
Appendix C: Full Sensitivity Analysis Results	73
Appendix D: SNOTEL Depth Data WY 2009 – WY 2018	75

LIST OF TABLES

Table 2-1:.....	19
Table 3-1.....	27
Table 3-2.....	30
Table 3-3.....	33
Table 3-4.....	34
Table 4-1.....	43

LIST OF FIGURES

Figure 2-1	7
Figure 2-2	8
Figure 2-3	9
Figure 2-4	11
Figure 2-5	11
Figure 2-6	12
Figure 2-7	14
Figure 2-8	15
Figure 2-9	17
Figure 2-10	20
Figure 2-11	21
Figure 3-1	24
Figure 3-2	24
Figure 3-3	26
Figure 3-4	28
Figure 3-5	29
Figure 3-6	29
Figure 3-7	31
Figure 3-8	32
Figure 3-9	34
Figure 3-10	36
Figure 3-11	36
Figure 3-12	38
Figure 3-13	39
Figure 4-1	46
Figure 4-2	49
Figure 4-3	50
Figure 4-4	51
Figure A-1	62
Figure A-2	62
Figure A-3	63
Figure A-4	64
Figure A-5	65
Figure B-1	68
Figure B-2	69
Figure B-3	70
Figure B-4	71
Figure B-5	72
Figure C-1	73
Figure C-2	73
Figure C-3	74
Figure D-1	75
Figure D-2	76
Figure D-3	76

CHAPTER 1 – INTRODUCTION

1.1 Background

Seasonal snow is a critical component of water supply in the American West and in arid regions around the world. Quantification of this seasonal snow allows for forecasting of spring and summer water availability in these areas. In the United States, snow measurement for the purpose of water forecasting has been taking place for over 100 years (NRCS, 2019) and for most of this time, snow water equivalent (SWE) and snow depth were measured manually each month at snow course sites (Church, 1914; Marr, 1940). This was challenging and labor-intensive as the sites are often difficult to access in the winter and may be located in hazardous terrain. This combined with a desire for greater temporal resolution led to the introduction of the automated snow telemetry (SNOTEL) stations that now comprise the bulk of snow data collection. (NRCS, 2019)

SNOTEL stations are expensive to build and maintain with two sensors using glycol-based pressure transducers, which has limited their abundance and distribution. Unlike the previous snow courses, the SNOTEL sites measure snowpack properties at a single point, and the representativeness of this point data across variable terrain can be poor (Rice & Bales, 2010; Meromy et al., 2013; Fassnacht & Hultstrand, 2015). Gathering snow data at additional points, especially in areas of varying slope, aspect, and cover type, increases available data at the watershed scale and improves basin storage estimates (Balk & Elder, 2000).

1.1.1 Snow Depth as an Indicator of Snow Water Equivalent

When measuring snowpack, SWE is usually considered to be the most important variable, as it is the amount of water stored on the landscape. Terrain features, such as slope, aspect, and

elevation, and vegetative cover are strongly indicative of the snow distribution across the landscape (Meiman, 1968 and references therein). Intensive measurements of snow depth, snow water equivalent, and snow density across diverse terrain can yield high confidence spatial estimates of snow water equivalent based on these terrain variables (Balk et al., 1998). Though snow depth is more variable across the landscape than other properties (Logan, 1973; López-Moreno et al., 2013), it is far simpler to measure than SWE or Snow Density (Peck, 1972). For this reason, many campaigns that gather measurements over a large area have focused on measuring snow depth (Clow et al., 2012).

Snow depth can be converted to snow water equivalent if density is known. Density measurements in the field require retrieving and weighing snow cores, which is time intensive especially over large areas. For that reason, a robust snow density model is required for estimating snow water equivalent from snow depth. Given sufficient data for particular sites, it is possible to construct a model for those sites which produces SWE estimates within the range of SWE variability at the plot scale of each site (Mizukami & Perica, 2008; Jonas et al., 2009). Sexstone and Fassnacht (2014) built a reliable snow density model for the upper Cache la Poudre watershed with two years of measurements distributed across varied terrain data combined with long-term measurements at a single point. This model provided a spatial estimate of snow density based on terrain variables and could be transferred to any watershed with a similar data availability.

1.1.2 Remote Sensing in the Context of Snow Science

Snow covered area and snow covered extent are fairly straightforward to measure from orbit but SWE, the main input for operational water forecasting in the western US, is much harder to retrieve, especially in mountainous terrain (Dozier et al., 2016). SWE retrieval from passive microwave techniques has advanced in recent years but is still limited by inaccurate readings in deep

snowpacks (>1m) and snowpacks with dense, frozen layers (Clifford, 2010). In addition, large pixel sizes (typically 25 km) limit use in complex mountainous environments where SWE can change significantly on the 10m scale (López-Moreno et al., 2013).

The National Aeronautic and Space Administration (NASA) has developed an airborne remote sensing technology that uses LIDAR to measure snow depth and passive microwave to measure snow water equivalent at a 50 m spatial resolution and a better than 10 cm depth resolution (ASO, 2019). Even at this high precision, ground-based measurements are required to verify the data (Bair et al., 2016), and as the technology is airborne it must be flown to collect data.

A low-cost technology that could provide a reliable verification to data from remote sensing would be valuable especially in remote areas. Most water forecasting in the complex terrain of the mountain west rely on ground measurements rather than remotely sensed data due to the limitations of the latter. Improving the reliability of remote sensing in this difficult terrain requires more measurements on the ground to verify data products, which are limited in complex alpine terrain (Clifford, 2010).

Some emerging approaches seek to measure snow depth or snow water equivalent on the plot scale more efficiently than traditional manual field collection. Structure from motion is a form of stereoscopy in which multiple images of the same scene taken from different locations are digitally analyzed to produce a three-dimensional model of the scene. Unmanned aerial vehicles (UAV's) fitted with cameras have been flown in complex alpine terrain before and during snow accumulation to retrieve snow depths with errors as low as 10 cm (Nolan et al., 2015; Vander Jagt et al., 2015). Synthetic Aperture Radar (SAR), a similar technique utilizing satellite observations, has been combined with daily satellite snow cover observations to capture 92% of snow depth variability observed on the ground in complex alpine environments (Rondeau-Genesse et al., 2016). A low-impact, high temporal resolution technique uses low-cost GPS receivers placed on the ground

surface to measure snow water equivalent of accumulating overlying snowpack at centimeter-scale accuracy for snowpacks with SWE less than 80 cm (Steiner et al., 2018).

1.1.3 Local-Scale Spatial Variability

Additional snow depth measurements can supplement historical or contemporaneous data to produce more robust estimates of snowpack across varying terrain. But even in small areas of consistent terrain snow can still vary as much as 50% (Fassnacht et al., 2018, Lopez-Moreno et al., 2011). Additional measurements taken close together (~10m) can produce estimates closer to the “true” snowpack properties of the plot and provide a quantification of the variability within that plot (Rice & Bales, 2010).

Plot scale variability changes based on landscape characteristics and time of year (Lopez-Moreno et al., 2011). Generally, variability is lower during accumulation and higher during snowmelt (Fassnacht and Hultstrand, 2015), while its interactions with landscape variables are more complicated. This variability directly affects how many measurements are required in a given area to approximate the “true” value. Fassnacht et al. (2018) showed that most often between 3 and 5 measurements are necessary to get within 5% of the mean snow depth. The spatial arrangement of measurements at a particular scale was shown to have little effect on the required number of measurements (Kashipazha, 2012).

1.1.4 Automated Image Recognition in the Context of Snow Science

In the last 10 years there has been interest in utilizing rapidly advancing computer vision techniques to gather data from terrestrial imagery of the snowpack. Some have focused on albedo estimation (Corripio, 2004) while others have worked toward retrieving snow covered area (Arslan et al., 2017). Utilizing a time lapse camera network in the Black Forest in Germany, automated

techniques were used to extract albedo, and a metric of canopy interception in the surrounding trees (Garvelmann et al., 2012). They also attempted to use an automated method to retrieve snow depth from a depth staff in the image but due to inconsistent results opted to measure it by reading off the snow depths manually. Recently, lapse imagery has been used to quantify snow covered area and local-scale snow cover depletion curves in a 30-by-30-meter area of relatively shallow snowpack in the Spanish Sierra Nevada (Pimentel et al., 2017).

Time lapse imagery is useful for tracking snow accumulation but manually reading depth from images is a time-consuming step and thus limits use at scale. The development of a reliable image recognition algorithm to automate this process would allow the deployment of low-cost and low-impact snow depth measurement stations capable of capturing data at a high temporal resolution. Multiple depth indicators could be placed at each camera for little additional cost to allow a greater spatial resolution at the plot scale.

1.2 Project Goals

Building on this previous work and recognizing the utility of a low-cost, off-the-shelf snow depth measurement technique, the research had the following objectives:

- Qualitatively assess challenges of installation and maintenance of a time lapse camera network over a winter season
- Develop a method to automatically convert images to snow depth
- Evaluate the method's accuracy and factors that affect it
- Estimate the distribution of hourly depth change at several study sites in order to identify spuriously large or small snow depth readings

CHAPTER 2 – METHODS

2.1 Site Selection

Several factors were considered in choosing locations for the time lapse cameras. Safe access during winter was of paramount importance as was compliance with public lands use regulations. These requirements were met by placing the sites alongside certain previously permitted sites maintained by other snow science groups. In addition, distributing the sites across different regions in the state of Colorado hedged against the potential of a poor snow season affecting data collection in a single region. Five camera and snow staff measurement locations were established (Figure 2-1). Each location had characteristics that made equipment setup simpler. Specifically, consistent snow cover from November through May, no canopy cover, and a slope of less than 10% (Table 2-1).

Three sites were located alongside Natural Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) stations, one was located at the Center for Snow and Avalanche Studies' Swamp Angel Study Plot, and one was located alongside the Watershed Function Scientific Focus Area <<http://watershed.lbl.gov/>> near the East River in Gunnison County.

Table 2-1: Site characteristics of snow measurement sites

	Joe Wright	Columbine	Swamp Angel	East River	Red Mountain
Elevation (m)	3085	2792	3371	2775	3414
Slope (%)	5%	13%	8%	19%	7%
Aspect (8 Point Cardinal Direction)	NW	E	E	NE	W
2018 Snow Cover (Days)	243	191	204	122	208

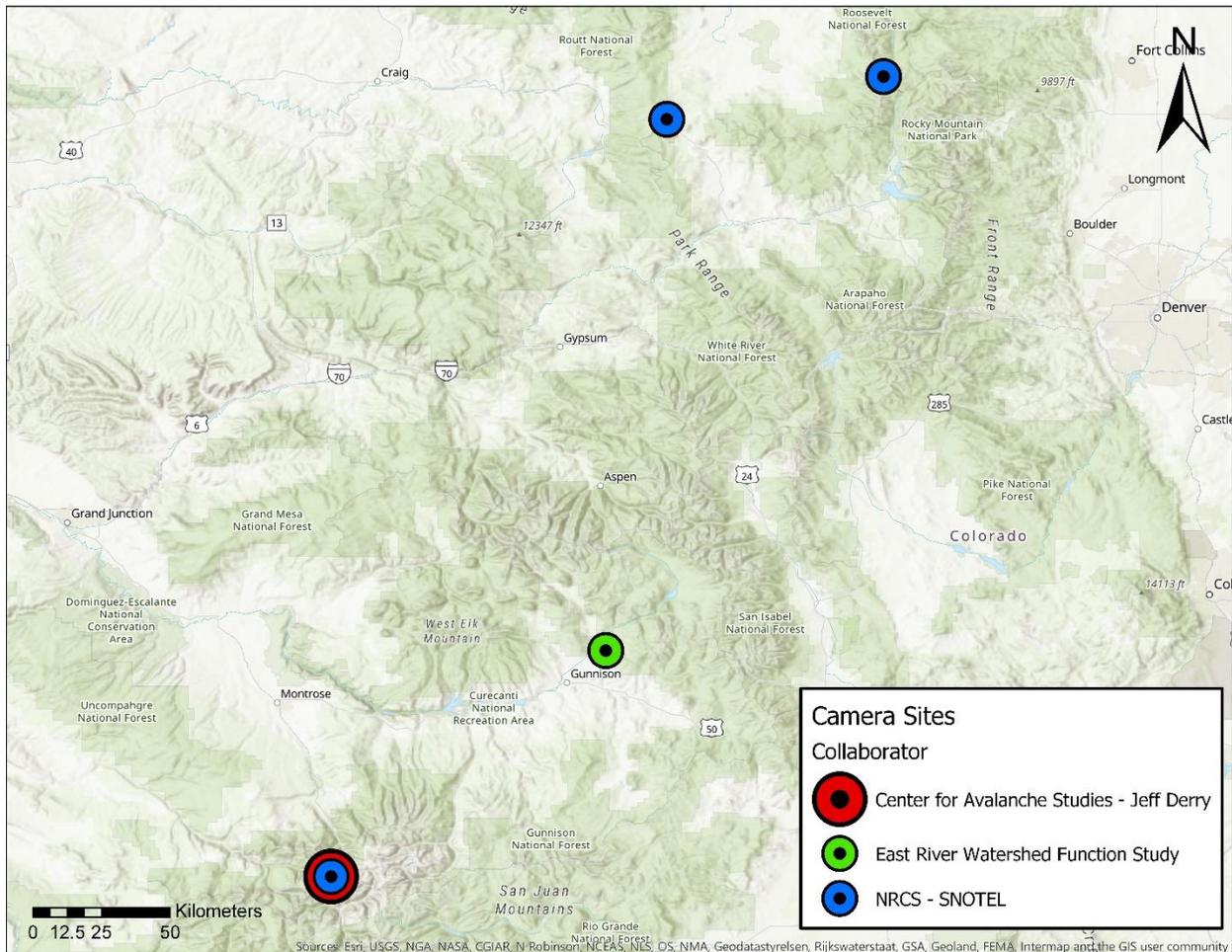


Figure 2-1: Site locations in Colorado color-coded by collaborator group

2.2 Equipment

Each site consisted of two components: a time lapse camera and a set of snow depth staffs. The Stealthcam G34 12 MP camera <stealthcam.com> was selected as it had weatherproofing, long battery life, and high resolution for a low cost (Figure A-1). Additionally, this camera model maintained an internal clock and wrote metadata into the image files, simplifying subsequent sorting and file management. At SNOTEL sites the camera was mounted on the SNOTEL meteorological tower (Figure A-3) facing toward a non-forested area (Figure 2-2). At sites without a meteorological tower, the cameras were affixed to a 3.3m steel T-post (Figure A-1).

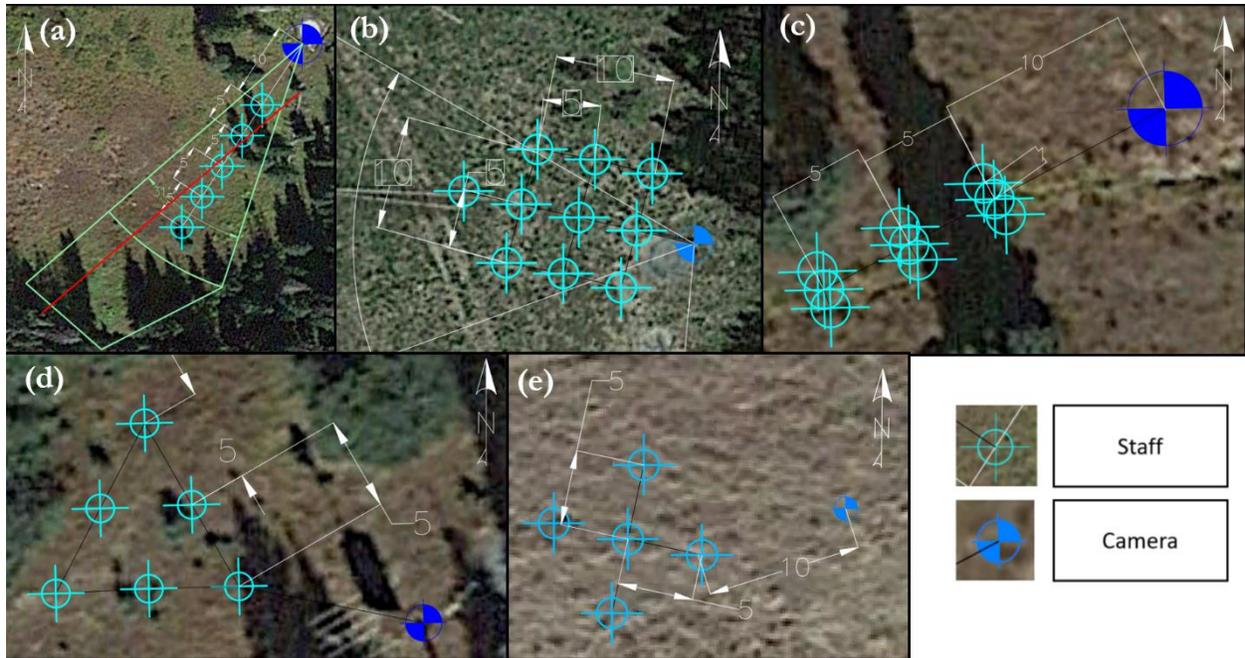


Figure 2-2: Site configuration summary, all distances are in meters. (a) The Joe Wright site was roughly aligned with the discontinued snow course shown in red, (b) Columbine was arranged in simple 5 meter grid (c) Red Mountain Pass was arranged in an “I”, as per Fassnacht et al., (2018), (d) the East River site was arranged in a simple plus to minimize footprint, (e) and the Red Mountain Pass Site was arranged in a triangle in order to best represent snow depths across the available cleared space. Refer to appendix B for more detailed maps.

The snow depth staffs were 3.3-meter steel fenceposts painted red (RGB ~ 164, 27, 53) (Figure A-2) to increase contrast with the background. Each staff was driven approximately 25 cm into the ground. The exact depth into the soil varied based on local conditions and was recorded for each staff.

2.3 Site Configuration

The staffs were installed in different configurations at each site (Figure 2-2). The configurations were based on local site conditions and sightlines as well as on snow depth sampling patterns used in previous work. The Joe Wright (Figure 2-2a) and Swamp Angel (Figure 2-2c) sites were arranged for comparison to previous work. At Joe Wright the staffs were arranged to roughly follow snow transects conducted by both the NRCS and subsequent researchers (Figure 2-2a, red

line). The swamp angel site was arranged to reflect a snow depth sampling strategy used to quantify snow depth variability at the plot scale (Fassnacht et al., 2018). The Columbine, (Figure 2-2b), East River (Figure 2-2d), and Red Mountain Pass (Figure 2-2e) sites were tightly constrained by available open space, with the Columbine site bounded by forest and a snowmobile trail, the East River site constrained by other experiments in the area, and the Red Mountain Pass site constrained by dense forest. The varying configurations allowed for evaluation of the effects of staff spacing, occlusion, and background characteristics on image recognition within the confined spaces.

The theoretical precision of snow depth readings from staffs in the image was considered when selecting site configurations. These values were calculated by placing objects of known size in front of a camera then calculating the object size in pixels. This value is the millimeters of depth per pixel for each snow depth staff and increases with distance from the camera (Figure 2-3). Even at 30m theoretical precision is sub-millimeter and only exceeds the millimeter scale at a distance of 50m.

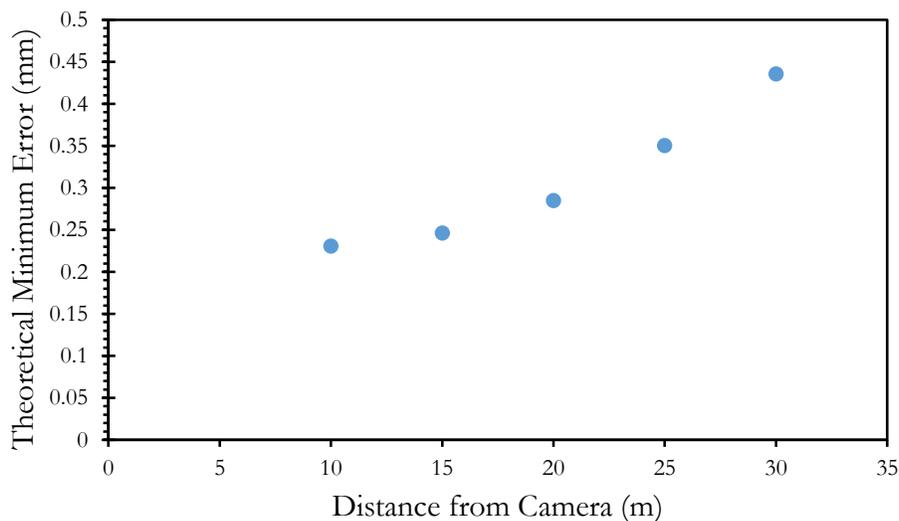


Figure 2-3: Theoretical maximum accuracy for the five staffs at the camera at the Joe Wright site

2.4 Site Maintenance

During the winter, camera batteries lasted about six weeks due to low temperatures, setting a site visit frequency to five to six-weeks. However, all sites and many routes to reach the sites were located in high elevation areas where sudden weather changes could make travel extremely hazardous. Thus, site maintenance was undertaken when possible, with return intervals ranging from four weeks to six weeks. The camera batteries lost power two times before a site visit (once at Red Mountain and once at Swamp Angel) which created gaps in the data at those sites.

2.4 Image Recognition

2.4.1 Image Recognition Overview

All image recognition and computer vision techniques have two stages: a calibration stage and a detection stage (Figure 2-4). In the calibration stage a database of images is prepared with the objects of interest in each image highlighted (Figure 2-5). The areas of interest are generally represented by the coordinates of the edges of a bounding polygon. This database is fed into a calibration protocol which produces a calibrated detector. This detector can then be applied to a set of images to produce a list of areas likely to contain the object which the detector was calibrated for (Figure 2-6).

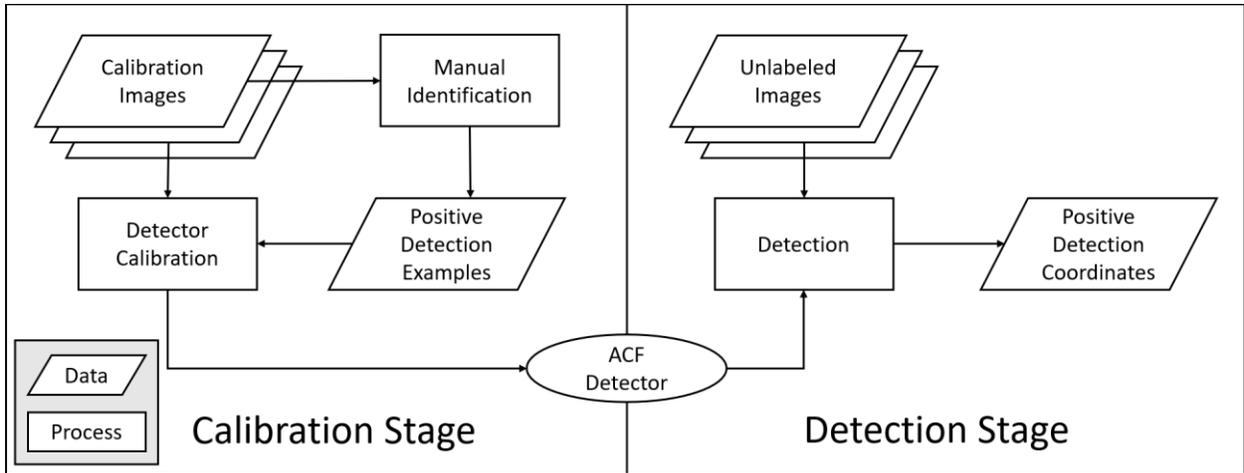


Figure 2-4: General calibration and detection process for image recognition techniques

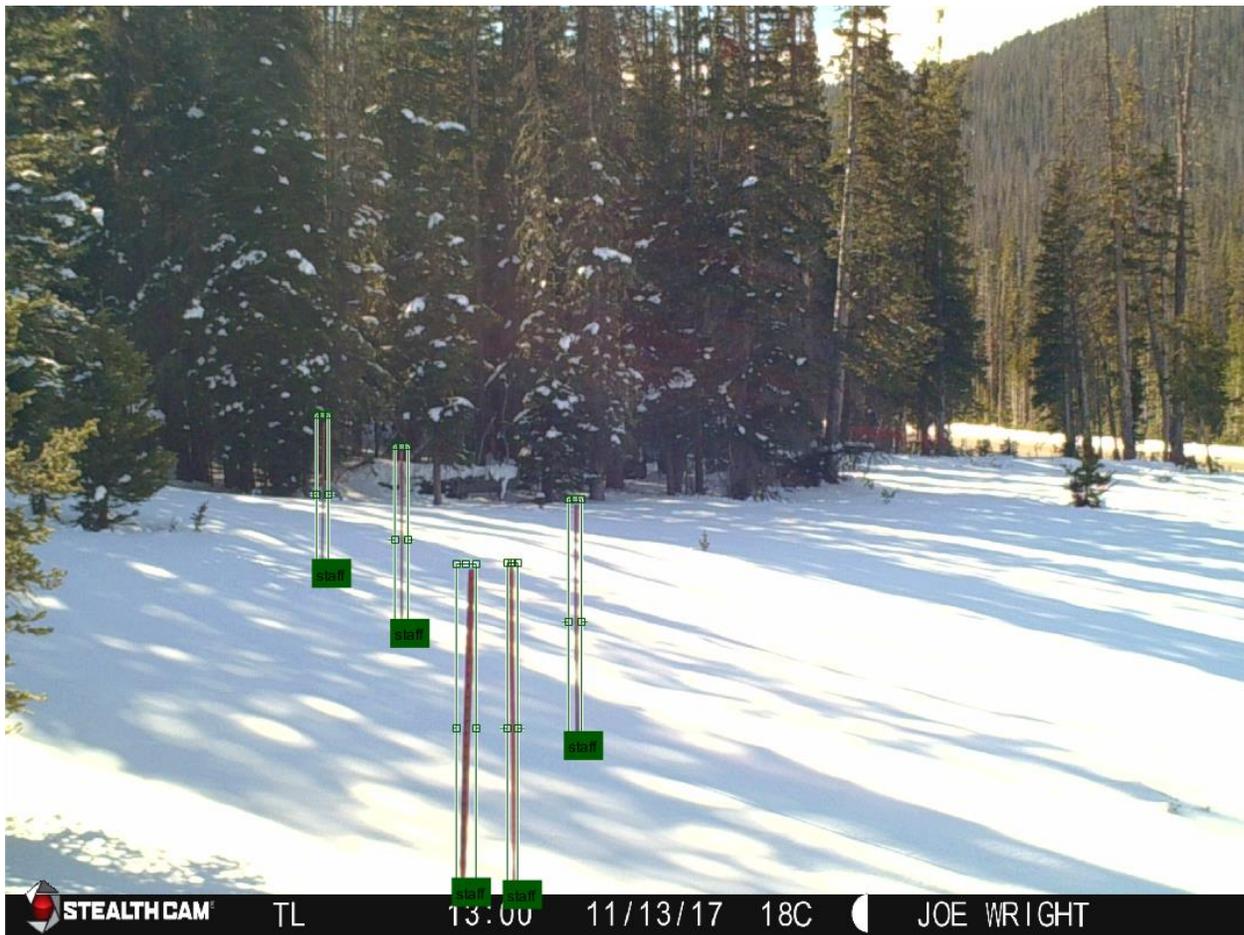


Figure 2-5: Image with user-created bounding boxes around objects of interest. These boxes are used as positive examples in the calibration stage

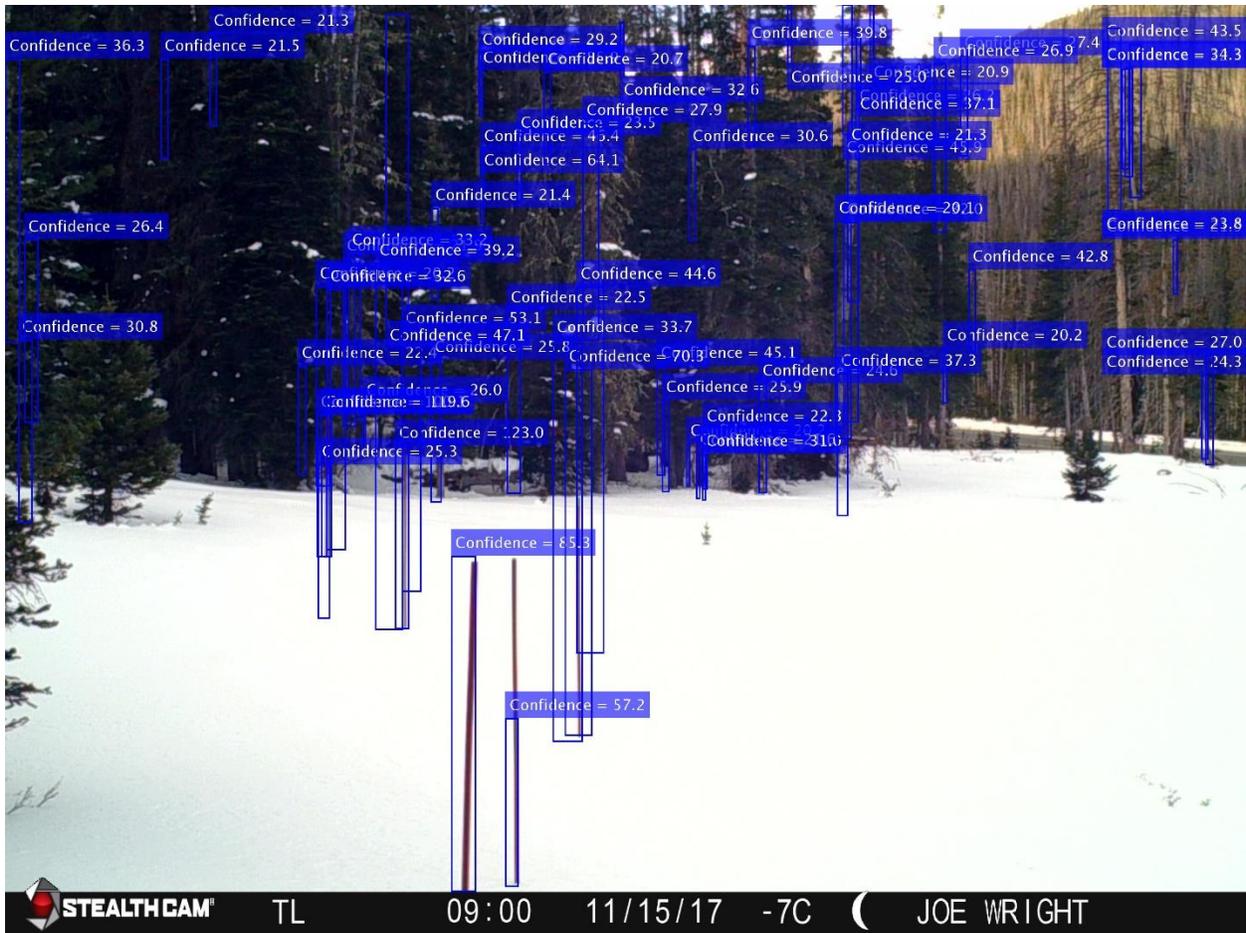


Figure 2-6: Raw output of a calibrated detector applied to an image with no user-supplied bounding boxes

2.4.2 Aggregate Channel Features

The Aggregate Channel Features (ACF) image recognition method (Dollar et al., 2014) was developed to efficiently determine the presence of faces in images (Yang et al., 2014) and is based on the earlier integrated channel features method (Dollar et al., 2009) as well as the Viola-Jones Pedestrian Detection method (Viola & Jones, 2001). The core process of ACF image recognition is a close examination of the image channels (here the red, green, and blue color channels) within a moving window of pixels. An iterative algorithm analyzes this window for features that appear consistently in the positive examples in the calibration dataset but are not present in areas outside the positive examples. These features can include high color contrast, similar pixels appearing in

specific spatial patterns, or dissimilar pixels appearing in a consistent arrangement. Once all the supplied images have been analyzed in this way, information on features that best differentiated positive and negative examples are stored in a detector object. This detector can then be applied to images without user-supplied positive examples in order to identify the target objects.

2.4.3 ACF to Snow Depth

The ACF detector outputs a list of possible positive detections alongside a “score” assigned to each. This score is an estimate of the likelihood of a given detection to be an actual snow staff and are useful for comparing returns within a particular image but cannot be used to compare detection strength between different images (Figure 2-6). There can be over 100 returns from the algorithm for a given image but only a few that correspond to the actual snow depth staffs.

Determining which returns to accept and which staff to assign them to is a critical component of the image recognition process.

The process chosen to accomplish this is depicted as a flowchart in Figure 2-7. First, a baseline bounding box for each staff was delineated manually. If either the camera or staffs moved significantly during the winter, additional baseline bounding boxes could be provided which would then be used from a specified date forward until a new set of bounding boxes was provided.

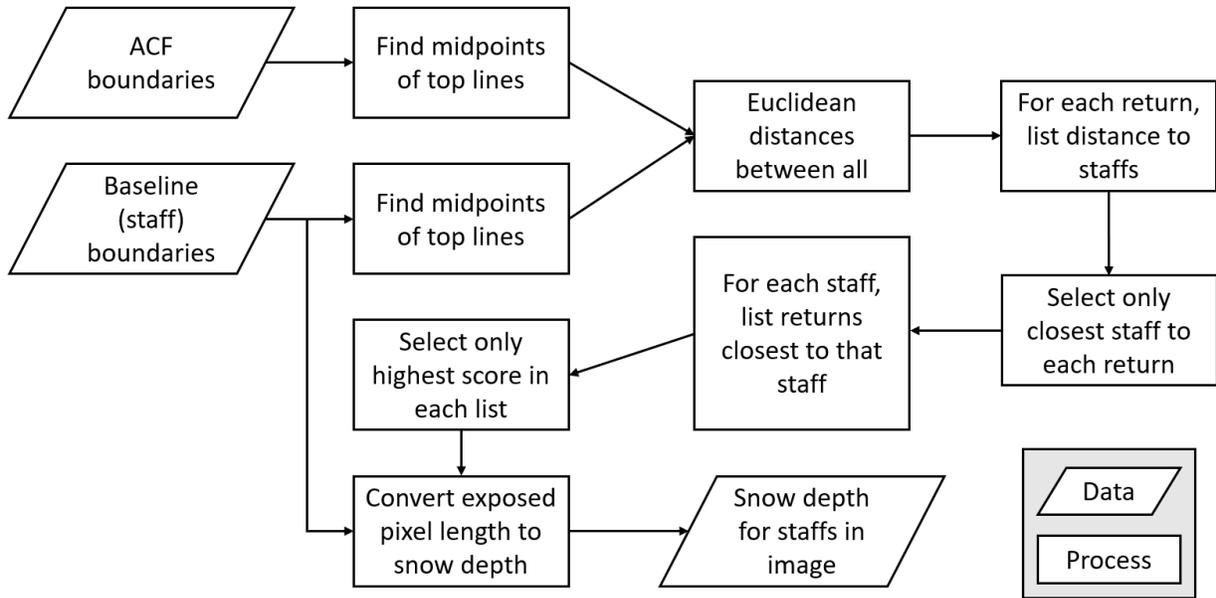


Figure 2-7: Process for assigning bounding boxes to individual staffs

These baseline boxes were used to determine the Euclidean distance (Pythagorean Theorem), in pixels, from the midpoint of the top line of each bounding box (Figure 2-8a) to the midpoint of the top line of each baseline box (Figure 2-8b). The top line midpoint was used instead of the geometric center of the box because the staff's visible length changes over the winter while the top of the staff remains in the same location. For each true staff location, a list of returns that were closest to that staff was generated. The scores associated with each return were used to sort these lists (Figure 2-8c) to produce the detection with the highest score in each of those lists, which was then selected as the final detected bounding box for that staff (Figure 2-8d).

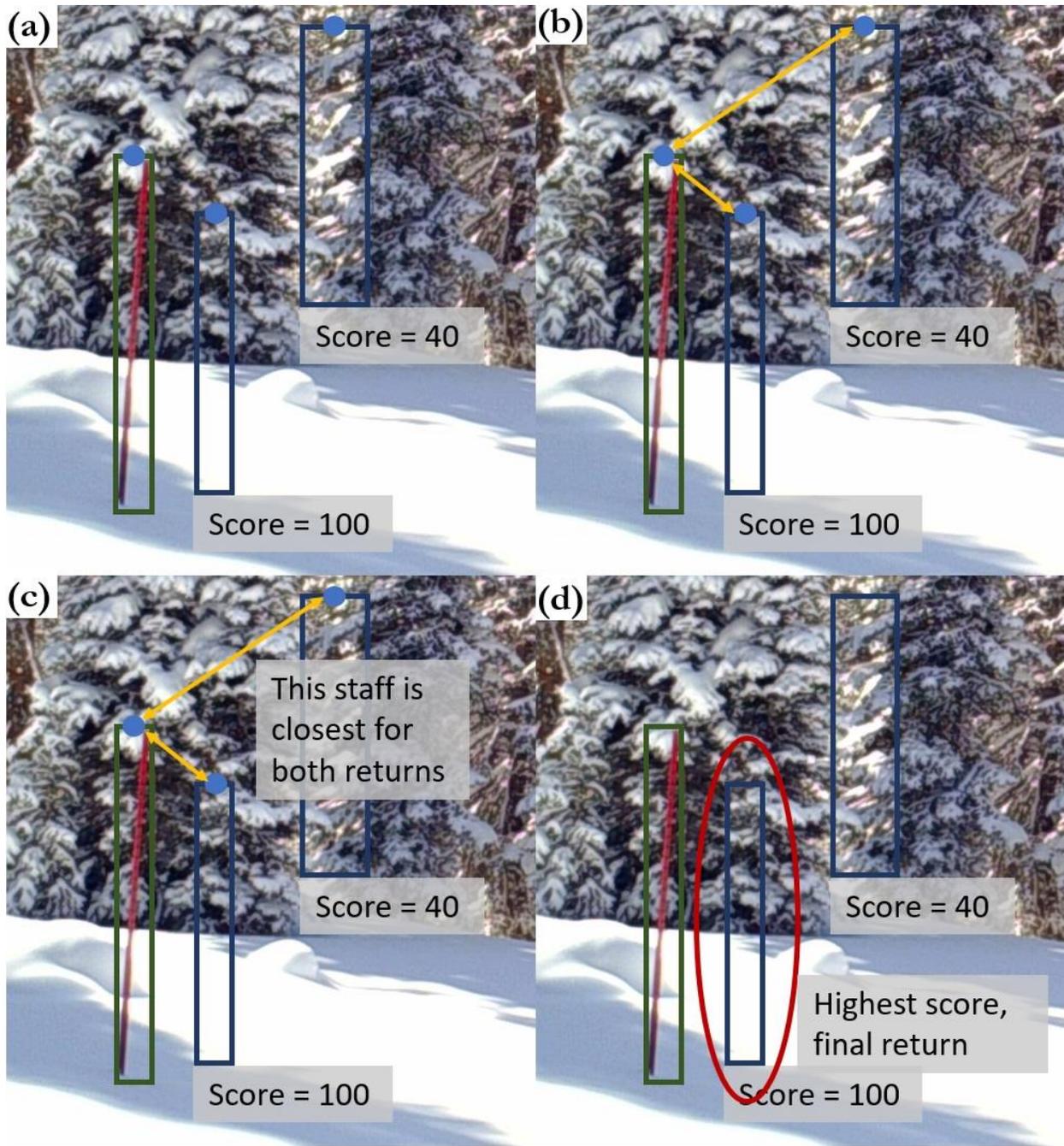


Figure 2-8: Examples of key stages in the bounding box to snow depth assignment process for a single staff and two returns created for demonstration purposes. In an actual image detection, the returns would likely be partially overlapping the staff, but these are distinct to make the demonstration clearer. The baseline staff is in green, two example returns are in blue (a) midpoint of top lines, (b) finding Euclidean distance (c) selection by highest score, (d) final accepted return

Once one bounding box was selected for each staff in the camera's view, it could be converted to a snow depth using the following equation to convert pixel measurements to snow depths:

$$d_s = L_0 - \frac{p_i}{p_0} L_0 \text{ or } d_s = L_0 \left(1 - \frac{p_i}{p_0}\right) \quad (1),$$

where d_s is snow depth in mm, p_0 is pixel length of staff with no snow, p_i is pixel length of exposed staff in image i , and L_0 is length of exposed staff with no snow in mm. The pixel length with no snow would be determined from an image early or late in the season when there was no snow present.

2.5 Data Analyses

2.5.1 Performance Metrics

Two methods were considered to evaluate the accuracy of snow depth estimation. The most straightforward method is simply to compare the algorithm-generated snow depth with snow depths generated manually from those same images. An alternative method frequently used in computer vision research would be to compare the pixel areas of the algorithm-generated bounding boxes and human-determined bounding boxes (Figure 2-9).

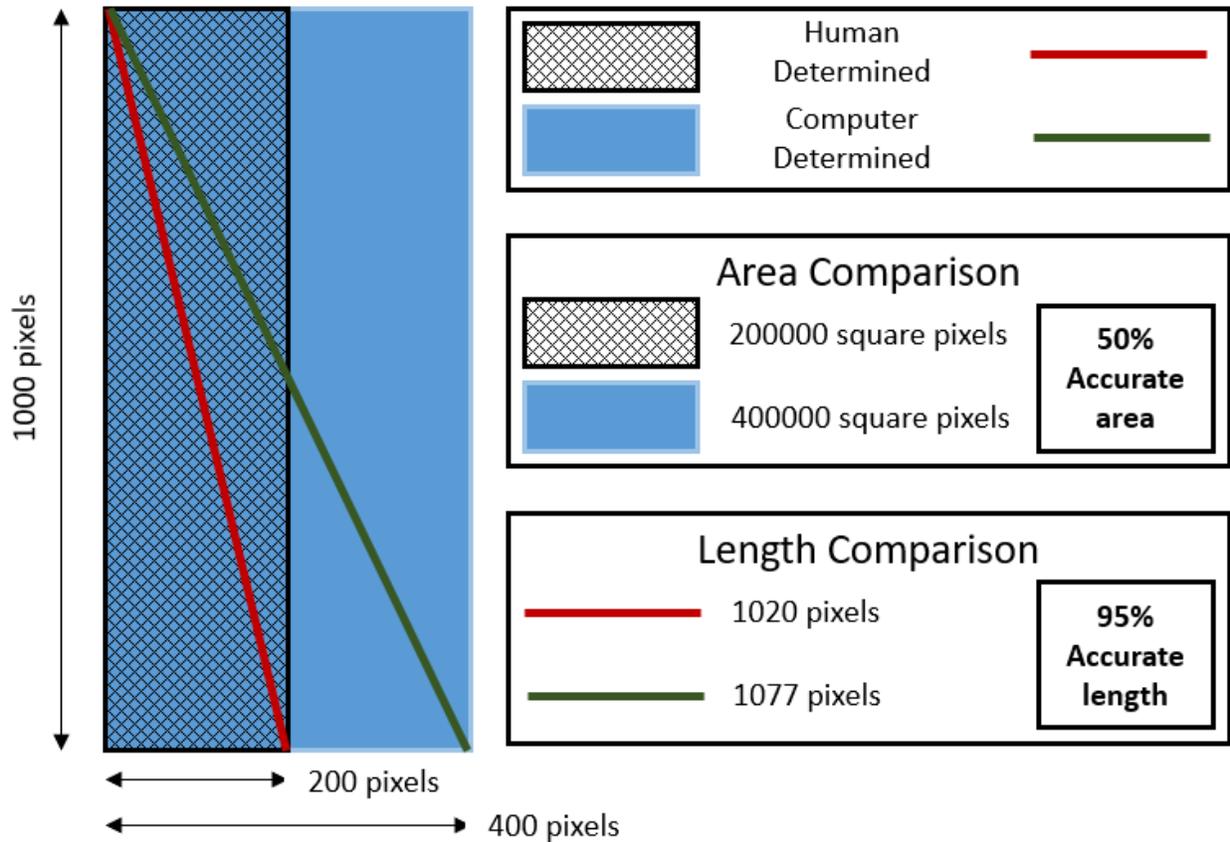


Figure 2-9: A comparison of error measurement based on identified areas to error measurement based on staff length using demonstration data. The shaded area and red line are the human identified staff location while the blue area and green line are the computer identified staff location

The detected areas can be drastically different while the snow depth varies by a much smaller margin (Figure 2-9). The central objective is to measure snow depth, not necessarily to exactly match the identified areas. Therefore, the measured snow depth was chosen to evaluate the accuracy of the image recognition process.

2.5.2 Detector Calibration Process Evaluation

In order to examine whether it was important to calibrate detectors for specific sites with images only from that site, a detector was trained with default parameters for each of four sites with a set of 100 labelled images. Then a detector was trained using a combined dataset with 200 images

taken from all sites. These detectors were then tested against a set of 50 reference images from each site for a total of 200 images.

2.5.3 Input Parameter Evaluation

The ACF Object Detector has a variety of parameters that can be changed in both the calibration and detection stages (Table 2-1). The optimal values of these parameters were found by performing a simple sensitivity analysis. For both calibration step variables and detection step variables, a collection of 200 calibration images and 200 reference images were compiled. Each of these sets was composed of 50 images from each of four sites (not including the East River site due to low snow accumulation). The images were a mix of early, middle, and late season images. The first set was used to train the algorithm and the second was used to evaluate the detection accuracy.

Table 2-1: a) Calibration and b) detection parameters examined for possible improvements in detection accuracy

Parameter	Description	Range
<i>a) calibration parameters</i>		
ObjectTrainingSize	A height and width in pixels specifying the size that objects are resized during calibration	
NumStages	Number of calibration stages in the iterative calibration process	1-8
NegativeSamplesFactor	Specifies how many times more negative examples will be extracted from the image than supplied positive examples	2-8
MaxWeakLearners	Number of Weak Learner objects inside the detector	256-8192

<i>b) detection parameters</i>		
NumScaleLevels	How many levels of downscaling are carried out at the start of detection	1-8
WindowStride	How many pixels the scanning window moves between detections	1-8
SelectStrongest	Use an internal method to deal with overlapping bounding boxes	True/False
MinSize	Height and width in pixels of the smallest area allowed to contain a positive detection	
MaxSize	Height and width in pixels of the largest area allowed to contain a positive detection	
Threshold	Parameter that can increase speed of detection at the cost of missing possible positives	-1 to 1

For calibration parameters, the process followed Figure 2-10. A detector was trained using the 200 images then tested against the 200 validation images. The absolute percent error between the detected depths and the manually delineated depths was recorded for each as well as the average elapsed time for calibrating that detector. For the detection parameters, the evaluation process followed Figure 2-11. A detector trained with the highest performing parameters from the previous step was applied to the store of images. The average absolute percent error in snow depths and the average time per detection were found for each parameter tested.

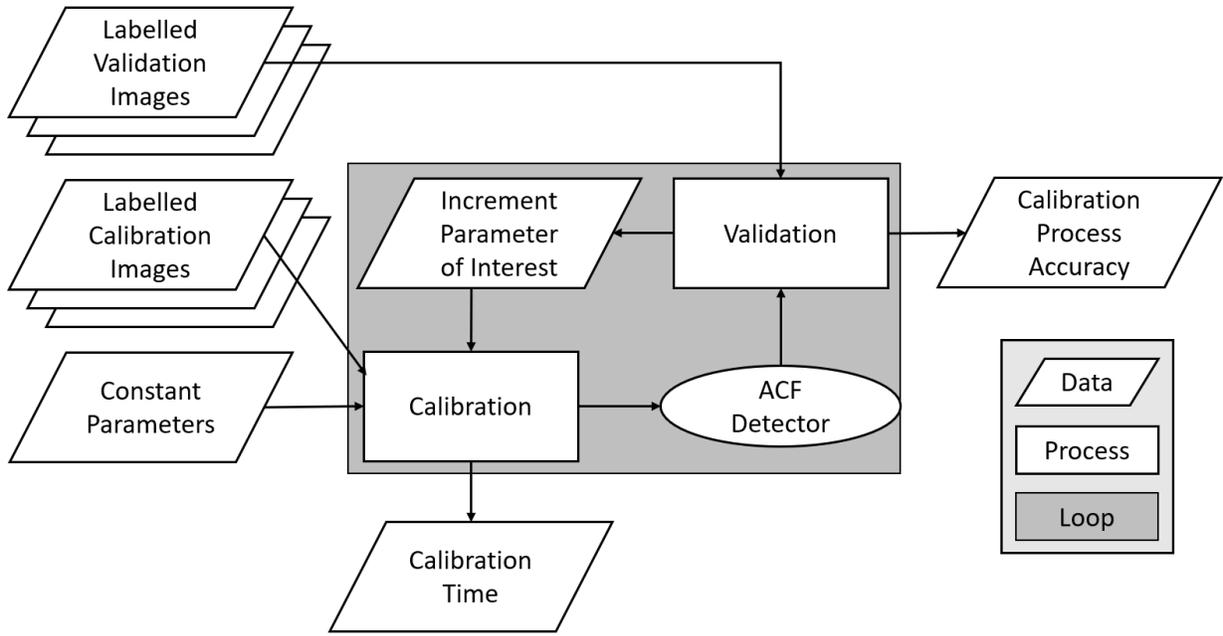


Figure 2-10: Calibration parameters evaluation process in which a detector is repeatedly trained while varying one parameter and applied to the validation dataset

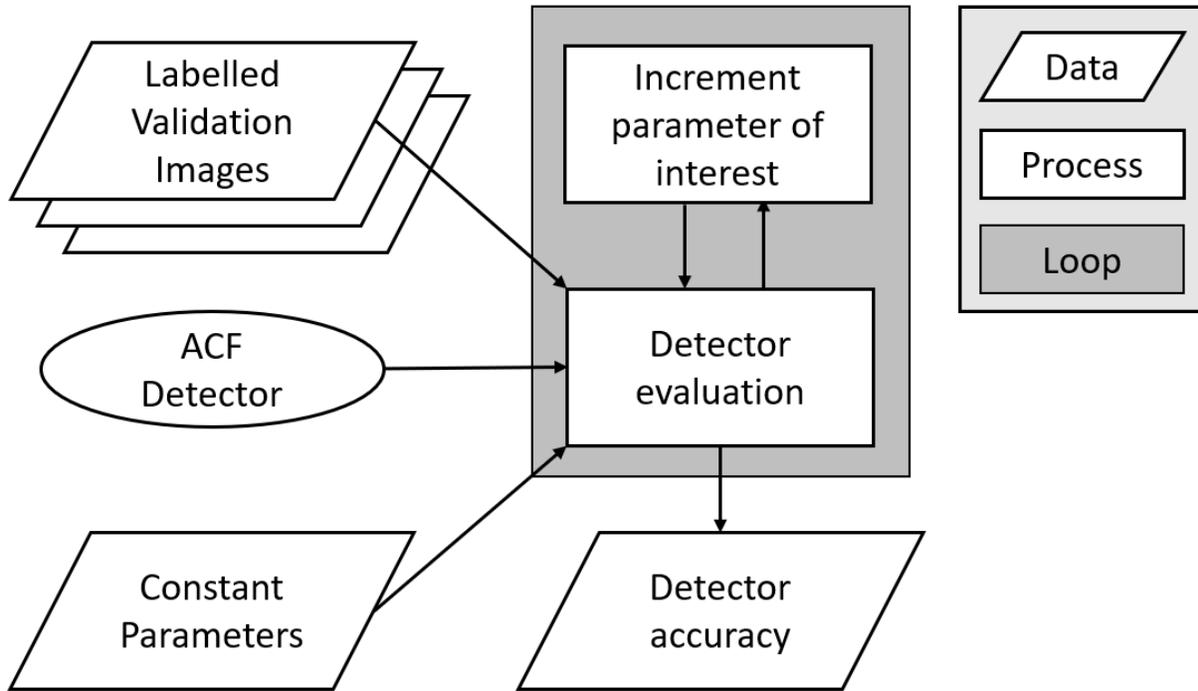


Figure 2-11: Detection parameters evaluation process in which a static detector is applied to the validation images dataset multiple times while varying the detection parameter of interest

This parameter evaluation process was very computationally intensive, especially the training parameter evaluation. It was conducted on high-performance, consumer-grade desktop computers available in the Warner College of Natural Resources Computer Labs. The computers had Intel Core i7 processors with integrated graphics processing as well as 16 GB of Random Access Memory (RAM) and 500GB of on-system Read Only Memory (ROM). The computers are specialized to geostatistical processing and devices with similar capabilities are generally available where geographic information systems and computer aided design are used frequently.

2.6 Snow Depth Data Filtering

The resultant snow depths from this process had a high level of noise so the data were further processed to remove outliers using the SNOTEL-measured hourly depth change during Water Year (WY) 2018. The SNOTEL maximum positive and minimum negative hourly snow

depth changes were assigned as the maximum and minimum allowable hourly changes at each site. Only snow depths which changed by less than this amount were accepted as final measurements. When a time step passed without any snow measurement accepted for a particular staff, the range of accepted change from the previous measurement for that staff would be increased linearly based on the number of previous time steps lacking data. This process eliminated outliers and reduced noise in the final snow depth product.

2.7 Joe Wright Snow Depth Data Further Evaluation

The depth staffs at Joe Wright were placed roughly along the snow course formerly taken at that site which allowed comparison of snow course data with snow depth staff readings (Figure 2-2a). While there were no snow courses taken while the depth staffs were present, independent snow depth transects along the same path were carried out during water years 2006, 2007, and 2008, overlapping with available SNOTEL depth sensor data. Using these two data sources, a robust relation between Joe Wright SNOTEL depth readings and snow depths was established and applied to WY 2018. These expected snow depths were compared to the detected data for that year to determine how consistently snow distribution in that season followed the pattern previously observed.

CHAPTER 3 – RESULTS

3.1 Snow Depth Data

3.1.1 Snow Depth Change Histograms

To further understand the increases and decreases in snow depth over short time periods, the hourly snow depth change was computed for each SNOTEL site based on the sites' hourly depth readings (Figure 3-1). Accumulation at all sites was most frequently 2 inches per hour, with a few increases as high as six inches per hour at all sites. Metamorphism at all sites most commonly occurred at 1 inch per hour with some time steps exhibiting metamorphism rates as high as 4 inches per hour. Rates this high could be due to wind redistribution since any reduction in snow depth before the peak SWE date was categorized as metamorphism. Melt rates were similar to metamorphism rates with most taking place at 1 inch per hour with a few hours having up to 5 inches of melt. The stated precision of the SNOTEL Judd snow depth sensor <juddcom.com> is 1 inch, but the operational accuracy is closer to 2 inches (Ryan et al., 2008).

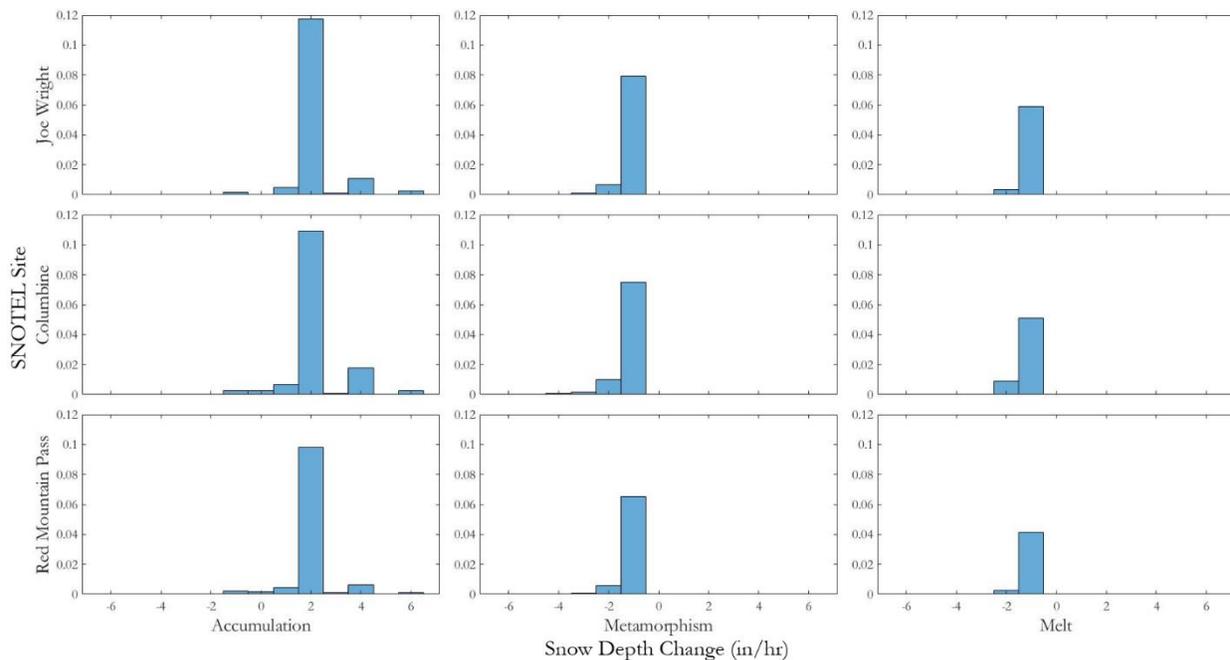


Figure 3-1: Frequency of hourly changes in SNOTEL measured snow depth due to snowfall, melt, and metamorphism for three sites. The Red Mountain Pass measurements were used for error identification at the Swamp Angel site. The histogram bars display the fraction of hours in water year 2018 in which the change was observed. Zero was the most common hourly change and was removed.

The general pattern of snow depth change between the sites is consistent. At all sites the highest accumulation was 6 inches in a single hour. The greatest decreases in snow depth varied between sites, but the general histogram shape was the same in metamorphism and melt. The detail in the histogram is limited by the 1-inch advertised precision of the Judd depth sensor <juddcom.com>, but the data still provide useful upper and lower bounds for maximum snow depth change at each site. These maximum change values were used to remove outlier data points in the automatically detected data where snow depth changed by a greater amount in a single hour, effectively treating the data with a low-pass filter informed by local site conditions. This process removed between 50 and 90 percent of the data points accepted in the detection stage, varying by site and snow depth.

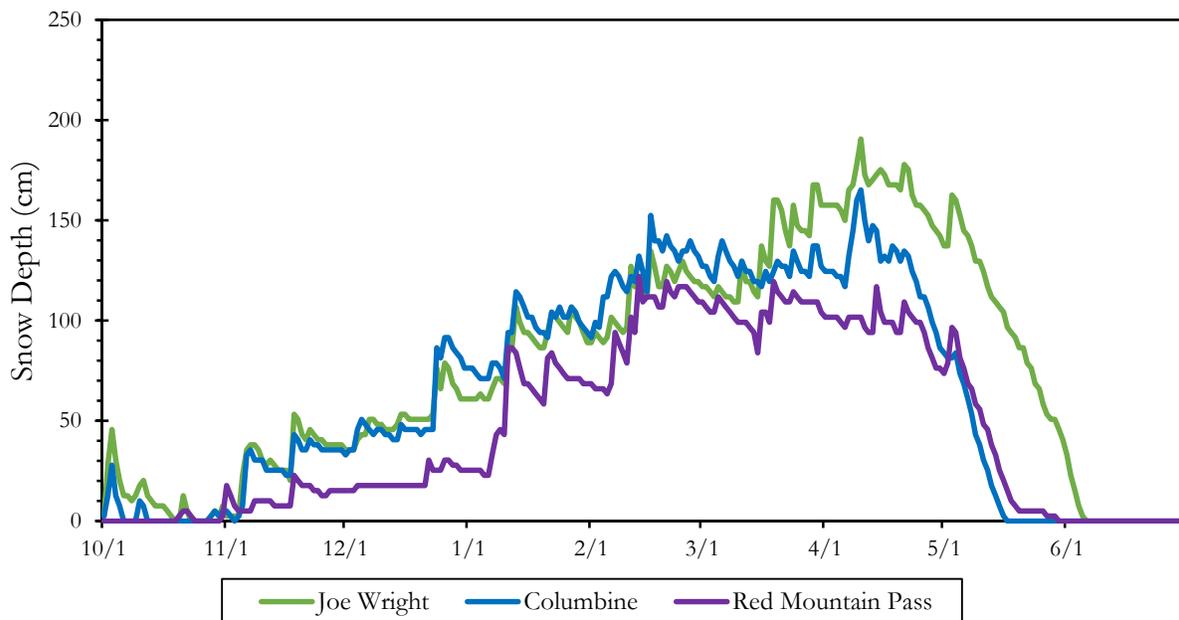


Figure 3-2: Daily SNOTEL snow depths for water year 2018 for Joe Wright, Columbine, and Red Mountain Pass

3.1.2 Aggregate Channel Features Data

The raw snow depth data from the ACF detector is noisy, necessitating filtering to remove outliers. The maximum hourly snow depth change from the ACF detector was constrained to the maximum increases and decreases observed at the SNOTEL station corresponding to each site (Figure 3-1). This method was effective at making the data more consistent (Figure 3-3b) but reduced the number of data points available across the period of record (Figure 3-3a). Applying a moving average with a window size equal to the number of images per day (15 at Joe Wright) to the raw data improved accuracy (Figure 3-3c, Table 3-1) but to a smaller degree than the filter method. Using a moving average on the filtered data (Figure 3-3d) did not produce improvements in most staffs but did drastically reduce the error of one staff (Table 3-1). The most successful method of error filtering was applying a second-order, one-dimensional median filter to the snow depths filtered by maximum possible change. This provided the best results for all staffs with improvements as large as ten percentage points over the maximum change filter alone.

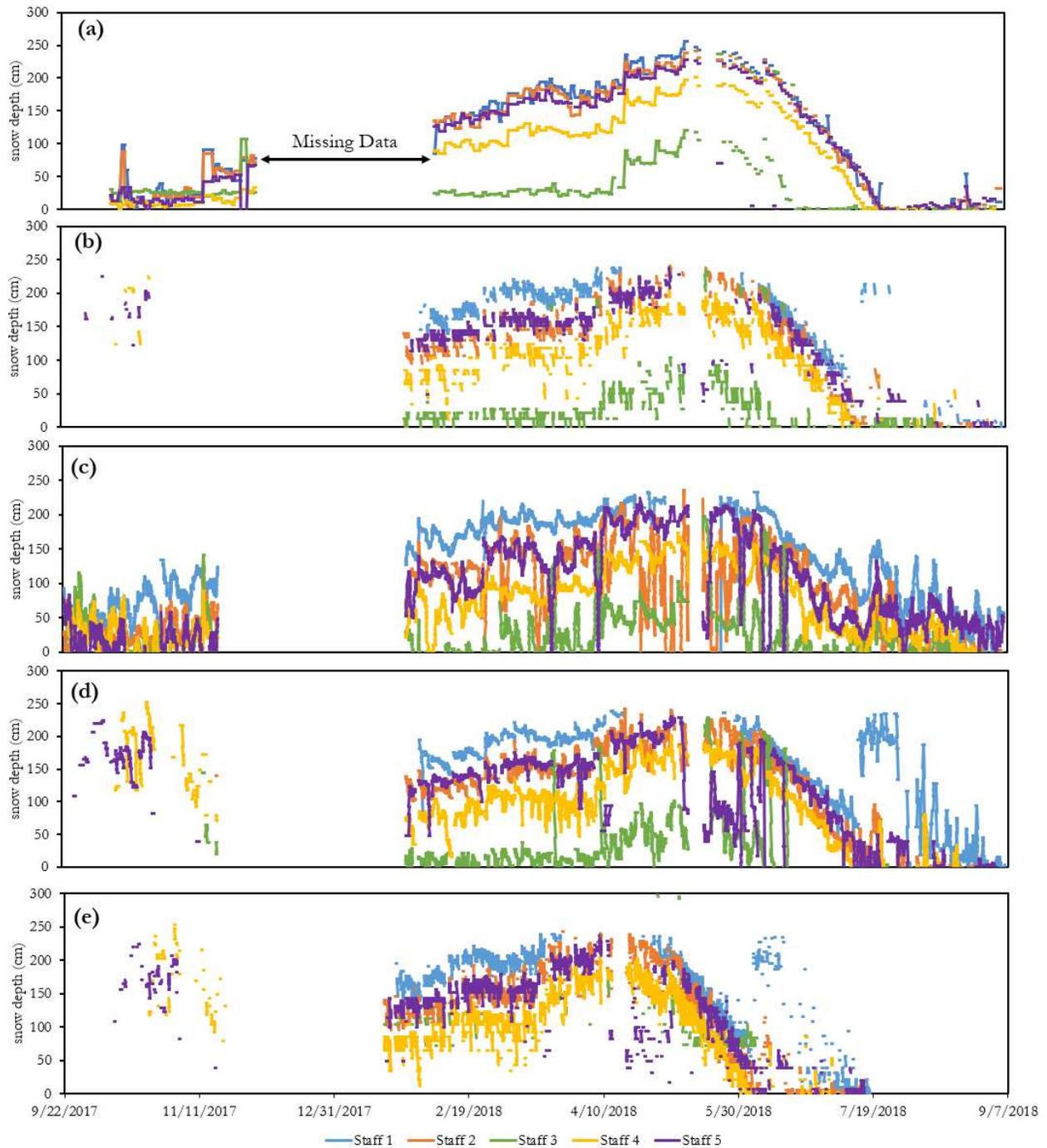


Figure 3-3: Joe Wright snow depths for individual staffs showing (a) manual delineation, (b) automated delineation with maximum allowable hourly change based on the site snow depths, (c) automated delineation using a moving average with a window size of 15, the number of images taken in a day, (d) a moving average of window size 15 applied to the filtered data, and (e) the filtered data after adding a third-order, one-dimensional median filter. The data gap in December was due to an SD card failure and the data gap in mid-April was due to staffs 1 and 2 being completely buried by snow.

Table 3-1: Mean absolute error in percent for Joe Wright by staff and error correction method

	Staff 1	Staff 2	Staff 3	Staff 4	Staff 5
Unfiltered	26%	49%	16%	19%	25%
Filtered	18%	36%	13%	16%	20%
Moving Average	22%	44%	11%	14%	21%
Moving Average of Filtered	21%	22%	11%	17%	29%
2nd Order Median Filter	19%	20%	9%	14%	20%

The Joe Wright results clearly show a difference in snow depths across the sampled area. Staffs 1, 2, and 5 are furthest from the camera and have snow depths that are between 50 and 100 cm deeper than the closer staffs (3 and 4 in Figure 3-3). This pattern is persistent across the season with the furthest staffs reaching zero snow depth up to four days later than the closest staff.

The Columbine site (Figure 3-4) had lower snow depths (Figure 3-2) than Joe Wright (Figure 3-3), and had multiple complete melt offs early in the season before the persistent snowpack started to accumulate. The Columbine camera was also mounted poorly and moved with wind, causing sudden changes in snow depth due to snow measurement misattribution. On February 24th the issue was resolved, leading to more consistent results before March 30th, when the mounting failed again.

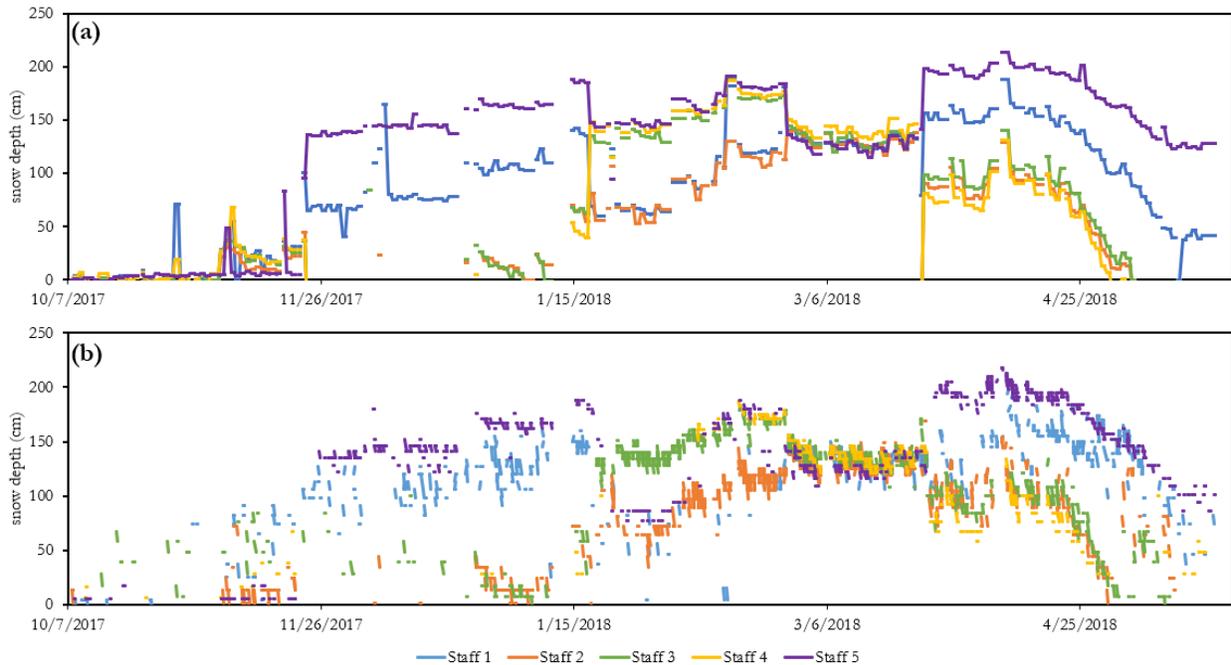


Figure 3-4: Columbine Staff Depths (a) via manual delineation and (b) with automated delineation with error correction using maximum allowable hourly change based on the site snow depth histogram

Red Mountain Pass and Swamp Angel were not conducive to the ACF method (Figures 3-5 & 3-6). Even with filtering, the data are effectively unusable. For Red Mountain, this is due to the forested background of the image yielding poor detections (Figure A-4). Swamp Angel's data are somewhat more useful later in the year when the background becomes consistently snow covered but are still noisy after filtering. The underlying issue at that site was the overlapping of the staffs from the perspective of the camera (Figure A-5). This made it difficult for the algorithm to differentiate staffs from each other and made the detection returns identification process less effective, since it relies on staffs being distant from each other in the image.

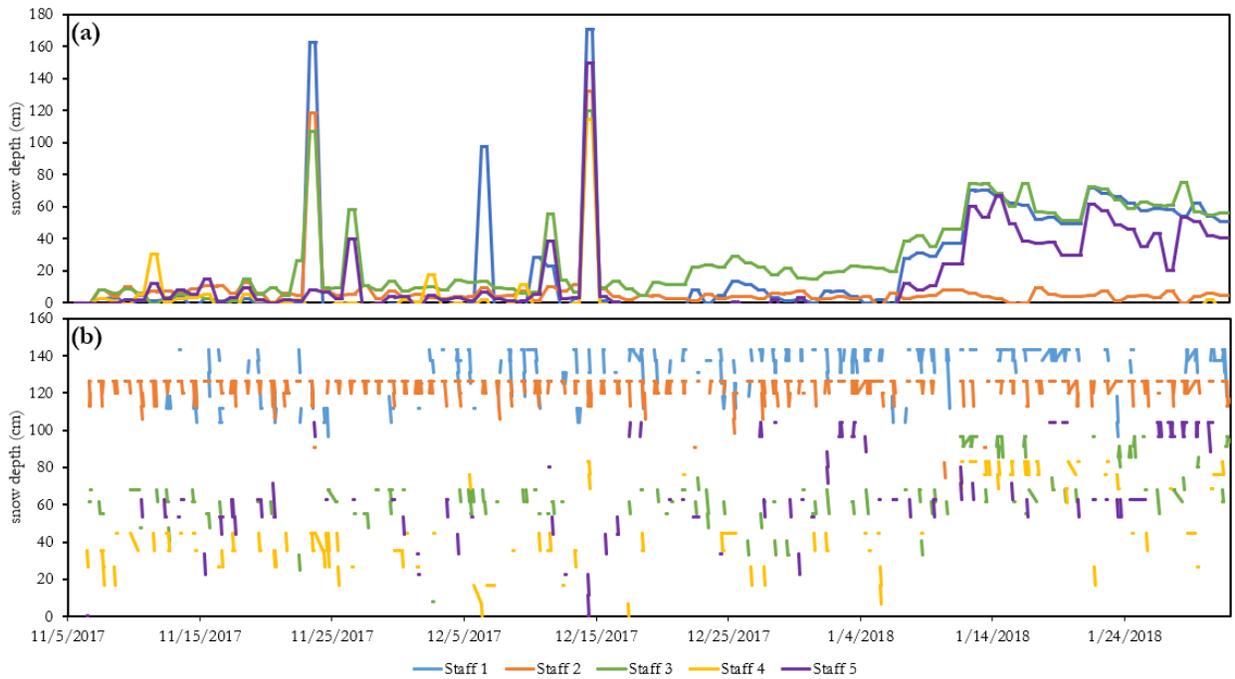


Figure 3-5: Red Mountain Pass staff depths (a) via manual delineation and (b) with automated delineation with error correction using maximum allowable hourly change based on the site snow depth change histogram

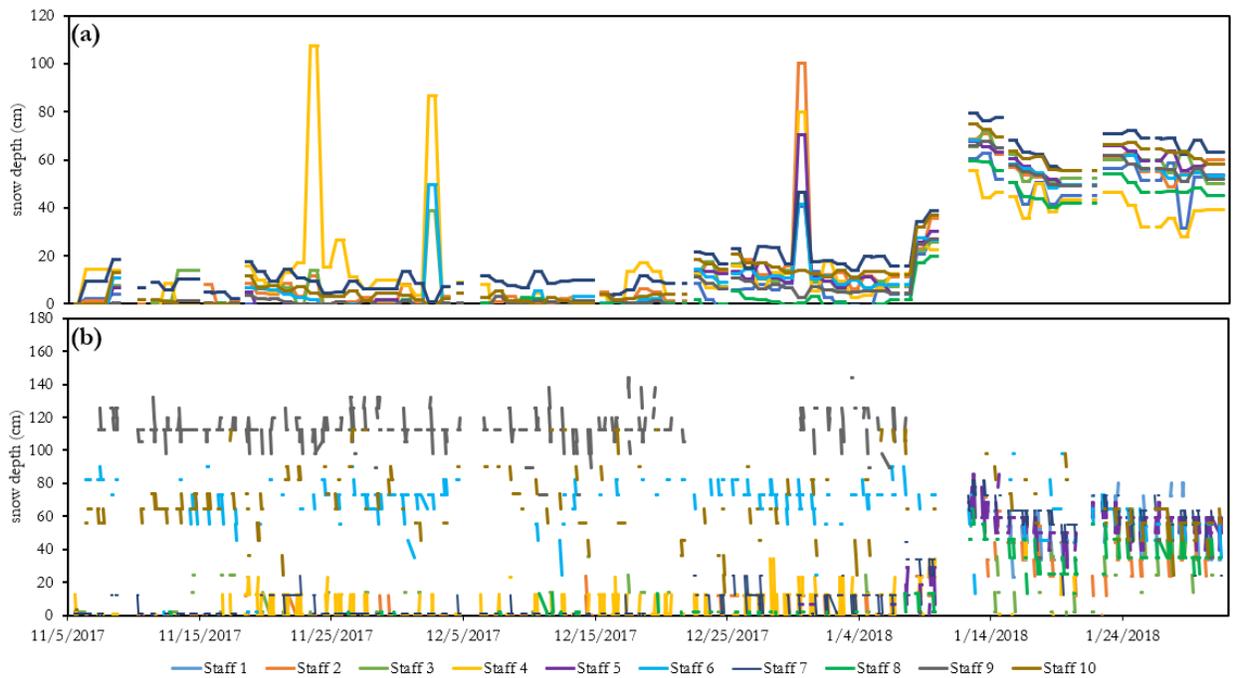


Figure 3-6: Swamp Angel staff depths (a) via manual delineation and (b) with automated delineation with error correction using maximum allowable hourly change based on the site snow depth change histogram

3.2 Site Accuracies

Accuracies varied both between sites and between staffs at the same site (Table 3-2). Joe Wright and Columbine had the best results with average absolute errors of 19 and 23 cm, respectively, when compared to manual delineation. Red Mountain and Swamp Angel both had much higher average absolute errors of 70 and 45 cm, respectively. The errors also vary widely between staffs at each site with Joe Wright having the most consistent performance across all staffs while the others showed larger disparities. The general pattern here reflects the apparent background of each staff and for what fraction of the year it was snow-covered. Red Mountain Pass displayed high errors as well as a wide range of error. The Swamp Angel accuracies ranged widely but a few of the staffs, generally those furthest from each other in the images, had errors between 20 and 30 cm.

Table 3-2: Average absolute error in cm for staffs at (a) Joe Wright, Columbine, and Red Mountain; and (b) Swamp Angel

(a)		Average Absolute Error in cm								
Site	Staff 1	Staff 2	Staff 3	Staff 4	Staff 5	Average				
Joe Wright	26	22	18	24	27	23				
Columbine	23	15	31	16	9	19				
Red Mountain	107	116	34	45	50	70				
(b)										
Staff Number	1	2	3	4	5	6	7	8	9	10
Swamp Angel	117	11	25	32	40	51	10	11	100	50
Swamp Angel Average	45									

3.2.1 Site Accuracies by Time of Day

For the Joe Wright and Columbine, the most successful sites, snow depth measurement accuracies were evaluated by time of day in order to examine whether restricting times of day could improve image recognition. At Joe Wright error was lower in the morning before 8:15 and in the evening after 15:45 (Figure 3-7a). The Joe Wright site is bounded on the east and west by steep mountains which causes the area to be illuminated by indirect light in the mornings and evenings. The Columbine site errors (Figure 3-7b) are also best in times of day with more indirect light and

fewer shadows since that site is bounded only on the west by mountains and at a further distance than Joe Wright.

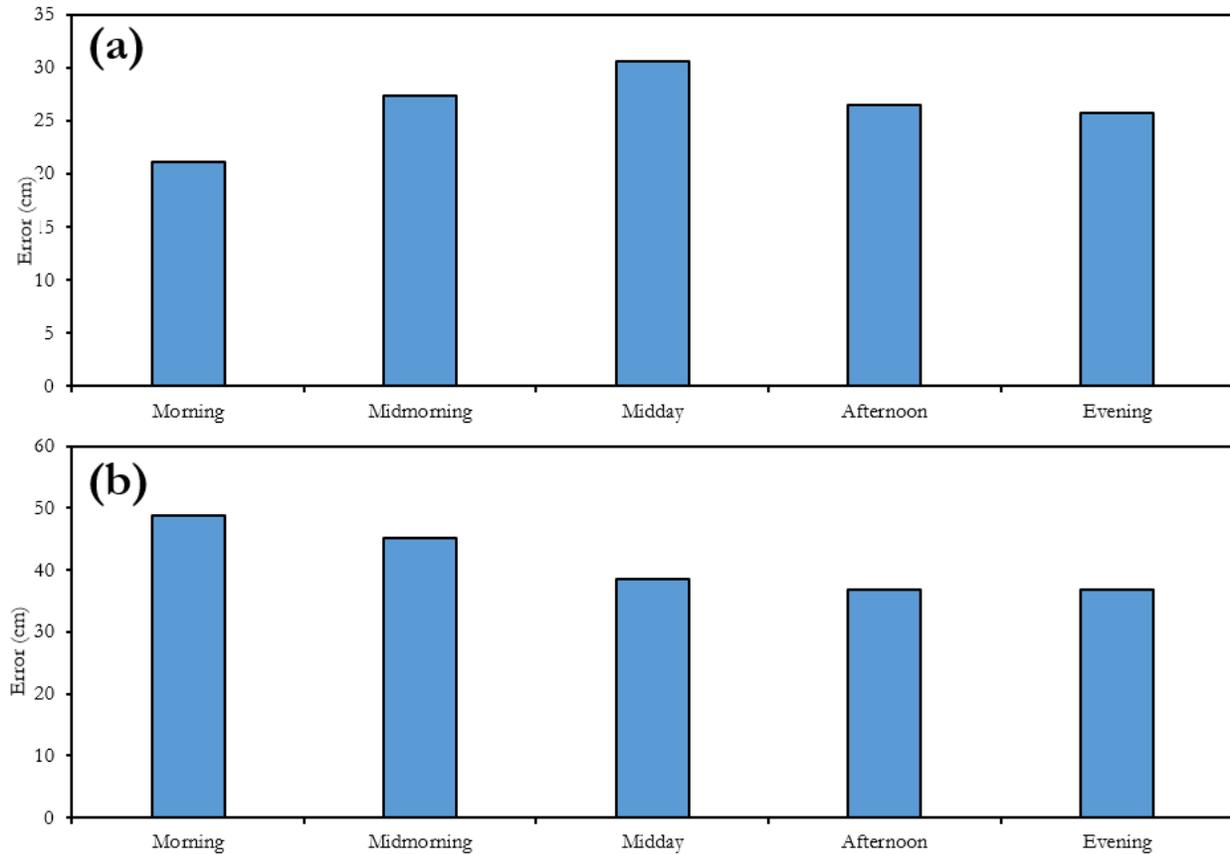


Figure 3-7: Site error by time of day at (a) Joe Wright and (b) Columbine. Morning is defined as 5:45-8:15, midmorning is defined as 8:15-10:45, midday is defined as 10:45-13:15, afternoon is defined as 13:15-15:45, and evening is defined as 15:45 to 18:15. Since the images were taken every 30 minutes from 6:00 to 18:00 a roughly equal number of images fall into each time category.

3.3 Optimal Calibration and Detection Process

3.3.1 Calibration Image Set Size and Content

An important question to consider before utilizing this technology is how many images must be manually delineated in order to produce a reliable detector in the calibration stage. The desktop computers used for this analysis (Methods 2.5.3) could process a maximum of 250 calibration images before running out of memory. Detectors trained with subsets a 250-image dataset

comprised of examples from all sites maintained consistent accuracies with sets as small as 60 images (Figure 3-8). Restricting calibration images by time of day only changed errors by 1% and didn't affect them in a consistent direction. In this application, calibrating with smaller datasets on lower performance computers can produce robust detectors.

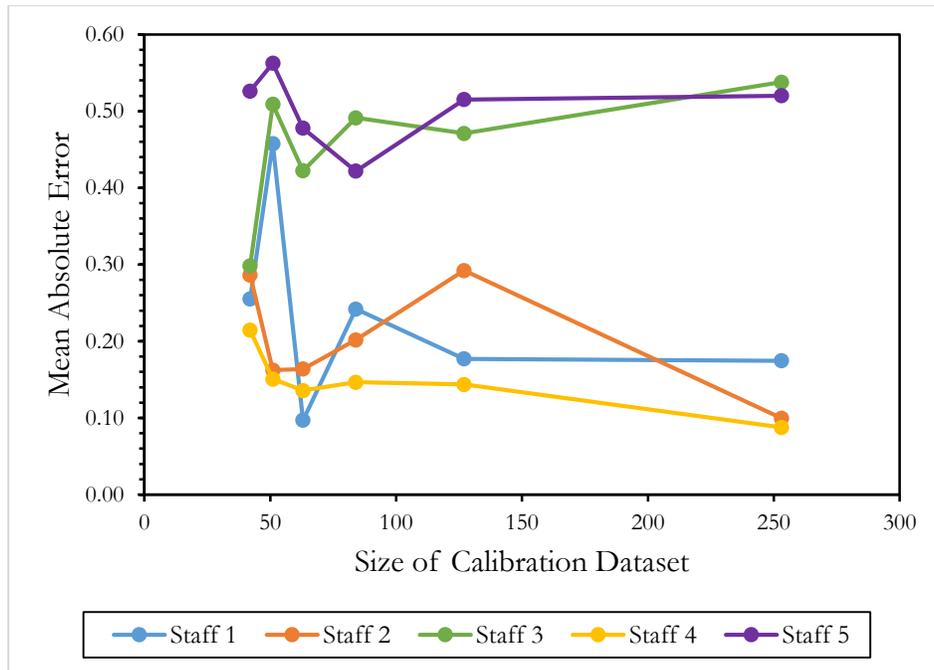


Figure 3-8: Calibration set size and detector performance for Joe Wright.

3.3.2 Cross Calibration

Another important question when using this technology at multiple sites is whether calibrating detectors individually for each site is superior to calibrating a single detector for all sites. Although Columbine and Swamp Angel each performed best when calibrated using only images from only those sites, the detector calibrated with images from all sites was only 5 percentage points less accurate at those sites and was the best detector for the Joe Wright site. Training a single detector from all image sets reduces computation time and simplifies the detection stage when working with multiple sites, but at the cost of a small performance loss at certain sites.

Table 3-3: MAE's in percent for detections calibrated with images from sites on the left and applied to the sites above. All sites performed best with a detector calibrated with an image set including all sites with the exception of Swamp Angel. Outside of the combined dataset, only Columbine did not perform best using a detector trained exclusively on images from that site.

Mean Absolute Errors		Validation Site		
		Joe Wright	Columbine	Swamp Angel
Calibration Image Set	Joe Wright	40%	16%	35%
	Columbine	47%	17%	23%
	Swamp Angel	61%	23%	14%
	Combined All Sites	30%	16%	19%

3.3.3 Optimal Parameters for Calibration and Detection

A sensitivity analysis on the input parameters for the calibration and detection stages showed that one of four calibration parameters and five of the six detection parameters had an effect greater than 5 percentage points on final snow depth accuracy as measured by mean absolute error (Table 3-4). In the calibration stage, the only variable that affected the detector accuracy was the number of iterative stages the detector underwent during calibration (Figure 3-9) with two or more stages drastically improving accuracy. In the detection stage, nearly every parameter had an effect.

Table 3-4: List of (a) calibration and (b) detection parameters which influenced detection accuracy by more than 5 percentage points of MAE with optimum ranges provided if detection accuracies were affected

	Affected Accuracies	Possible Range	Optimum Range
<i>a) calibration parameters</i>			
ObjectTrainingSize	No	Any	
NumStages	Yes	1-8	> 2
NegativeSamplesFactor	No	2-8	
MaxWeakLearners	No	256-8192	

<i>b) detection parameters</i>			
NumScaleLevels	Yes	1-8	>4
WindowStride	Yes	1-8	<4
SelectStrongest	Yes	True/False	True (on)
MinSize	Yes	Any	Set to null
MaxSize	Yes	Any	Size of largest staff at no snow
Threshold	No	-1 to 1	

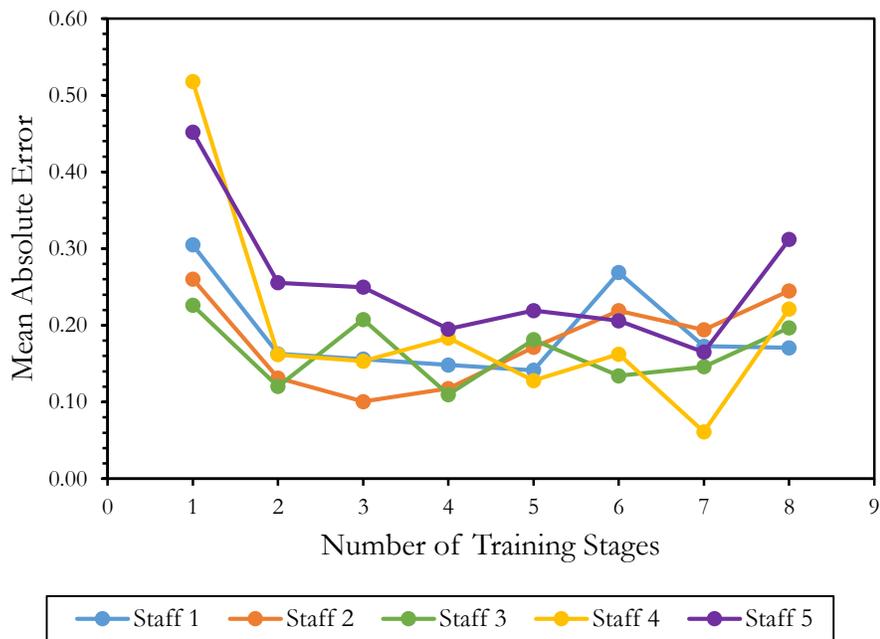


Figure 3-9: Effect of number of calibration stages on detector accuracy based on the Joe Wright site

Two parameters with clear and large effects on detection accuracies were activating the detector’s internal “select strongest” method for reconciling overlapping bounding boxes and increasing the number of scale levels (Figure 3-10). Lower scale levels had reduced accuracies and a

wider spread for all staffs in the image while the highest possible setting caused a slight increase in error. The “select strongest” result was much more straightforward, increasing accuracies by approximately 30% when in use. Detailed accuracy effects of other parameters are displayed in detail in Appendix C.

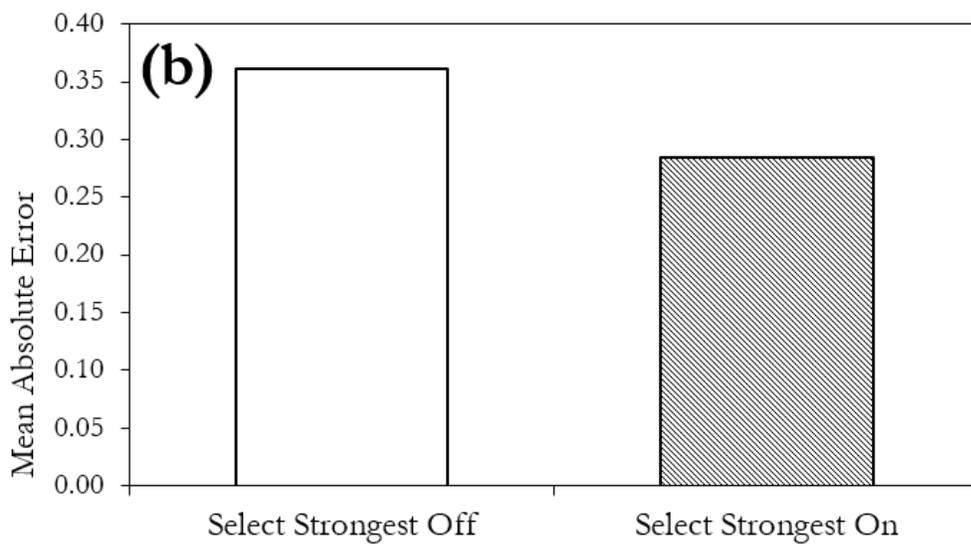
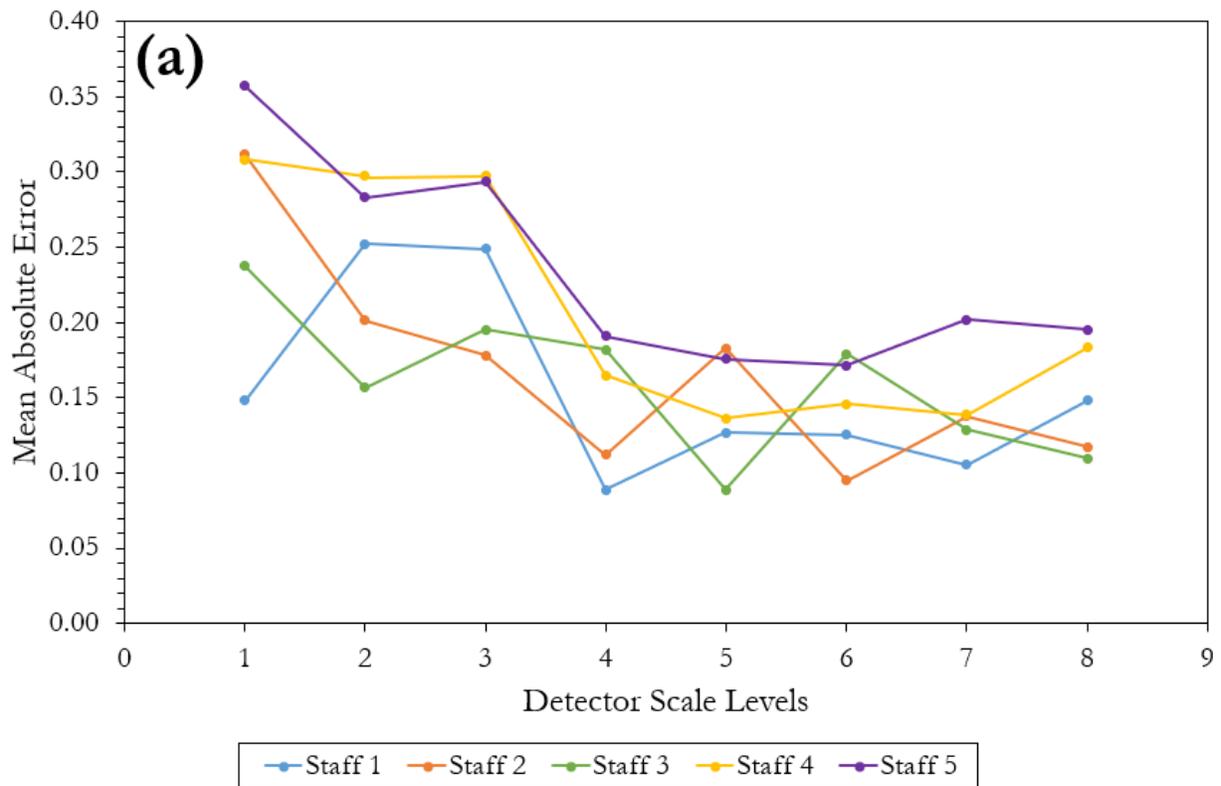


Figure 3-10: Detection parameters that affect accuracy: (a) Number of scale levels and (b) activating the detector’s internal method for reconciling overlapping returns.

Histogram equalization of calibration images has been used to increase performance of image recognition algorithms (R. Fernandes, personal communication, Feb 27 2019) and was evaluated with the Joe Wright dataset. Histogram equalization of images had a negative effect on accuracy when applied to calibration and detection separately, as well as when applied to both steps individually (Figure 3-11).

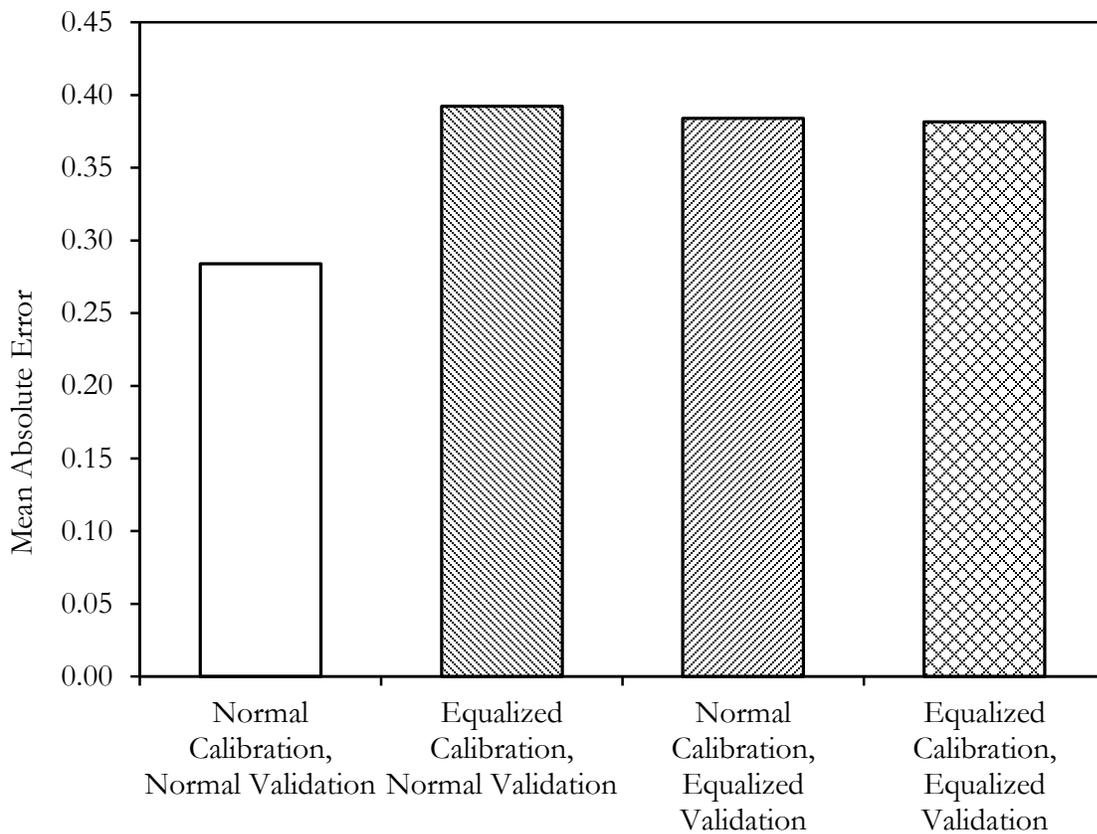


Figure 3-11: Accuracies with no histogram equalization, with histogram equalization on the calibration dataset only, histogram equalization on the detected images only, and with histogram equalization on both sets of images

3.4 Historical Joe Wright Snow Depth Distribution

Joe Wright was the most successful site and showed the greatest spatial variability with increased snow depth at distances further from the SNOTEL snow pillow and camera. The staffs at this site were arranged to roughly follow the discontinued snow course which was discontinued in 1989 (Figure 2-2a). A depth transect was measured along this snow course in water years 2006, 2007, and 2008 by Steven Fassnacht. These snow courses took depth measurements in 0.25 to 1 meter increments and could approximately be matched with the Joe Wright snow staffs present in water year 2018.

The 2006-2008 measurements at the distances matching the snow depth staffs showed that the snow depths along the transect were linearly correlated with the SNOTEL depth measurements (Figure 3-12). These strong linear correlations allow a scaling factor to be found for each distance which can approximate snow depths at that location from the SNOTEL snow depth. Since the lines have intercepts of 0, multiplying the SNOTEL snow depth by the slope of the historical best fit line of a given distance produces an estimate of snow depth at that distance.

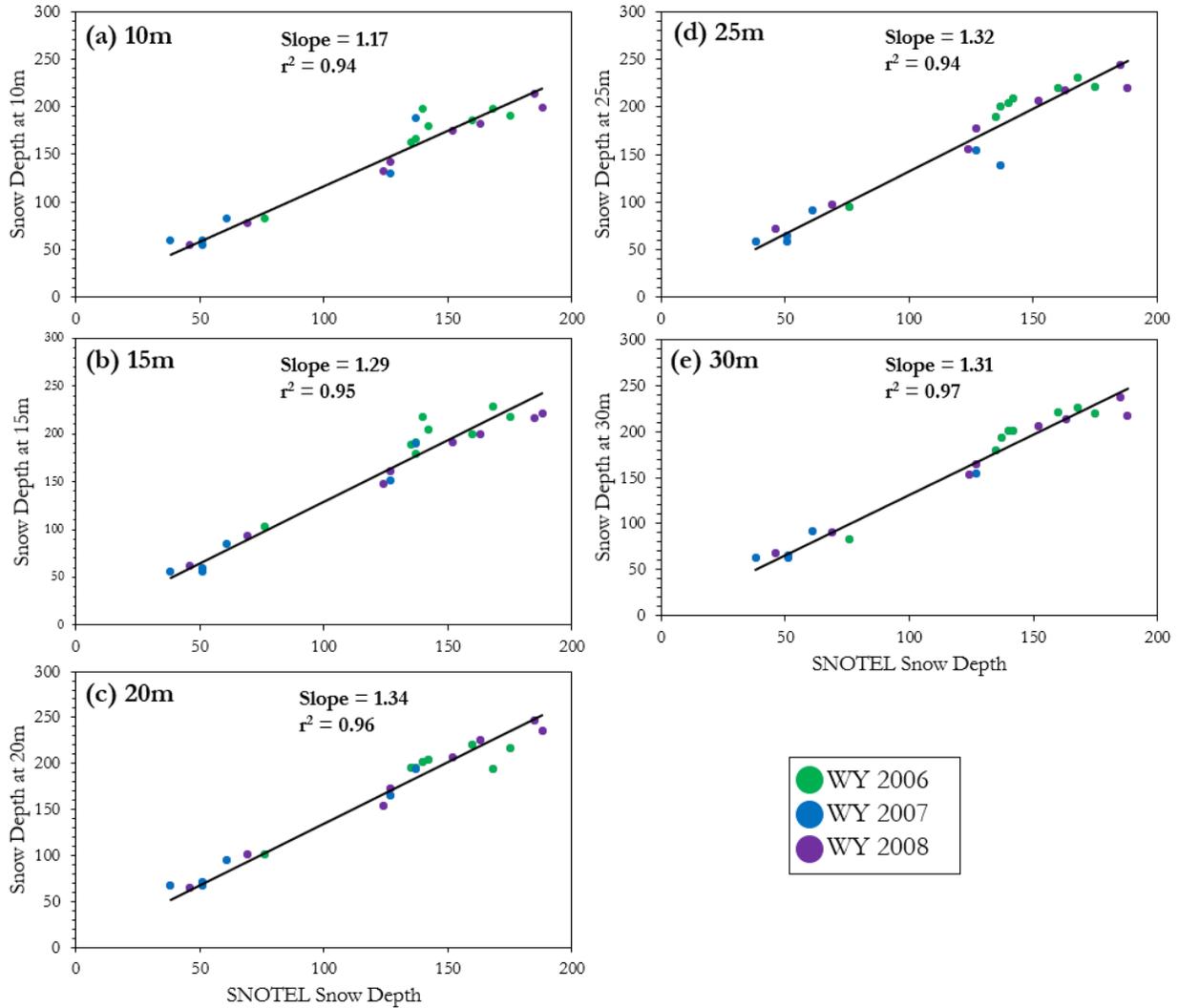


Figure 3-12: Water Year 2006-2008 transect data for each of the distances represented by a depth staff. Each point represents a depth reading at that distance.

The Joe Wright automated detection data showed snow depths increasing as distance to the snow pillow increased (Figure 3-2b). This pattern was roughly consistent with the WY 2006-2008 data but the exact scaling factors at each distance were different. (Figure 3-13).

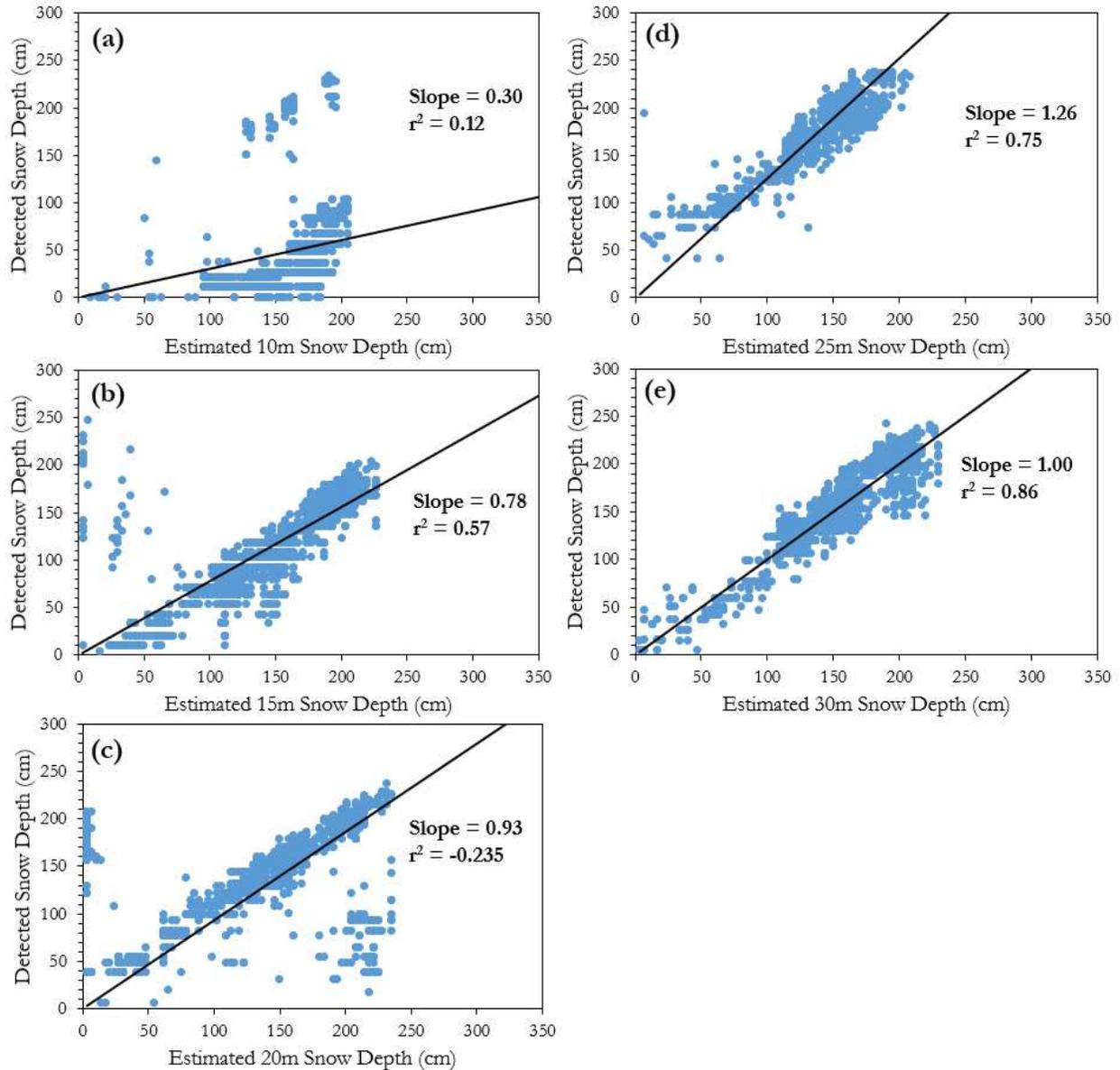


Figure 3-13: Joe Wright detected data at each distance plotted against snow depths estimated from the historically-based scaling factors (Figure 3-10) and SNOTEL depth data from WY 2018. Each point is an accepted detection return.

The staff at 10 meters from the snow pillow had very poor correlation with the expected snow depths and with the SNOTEL snow depths. The staff at 15 meters had an acceptable correlation and while lower than the expected snow depth (Slope = 0.78), matched well with the unadjusted SNOTEL depth sensor data (Slope = 1.01). The staff at 20 meters had an extremely poor correlation, but the best fit line still describes much of the data. The readings on this staff were

lower than expected from the historical estimate for that location. The staffs at 25 and 30 meters showed robust correlations with the estimated depth values. The 25 meter location in WY 2018 exceeded estimates while the depths at 30m nearly exactly matched them (Slope = 1.00).

CHAPTER 4 – DISCUSSION

An objective of this study was to develop a robust method to automatically convert images of snow depth staffs into snow depth. At the best performing sites mean absolute errors were between 10-30cm while less successful sites had much higher errors. The time lapse photography and image recognition snow depths may still be useful given that snow depths can vary by as much as 25% over distances as short as 10 meters (Figure 3-2) (Lopez-Moreno et al., 2013; Sturm et al., 2010; Fassnacht et al., 2018) and as much as 50% at the watershed scale based on terrain variables (Erxleben et al., 2002). Image recognition could be used in locations not well suited to traditional snow depth measurement techniques or in complex, unsafe terrain to produce a low-cost snow depth estimate.

4.1 Snow Depth Measurement Technologies Comparison

The ultrasonic JUDD sensor currently installed at many SNOTEL sites measures snow depths via ultrasonic reflection <juddcom.com>. Currently, these sensors cost approximately \$700 and have errors on the order of 2-3 cm (Ryan et al., 2008). These sensors have a built-in temperature sensor to adjust the sonic travel time by the ambient temperature. The Campbell SR50A-L <campbellsci.com> tends to be more accurate (Ryan et al., 2008), but is more expensive and requires an external measurement of temperature to adjust the sonic travel time. These sonic depth sensors can produce inaccurate readings during snowfall events, since the falling flakes scatter and reflect the emitted sound. They also can be inaccurate during large temperature gradients between the sensor and the snow surface (Ryan et al., 2008). A laser distance sensor is now commercial available, which has quoted accuracies on the millimeter scale, but prices upwards of \$2500 <lufft.com>.

An advantage of the image recognition and snow staff technique is scalability. Taking additional measurements on a local scale requires only placing another depth staff, rather than an additional distance sensor with associated power supply and datalogger. A network of depth sensors can produce high spatial and temporal resolution estimates of snow depth (Rice & Bales, 2010). These networks can be constructed at a lower cost by having one datalogger at a central location receiving data wirelessly from other stations

The high cost of commercial depth sensors can be significantly reduced by constructing the sensors with mass-produced parts and open-source software. Arduino-compatible sensors <www.arduino.cc> have far lower costs than the prebuilt sensors discussed above (Table 4-1) and offer unique flexibilities in power supply and data recording. Certain low-power Arduino boards can switch themselves off between measurements, greatly reducing power consumption (Arduino.cc, 2019). Additionally, constructing the sensors with identical parts means repairs can be completed at a lower cost by the researcher instead of the manufacturer.

Table 4-1: Parts breakdown and pricing for an Arduino-based snow depth measurement and transmitter units based on ultrasonic (a) and laser (b) ranging. The same components are usable for a measurement and receiver unit, with the addition of a datalogger (c).

Part	Price (\$)	Quantity
(a) Ultrasonic Transmitter Unit		
HC-SR04 Ultrasonic Distance Sensor	1.50	1
DHT22 Temperature Sensor	3.59	1
Adafruit Rechargeable Battery Pack	7.95	1
Weatherproof Enclosure	~20	1
Arduino NRF24L01 Transceiver	1.09	1
Adafruit Feather (Atmega328P)	12.50	1
Total	34.13	
(b) Laser Depth Transmitter Unit		
TF Mini LiDAR(ToF) Laser Range Sensor	37.90	1
Adafruit Rechargeable Battery Pack	7.95	1
Weatherproof Enclosure	~20	1
Arduino NRF24L01 Transceiver	1.09	1
Adafruit Feather (Atmega328P)	12.50	1
Total	79.44	
(c) Receiver Unit		
+ MicroSD Breakout Board	7.50	1
+ MicroSD Card (2GB)	1.60	1
Total	+43.23	

The HC-SR04 ultrasonic distance sensor (Table 4-1) has quoted accuracy of 1cm, equivalent to the quoted accuracy of the JUDD sensor and like the JUDD sensor, the operational accuracy is likely lower. The transceiver unit included here has a quoted 100m range, and while this would likely be lower in a complex mountain environment, it is suitable for the plot scale depth variability measurements examined here. Although this price is much lower than commercial snow depth sensors, it still scales linearly with the number of sensors installed. With the time lapse camera and image recognition method, after an initial \$100 camera, each depth staff only costs an additional \$8. Collectively, these snow depth measurement techniques are suited to different objectives and field conditions. The sensors described above need to be mounted on an arm above the snow and have batteries replaced individually, increasing installation and upkeep time. The image recognition method, while less accurate than these sensors, has the benefit of simpler installation and only

requiring a single electronic device in the field. In some locations, the camera images can be telemetered using the cellular network where transmission bandwidth is not restrictive.

4.1.1 Spatial and Temporal Resolution

The precision available in laser ranging technologies may not be entirely necessary in the context of snow science. Water forecasts are currently made with SWE measurements with precision of 0.1 inch (NRCS, 2019) and snow depth measurements with operational precision of around 2 inches (Ryan et al., 2008). For a snowpack with over a meter of depth, common in Colorado (Figure 3-2), a variance of 1 inch corresponds to less than 3% of the total, smaller than the inherent year-to-year variability of snow depths on the plot scale (Fassnacht & Patterson, 2013).

Some of these technologies also provide the ability to achieve a high temporal resolution, such as hourly, but currently quality controlled daily snowpack data are provided by NRCS and monthly data are used operationally in water forecasting in the United States (NRCS, 2019). Higher temporal resolution data are still useful for understanding snowpack processes such as hourly change due to melt, metamorphism, and precipitation presented earlier (Figure 3-1). In avalanche research hourly snow depth has been integrated into models that offer high temporal resolution avalanche forecasts (Brun et al., 1989). Hourly depth data is also very important in layer-based snowpack modelling as the accumulation rate and density of individual snow events is crucial to understanding layers formed during each event and modelling their evolution (Sturm et al., 2010).

4.1.2 Other Image Recognition Techniques Examined

Prior to the selection of the Aggregate Channel Features approach, two other methods were examined. The first was an algorithmic approach while the second involved use of a Regional Convolutional Neural Network (RCNN) (Girshick et al., 2014). The algorithmic approach was focused on combining edge detection, relative pixel color, and additional methods in order to detect

the snow staffs. Edge detection and relative redness could identify staffs from images well but required fine tuning of parameters for each image and small, hour-to-hour changes in lighting required recalibration of parameters. This could take up to an hour per image and often wouldn't result in the correct detection of all staffs. The Hough Transform (Duda & Hart, 1972) was also incorporated into this method to extract linear shapes in the images but performance was inconsistent.

The RCNN method uses a neural network architecture generally used for image classification and applies it to region detection. Many recent developments in computer vision have been based on this method (Girshick, 2015, Li et al., 2015) but using it properly requires advanced construction and fine-tuning of several layers of parameters. In addition to the high complexity of designing a neural network, its computational cost is higher than the ACF method, allowing for fewer images to be processed in the calibration stage using the same hardware.

Compared to these two approaches, the Aggregate Channel Features method is generalized, flexible, and computationally inexpensive. While its accuracy varied by time of day (Figure 3-7) it was able to detect staffs in diverse lighting conditions and without manual tuning of parameters. It also was able to ingest double to triple time images in the calibration stage than RCNN given the same computing power.

4.2 Snow Depth Variability

At the Joe Wright site, snow depths varied by as much as 50cm over distances of 5m (Figure 3-3) and this variability had a consistent pattern through multiple snow seasons (Figure 3-11). Snow depth variability can be high at local scales (Lopez-Moreno et al., 2013; Sturm et al., 2010; Fassnacht et al., 2018) and is usually caused by snow distribution factors such as canopy cover, wind, and solar radiation (Liston & Sturm, 1998). The consistency of the snow depth pattern at the Joe Wright site

within a snow season (intra-annual variability) and between different snow seasons (inter-annual variability) (Fassnacht et al., 2018) suggests that the cause is specific to that site and is consistently present through the snow season (Sturm et al, 2010). By contrast, at the Columbine site, variability was much lower, with measurements across 20 meters varying by less than 10cm throughout WY 2018 (Figure 3-4).

The consistent snow depth pattern at Joe Wright is likely due to snow transport and deposition at the site from along the nearby road and from Cameron Pass just to the south (Figure 4-1b) which was observed at multiple points in the winter during fieldwork. While the Columbine site is located in a similarly sized clearing to Joe Wright, there is a small area of mixed lodgepole pine and spruce-fir forest between the road and the SNOTEL site (Figure 4-1a). These sites are similar in slope, aspect, and vegetative cover, but vary greatly in snow depth and snow depth variability (Figures 3-2 & 3-3). Low-cost methods that capture snow depth discrepancies in areas with similar landscape characteristics can improve basin-wide snowpack storage estimates by providing a better estimate of variability (Balk & Elder, 2000).



Figure 4-1: True color satellite image of the (a) Columbine and (b) Joe Wright showing proximity and orientation of roads and forests

4.3 Snow Measurement via Terrestrial Photography

Previous studies on image recognition for snow science have focused on measurement of albedo (Corripio, 2004) and snow covered area (Arslan et al., 2017). These projects deployed time

lapse camera networks and used automated methods to convert images to data but the approach differed the project presented here. Rather than using machine learning, they manually identified regions of interest in the images from each camera. The average pixel colors in these regions were evaluated against a plate of known color in the scene to estimate albedo (Corripio, 2004). Snow covered area was retrieved by classifying each pixel as snow or non-snow (Arslan et al., 2017).

These area of interest based approaches are less applicable to snow depth retrieval because the apparent size of the staff in the image is not constant. Additionally, issues arising from shadows moving across the scene through each day are reduced by choosing regions of interest in the image with consistent lighting. The known color plate placed in the scene also allows for a baseline light condition to be established for each image, simplifying the snow presence determination.

Snow surface roughness has also been retrieved automatically from images (Manninen, 2012). These were not time lapse images at a particular site but rather photographs of a purpose-built snow roughness board placed partially into the snow at locations of interest. This board had two stripes of checkerboard markings around its edge which allowed for an algorithm tuned to locate small areas with high contrast to recognize the board's location and orientation automatically in 99% of cases. Without the checkerboard markings, initial image analysis required manual cropping of each photograph (Fassnacht et al., 2009). Once these were known, the snow surface could be identified and its roughness estimated. The success of this automated technique suggests that distinctive and high contrast patterns could help improve detection accuracy of snow depth staffs.

Garvelmann et al. (2012) set out to measure snow depth, albedo, and canopy interception at a network of 45 time lapse cameras. They reduced the cost of the cameras by using low-cost digital cameras mounted inside purpose-built housing. These cameras and housings had issues with snow buildup and occasional hardware failures due to low temperatures and moisture, but most cameras functioned through the winter. The same process used in previous studies was used to retrieve

albedo and an automated snow pixel identifying method was used to estimate canopy interception based on manually defined regions of interest positioned in the background tree canopy. They attempted to use the same snow pixel identification process to differentiate snow depth staffs placed in the cameras' field of view from the background but had inconsistent results and high errors. They manually delineated snow depths for their final assessment.

Another group (Pimentel et al., 2017) used snow depth staffs and a time lapse camera to produce a Snow Accumulation/Depletion Curve (ADC) relating snow covered area to snow depth. Using the techniques similar to Arslan et al. (2017) and Garvelmann et al. (2012) they sorted pixels in the plot into snow and non-snow and computed snow covered area over the plot. They also used a red-painted depth staff to find snow depth at a key point in the plot by applying the same snow pixel differentiation algorithm. The application of this algorithm was simplified by specifying a small area of interest and by placing the staff so that the background was entirely snow covered even at low snow depths.

In the context of previous methods, this research is the first time that image recognition and machine learning have been applied to snow depth measurement from terrestrial imagery. This technique has accuracies within 10-20cm in ideal conditions (Table 3-2). Operationally, unlike the techniques based on classifying pixels previously used to determine albedo and snow covered area, it does not require a baseline color reference to be placed in the image or an area of interest to be defined by the user.

4.4 Recommendations

Operating a network of varied sites across the state of Colorado during the winter of 2017-2018 allowed the challenges associated with establishing and maintaining these sites to be examined. There are a number of recommendations that should be followed when setting up and maintaining

future sites, as well as when processing images in order to achieve best results. In addition, there are additional resources that future researchers could utilize to improve upon this technique.

4.4.1 Site Setup Recommendations

The installed site hardware worked as intended with the snow depth stakes remaining upright and the cameras working through a season in cold, wet, and windy conditions. Although the physical sites performed well, certain aspects of their configuration interfered with the image recognition process. The most straightforward impediment to image recognition was snow staffs occluding each other. The Swamp Angel site had low accuracies because the staffs were placed too closely and overlapped each other (Figure 5-1a). Other sites such as Columbine with more widely placed staffs were more successful (Figure 5-1b).

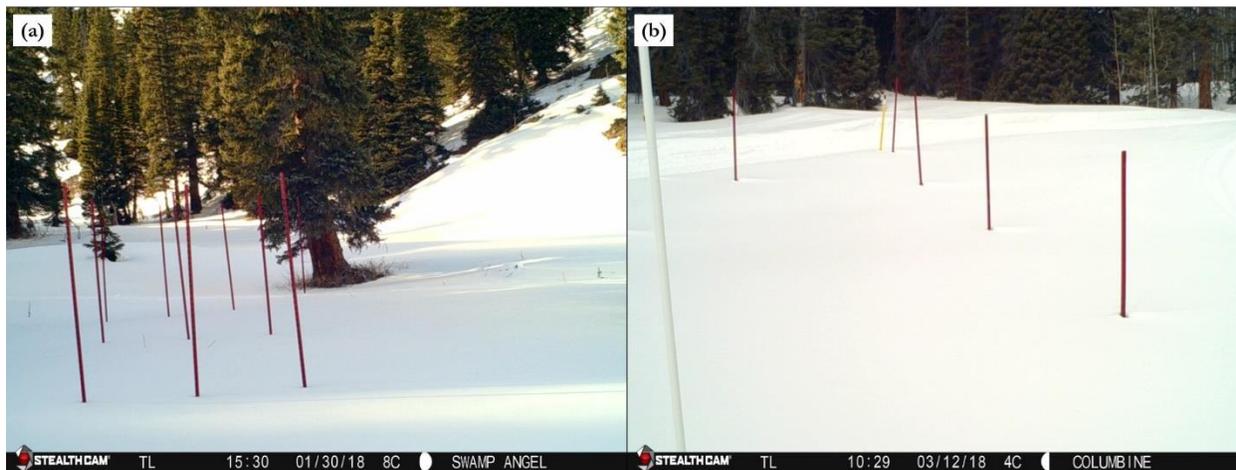


Figure 4-2: (a) Overlapping staffs at Swamp Angel and (b) separate staffs at Columbine

Staff accuracy was strongly affected by the staffs' apparent backgrounds with snow-covered backgrounds yielding higher accuracies than those with forested backgrounds. This is due to the near constant color of the snow having a high contrast with the red staffs in all lighting conditions whereas the grays and browns of the forest background were quite similar to the staff red, especially

when the staffs are in shade (Figure 4-3). This is also the reason accuracies were higher when the snowpack was deeper, as a deeper snowpack made many staff backgrounds more uniformly colored.

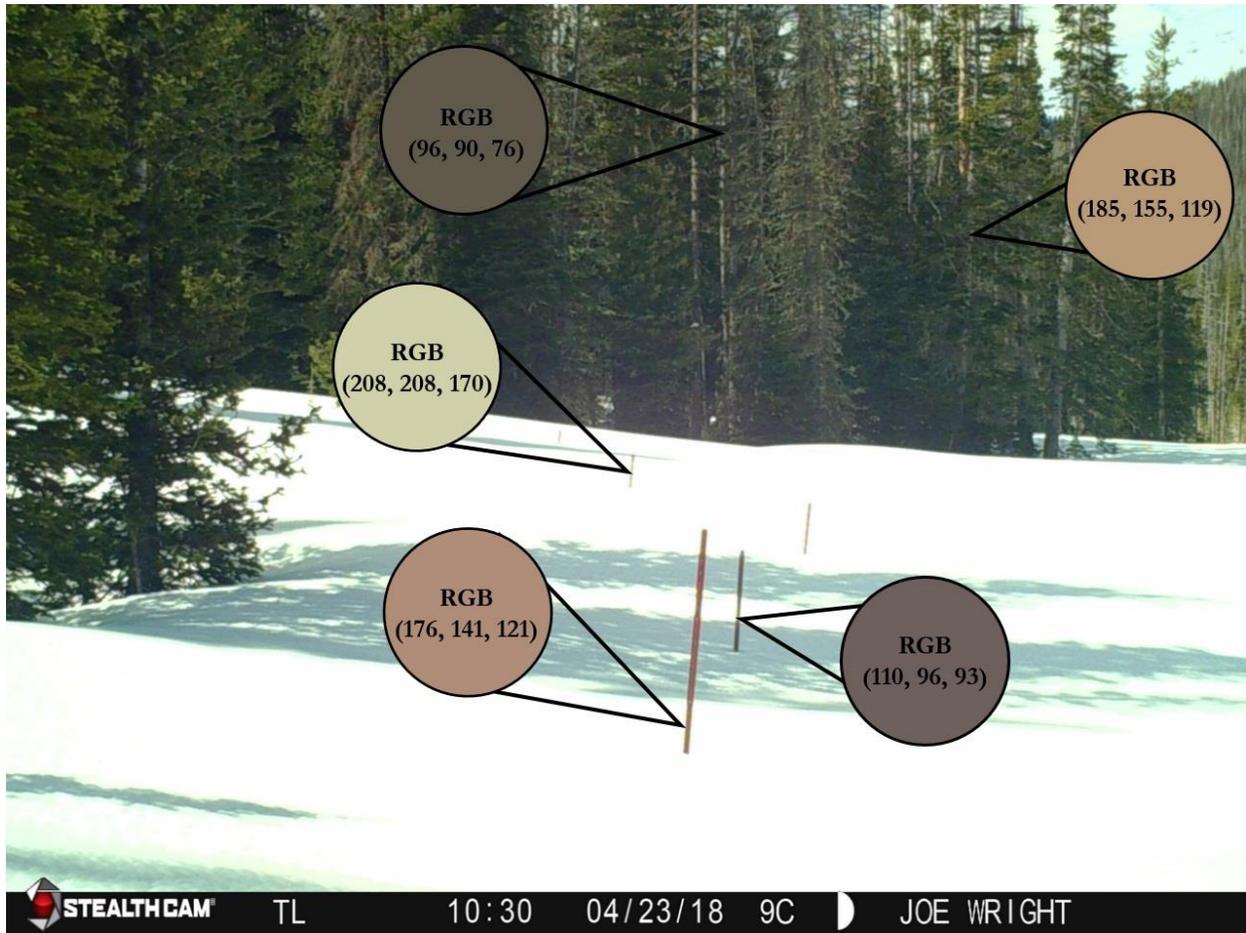


Figure 4-3: Color comparison between dense forest background (upper left & upper right) and three snow staffs (lower right, middle left, lower left) demonstrating similarities in forest background colors and the wide range of colors found in the foreground staffs

4.4.2 Site Equipment Recommendations

Placing the staffs and camera to maximize the amount of snow-covered background is straightforward when working with a sloping or expansive site. When faced with a confined or flat site, this can be more difficult. At Columbine and Joe Wright, there were expansive open areas (Figure 2-2b, 2-2a) that allowed the furthest staffs to be placed at least 10m from the nearest

forested area. At Red Mountain Pass, the forest edge was less than 2 meters from the furthest line of staffs, causing a larger fraction of the background to be forested (Figure 2-2e).

At sites with limited open space, placing the camera at a higher angle above the staffs (Figure 4-2) can reduce the forested fraction of the background. This was difficult to perform at the SNOTEL sites since the camera mount on the meteorological tower was temporary and climbing on the tower to a sufficient height would have been dangerous. Utilizing a more robust commercial camera mount rather than the fabric strap used in this first season would allow the camera to be placed at a higher angle looking down at the site, increasing the amount of snow in the background of each of the staffs.

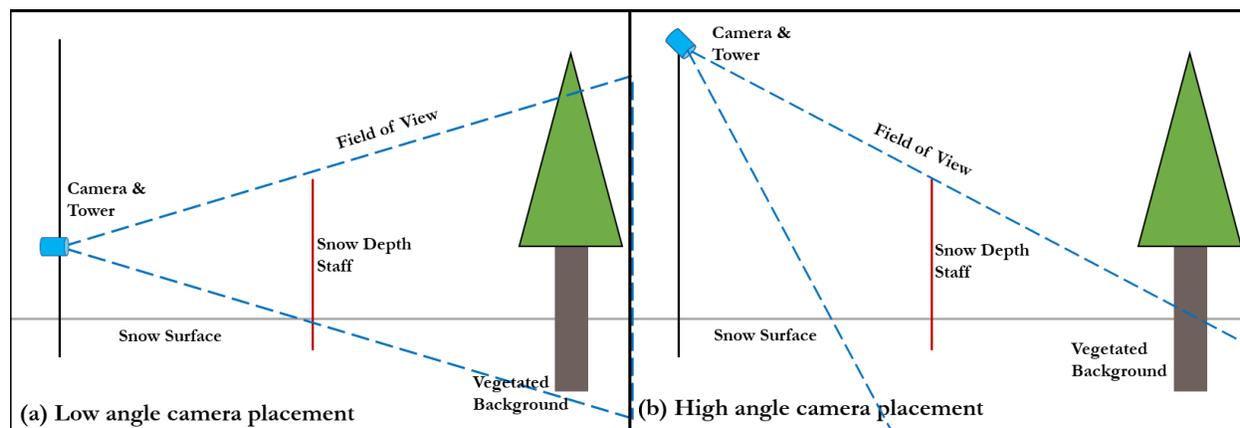


Figure 4-4: Cameras mounted at (a) staff height and at a (b) higher angle, demonstrating that the field of view of the camera at a higher placement will include a larger area of snow and a smaller area of vegetation in the background.

The steel fencepost staffs worked well in this application, remaining upright despite frequent high winds, surface water flow, and soil frost heaving. However, each staff weighs over 7 kilograms, limiting the distance they can be carried. Their 3-meter length is necessary for peak snow accumulation at many high elevation Colorado sites (Appendix D) but also makes handling cumbersome and vehicle transport difficult. Several alternative staff materials were considered and utilizing bamboo or plastic pipes is recommended. They may move more than the steel fenceposts but are far lighter and could be split into smaller sections to be assembled on site, which would

improve portability. The lighter plastic pipes or bamboo could be attached to a piece of rebar driven into the ground, requiring less hardware and heavy lifting than pounding the fenceposts into the ground.

The red color of the staffs differentiated them from the background sufficiently to enable consistent detections of staff presence in images even if the staff dimensions weren't highly accurate (Table 3-2). Another investigation into snow depth via automated image recognition (R. Fernandes, personal communication, Feb 27 2019) painted red and black squares on their staffs instead of keeping a constant color and saw increased accuracy. A study on snow surface roughness was able to automatically determine that parameter from images using a purpose-built image recognition program and used a pattern of light and dark squares around the board's edges to make the board easier to detect (Manninen et al, 2012). Unique patterns in the scene are also simpler to recognize for computer vision algorithms which could reduce false detections and improve snow depth data.

4.4.3 Image Recognition Recommendations

If this method were deployed operationally, some of the steps that were done manually in this research could be automated. The detector performs best when snow is present and in this work the dates of snow appearance and disappearance were manually found. Choosing an area of interest on the ground and finding its albedo relative to the image could determine whether snow is present and decide whether to use the detector or set all snow depths to zero. Additionally, if strong trends were observed in accuracies by time of day, images could be filtered to only make detections in images within a certain time range.

One of the limitations of this investigation was the use of desktop computers for detector calibration and evaluation. This allows the technique to be easily and inexpensively reproduced but limits the size of calibration sets and image processing speed. High-performance computing could

greatly increase the speed and accuracy of this technique, but supercomputers have been expensive and technically challenging to operate. Recently, technology companies such as Google, Amazon, and others have introduced low-cost, user-friendly, cloud-based high-performance computation services, many of which are optimized for deep learning and image recognition (Hale, 2018). Using these services would allow detector calibration and image processing to be done from any device with an internet connection.

In addition to these cloud-based solutions, there are a variety of purpose-built, proprietary image recognition software packages that could likely increase detection accuracy and speed over the Matlab implementation of Aggregate Channel Features technique used in this work. There are also generalized pretrained detectors which can be specialized to a target object with a small number of additional calibration images. These pretrained detectors can achieve higher accuracies but at a much higher computational cost especially when applied to large, high resolution images.

CHAPTER 5 – CONCLUSION

The pilot study conducted during the winter of 2017-2018 is a first step in the use of image recognition for snow depth estimation. The resultant data had high error but provided information on factors that could be controlled in the physical site setup and later image processing in order to minimize error. Operation of camera sites across the state of Colorado additionally provided important insight into best practices for future studies. The technology at this stage could be used operationally in areas poorly suited to current measurement techniques where a rough estimate of snow depth is needed.

The raw data returned from the automated detection protocol had mean error as high as 70cm, necessitating a robust process of error identification. Though error identification lowered mean error to around 20cm it eliminated data points and made data availability for a specific time step uncertain. More advanced and computationally intensive image recognition algorithms could be more successful in accurately identifying staffs thereby increasing accuracy and data representativeness across the period of record.

Maintenance of the camera sites consisted of only changing batteries and an SD card while installation required no power tools or specialized equipment. Only one staff across all sites visibly moved despite most sites being subjected to surface flow and frost heaving during the winter. This simplicity, reliability, and lack of moving parts make this technology an option for snow depth measurement in remote areas and designated wilderness areas.

The most exciting future possibility based on this work is utilization of cloud-based image recognition services. These could provide more accurate detectors, faster image processing, and allow any device with an internet connection to convert images into snow depths. A cloud-based solution would also open this technology to the public, allowing images of specific locations or

landmarks to be converted to snow depths, providing valuable data in areas that are not well represented in current measurement networks.

In its present state this method can provide snow depth data with absolute errors as low as 20cm on a properly arranged site. And while this study does provide guidelines for error minimization in both site setup and data processing, the most exciting next steps involve moving the image recognition to cloud-based services. Not only would this improve accuracies and reduce computer hardware requirements for the user, it would open the possibility of snow depth retrieval from images taken in remote areas by the anyone, rather than only from time lapse cameras installed at dedicated sites.

REFERENCES

- Arslan A. N., Cemal Melih Tanis, Sari Metsämäki, Mika Aurela, Kristin Böttcher, Maiju Linkosalmi, & Mikko Peltoniemi. (2017). Automated Webcam Monitoring of Fractional Snow Cover in Northern Boreal Conditions. *Geosciences*, 7(3), 55.
- Arduino. (2019). Arduino Sleep. Retrieved from <https://playground.arduino.cc/Learning/ArduinoSleepCode/>
- Bair, E., Rittger, K., Davis, R., Painter, T., & Dozier, J. (2016). Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory. *Water Resources Research*, 52(11), 8437-8460.
- Balk, B., Elder, K., & Baron, J. (1998). Using geostatistical methods to estimate snow water equivalence distribution in a mountain watershed. In *Western Snow Conference* (pp. 100-111).
- Balk, B., & Elder, K. (2000). Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed. *Water Resources Research*, 36(1), 13-26.
- Brun, Martin, Simon, Gendre, Coleou, & Brun, E. (1989). An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of Glaciology, Cambridge, England*, 35(121), 333-342.
- Church, J. E. (1914). Recent Studies of Snow in the Western United States. *Quarterly Journal of the Royal Meteorological Society*, 40, 43-52. doi:10.1002/qj.49704016905
- Clifford, D. (2010). Global estimates of snow water equivalent from passive microwave instruments: History, challenges and future developments. *International Journal of Remote Sensing*, 31(14), 3707-3726.

- Clow, D., Nanus, L., Verdin, K., & Schmidt, J. (2012). Evaluation of SNODAS snow depth and snow water equivalent estimates for the Colorado Rocky Mountains, USA. *Hydrological Processes*, 26(17), 2583-2591.
- Corripio, J. (2004). Snow surface albedo estimation using terrestrial photography. *International Journal of Remote Sensing*, 25(24), 5705-5729.
- Dollár, P., Tu, Z., Perona, P., & Belongie, S. (2009). Integral channel features.
- Dollár, P., Appel, R., Belongie, S., & Perona, P. (2014). Fast Feature Pyramids for Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8), 1532-45.
- Dozier, J., Bair, E., & Davis, R. (2016). Estimating the spatial distribution of snow water equivalent in the world's mountains. *Wiley Interdisciplinary Reviews: Water*, 3(3), 461-474.
- Elder, K., Dozier, J., & Michaelsen, J. (1991). Snow accumulation and distribution in an alpine watershed. *Water Resources Research*, 27(7), 1541-1552.
- Erxleben, J., Elder, K., Davis, R., Hardy, J. P., Pomeroy, J. W., Elder, K., & Klein, A. G. (2002). Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains. *Hydrological Processes*, 16(18), 3627-3649.
- Fassnacht, S. R., & Patterson, G. G. (2013). Niveograph interpolation to estimate peak accumulation at two mountain sites. In *Cold and Mountain Region Hydrological Systems Under Climate Change: Towards Improved Projections* (pp. 59-64). Gothenburg, Sweden: Proceedings of H02, IAHS-IAPSO-IASPEI Assembly.
- Fassnacht, S. R., & Hultstrand, M. (2015). Snowpack variability and trends at long-term stations in northern Colorado, USA. *Proceedings of the International Association of Hydrological Sciences*, 371, 131-136. doi:10.5194/piahs-371-131-2015

- Fassnacht, S. R., Brown, K. S. J., Blumberg, E. J., López-Moreno, J. I., Covino, T. P., Kappas, M., ... Kashipazha, A. H. (2018). Distribution of snow depth variability. *Frontiers of Earth Science*, 12(4), 683-392. doi:10.1007/s11707-018-0714-z
- Garvelmann, J., Pohl, S., & Weiler, M. (2012). Applying a time-lapse camera network to observe snow processes in mountainous catchments. *Hydrology and Earth System Sciences Discussions*, 9(9), 10687-10717.
- Hale, J. (2018, October 29). Best Deals in Deep Learning Cloud Providers. Retrieved from <https://towardsdatascience.com/maximize-your-gpu-dollars-a9133f4e546a>
- Jagt, B., Lucieer, A., Wallace, L., Turner, D., & Durand, M. (2015). Snow Depth Retrieval with UAS Using Photogrammetric Techniques. *Geosciences*, 5(3), 264-285.
- Jonas, Marty, & Magnusson. (2009). Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology*, 378(1), 161-167.
- Kashipazha, A., Fassnacht, Steven, Arabi, Mazdak, Kampf, Stephanie, & Laituri, Melinda. (2012). Practical Snow Depth Sampling around Six Snow Telemetry (SNOTEL) Stations in Colorado and Wyoming, United States, ProQuest Dissertations and Theses.
- Liston, G., & Sturm, M. (1998). A snow-transport model for complex terrain. *Journal of Glaciology*, 44(148), 498-516.
- Logan, L. (1973). Basin-wide water equivalent estimation from snowpack depth measurements. *Role of Snow and Ice in Hydrology*, IAHS AIHS Publ, 107, 864-884.
- López-Moreno J. I., Fassnacht S. R., Beguería S., & Latron J.. (2011). Variability of snow depth at the plot scale: Implications for mean depth estimation and sampling strategies. *The Cryosphere*, 5(3), 617-629.
- López-Moreno J. I., Fassnacht S.R. , Heath J.T. , Musselman K. , Revuelto J., Latron J., . . . Jonas T. (2012). Small scale spatial variability of snow density and depth over complex alpine terrain:

- Implications for estimating snow water equivalent. *Advances in Water Resources*, 55(C), 40-52.
- López-Moreno, J., Pomeroy, J., Revuelto, J., & Vicente-Serrano, S. (2013). Response of snow processes to climate change: Spatial variability in a small basin in the Spanish Pyrenees. *Hydrological Processes*, 27(18), 2637-2650.
- Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, 10(11), 114016.
- Manninen, T., Anttila, K., Karjalainen, T., & Lahtinen, P. (2012). Instruments and Methods Automatic snow surface roughness estimation using digital photos. *Journal Of Glaciology*, 58(211), 993-1007.
- Marr, J. C. (1940). *Snow Surveying*. Washington, D.C.: U.S. Dept. of Agriculture Misc Pub No 380. 45pp.
- Meiman, James R., Froehlich, Henry A., & Dils, Robert E. (1968). Meiman, J. R., H. Froehlich, and R. E. Dils. *Snow Accumulation in Relation to Elevation and Forest Canopy*, Original: Report. 15 p.
- Meromy, L., Molotch, N. P., Link, T. E., Fassnacht, S. R., & Rice, R. (2012). Subgrid variability of snow water equivalent at operational snow stations in the western USA. *Hydrological Processes*, 27(17), 2383-2400. doi:10.1002/hyp.9355
- Mizukami, N., & Perica, S. (2008). Spatiotemporal Characteristics of Snowpack Density in the Mountainous Regions of the Western United States. *Journal of Hydrometeorology*, 9(6), 1416-1426.
- NASA. (2019). NASA Airborne Snow Observatory. Retrieved from <https://aso.jpl.nasa.gov/>

- Nolan M., Larsen C., & Sturm, M. (2015). Mapping snow depth from manned aircraft on landscape scales at centimeter resolution using structure-from-motion photogrammetry. *The Cryosphere*, 9(4), 1445-1463.
- NRCS. (2019). Program History. Retrieved from <https://www.nrcs.usda.gov/wps/portal/wcc/home/aboutUs/history/>
- Pimentel, R., Herrero, J., & Polo, M. (2017). Subgrid parameterization of snow distribution at a Mediterranean site using terrestrial photography. *Hydrology and Earth System Sciences*, 21(2), 805-820.
- Rice, R., & Bales, R. C. (2010). Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California. *Water Resources Research*, 46(3). doi:10.1029/2008wr007318
- Rondeau-Genesse, Trudel, & Leconte. (2016). Monitoring snow wetness in an Alpine Basin using combined C-band SAR and MODIS data. *Remote Sensing of Environment*, 183, 304-317.
- Ryan, W., Doesken, N., Fassnacht, S., Buttle, James M., Klein, Andrew G., & Pomeroy, John W. (2008). Preliminary results of ultrasonic snow depth sensor testing for National Weather Service (NWS) snow measurements in the US. *Hydrological Processes*, 22(15), 2748-2757.
- Sexstone, G., & Fassnacht, S. (2014). What drives basin scale spatial variability of snowpack properties in northern Colorado? *The Cryosphere*, 8(2), 329-344.
- Steiner L., Meindl M., Fierz C., & Geiger A. (2018). An assessment of sub-snow GPS for quantification of snow water equivalent. *The Cryosphere*, 12(10), 3161-3175.
- Sturm, M., & Wagner, A. (2010). Using repeated patterns in snow distribution modeling: An Arctic example. *Water Resources Research*, 46(12), N/a.
- Sturm M, Taras B, Liston GE, Derksen C, Jonas T, Lea T. (2010) Estimating snow water equivalent using snow depth data and climate classes. *J Hydrometeorol* 2010;11:1380–94.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.

CVPR (1), 1, 511-518.

Yang, B., Yan, J., Lei, Z., & Li, S. Z. (2014, September). Aggregate channel features for multi-view

face detection. In *IEEE international joint conference on biometrics* (pp. 1-8). IEEE.

APPENDICES

Appendix A: Miscellaneous Figures



Figure A-1: Stealthcam time lapse game camera used at all sites



Figure A-2: Steel T posts used as snow staffs and as mounting points at sites without



Figure A-3: Camera mounted on a SNOTEL meteorological tower at Joe Wright. A similar setup was used at Columbine and Red Mountain Pass

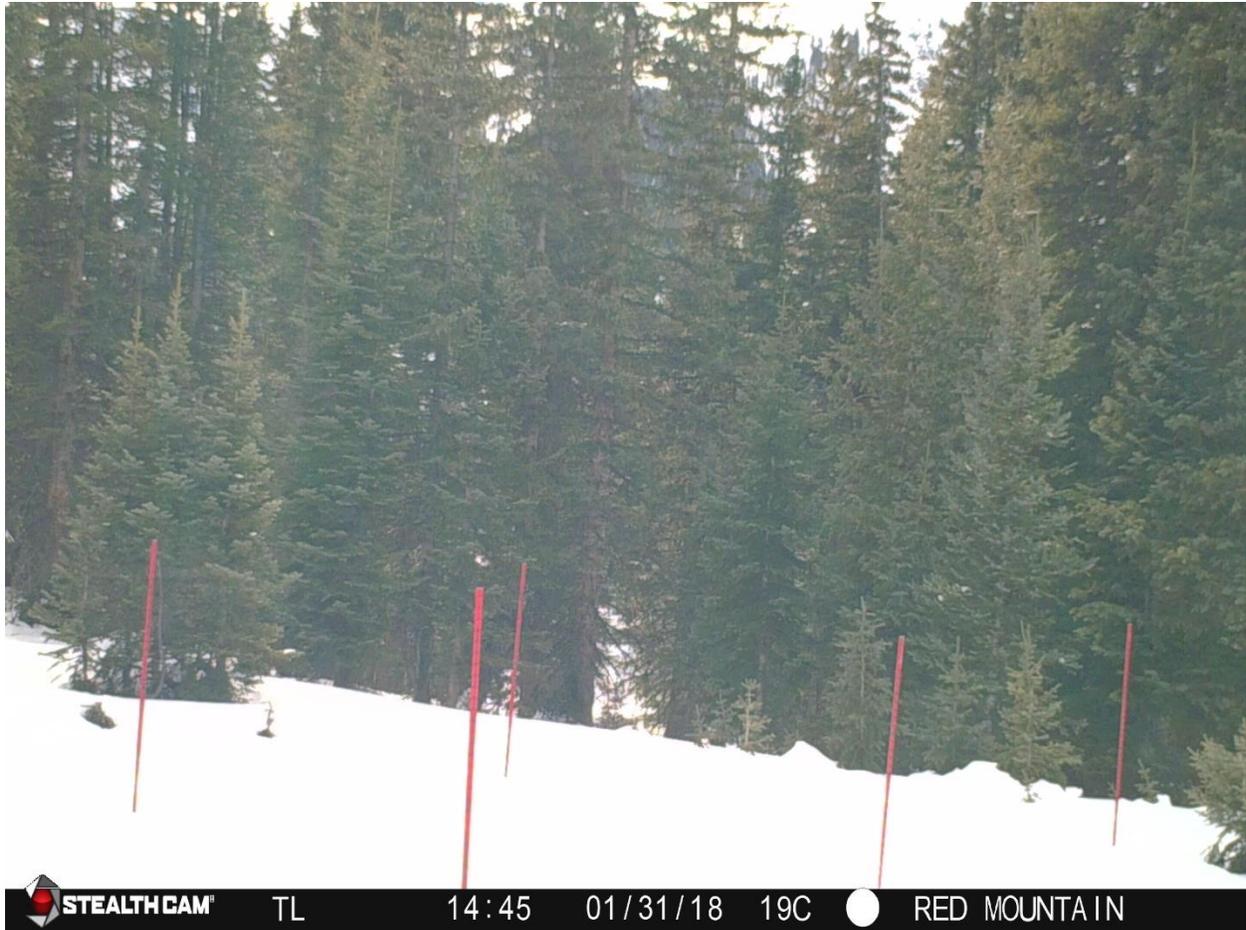


Figure A-4: View from the Red Mountain Pass camera, note the high percentage of forested background



Figure A-5: View from the Swamp Angel camera, note the close placement of staffs

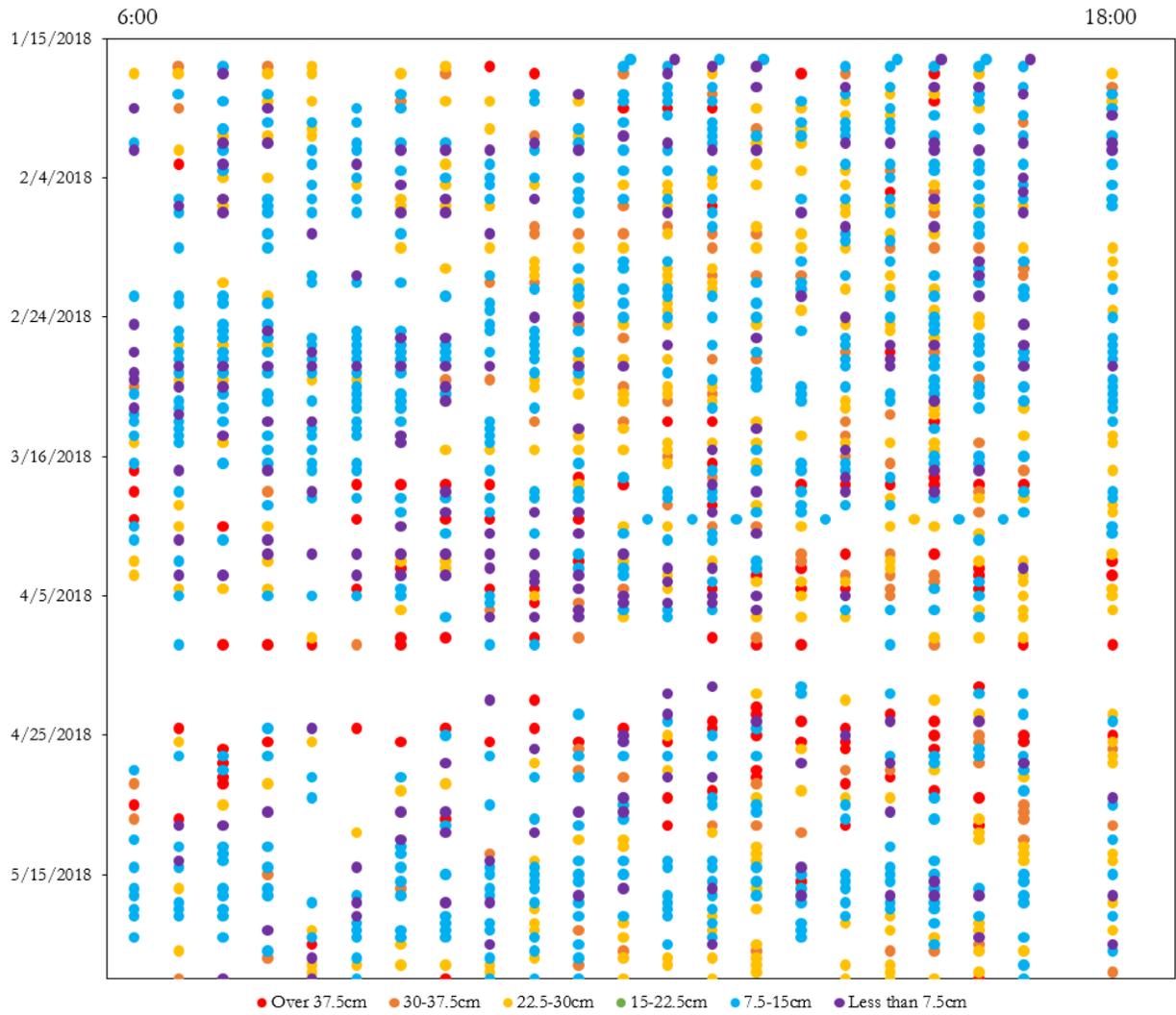


Figure A-6: Full time of day errors for Joe Wright

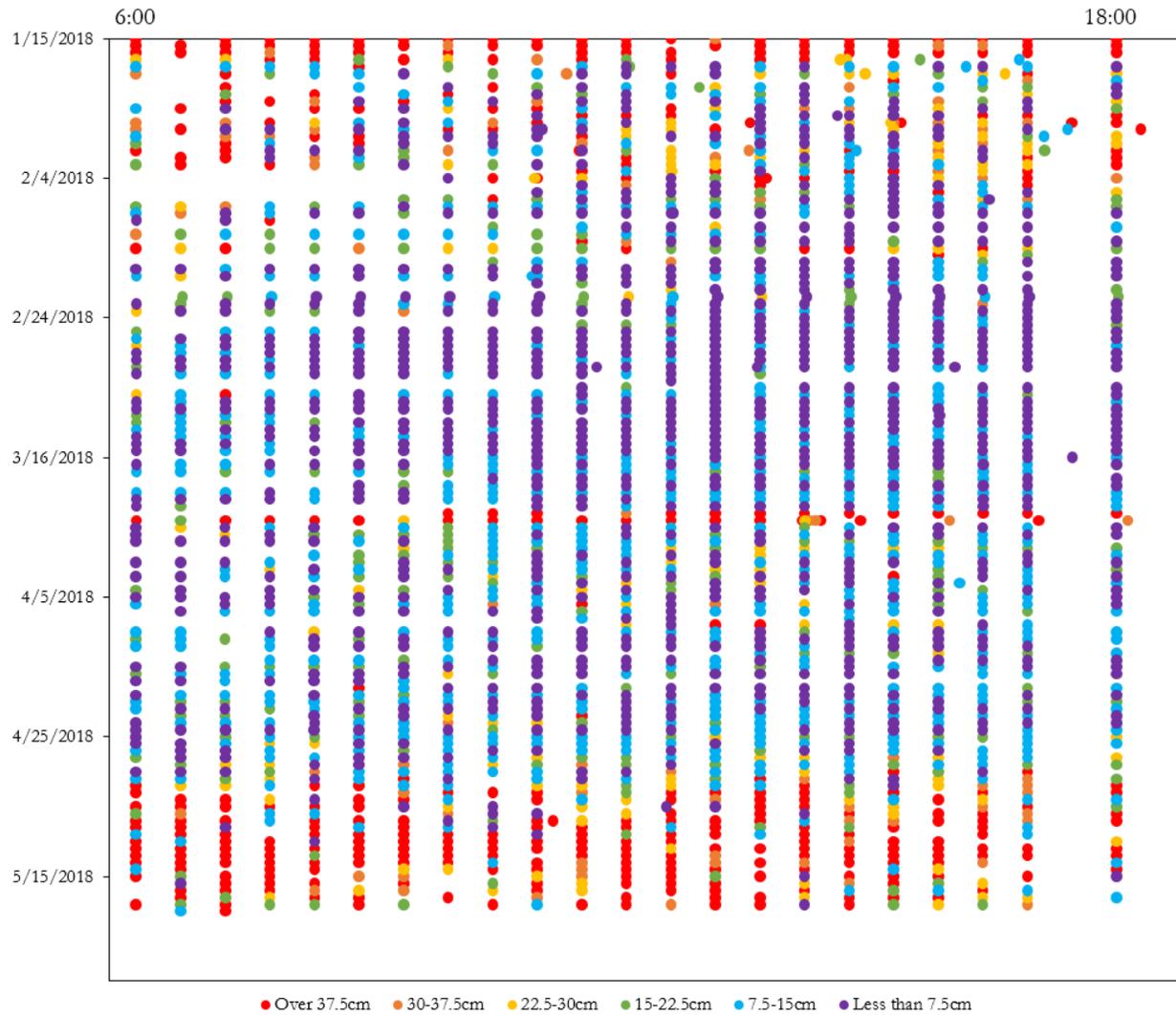


Figure A-7: Full time of day errors for Columbine

Appendix B: Site Configurations

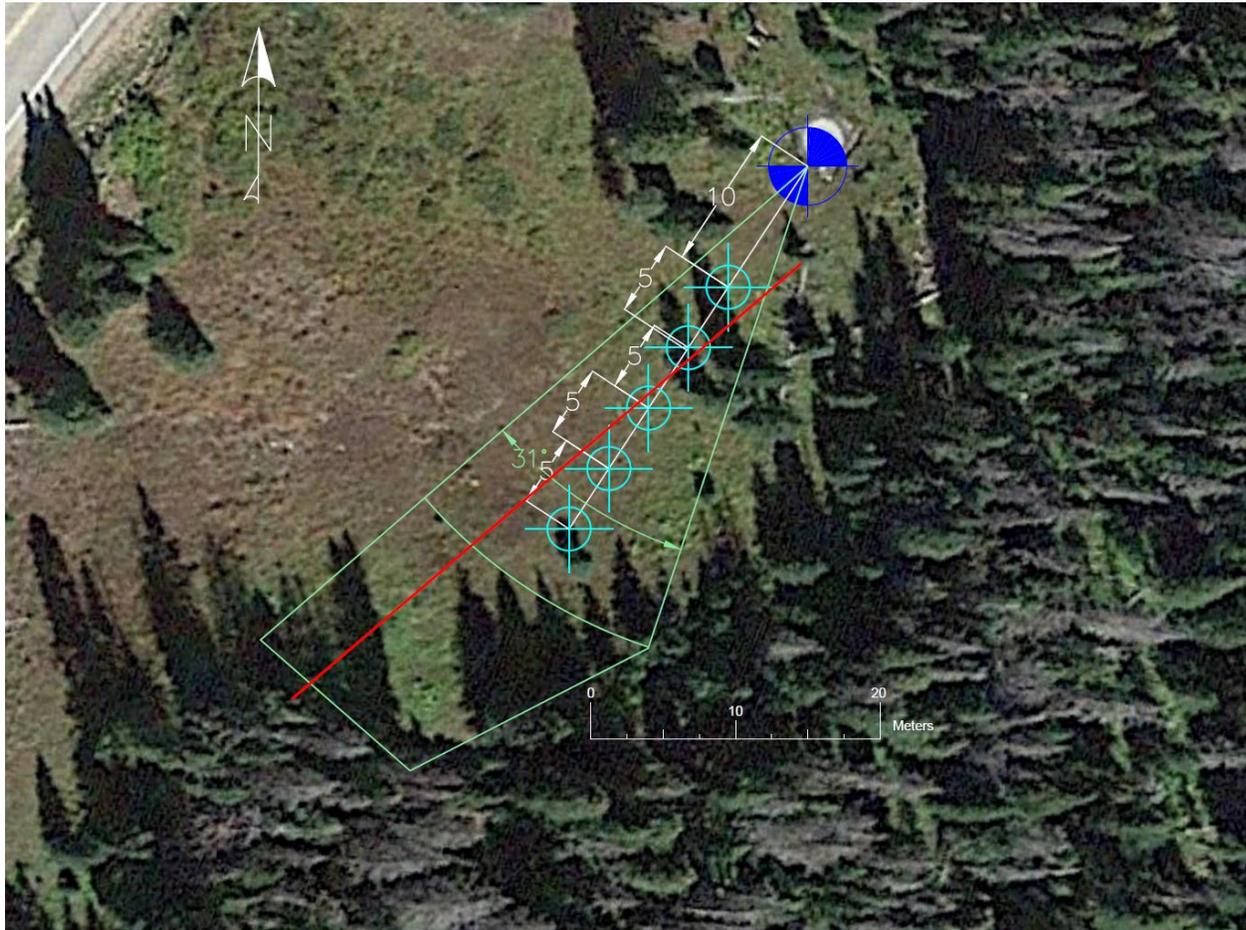


Figure B-1: Joe Wright site map, light blue symbols are staffs, dark blue are cameras, path of WY 2006 – WY 2008 transects shown in red

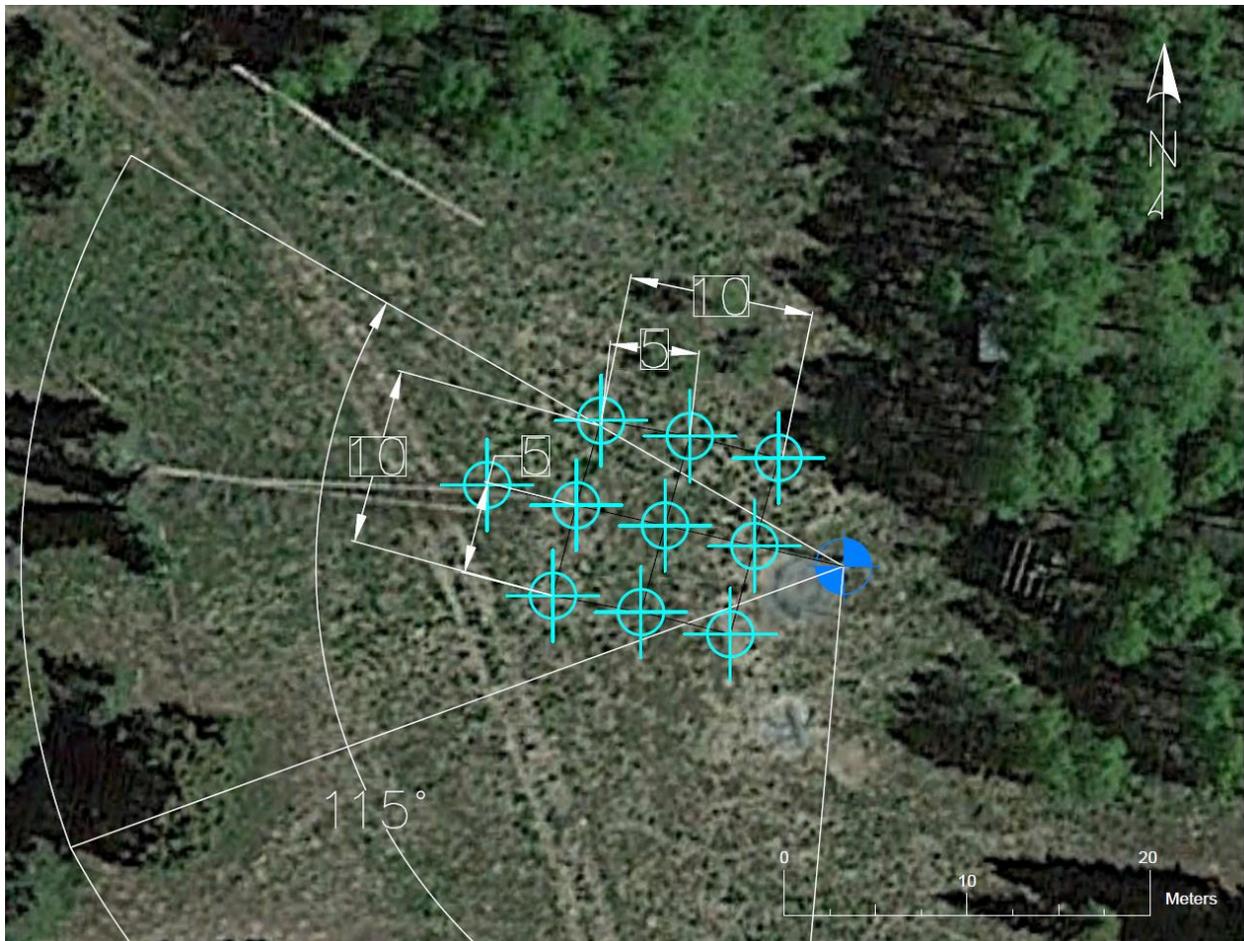


Figure B-2: Columbine site map, light blue symbols are staffs, dark blue are cameras



Figure B-3: Swamp Angel site map, light blue symbols are staffs, dark blue are cameras

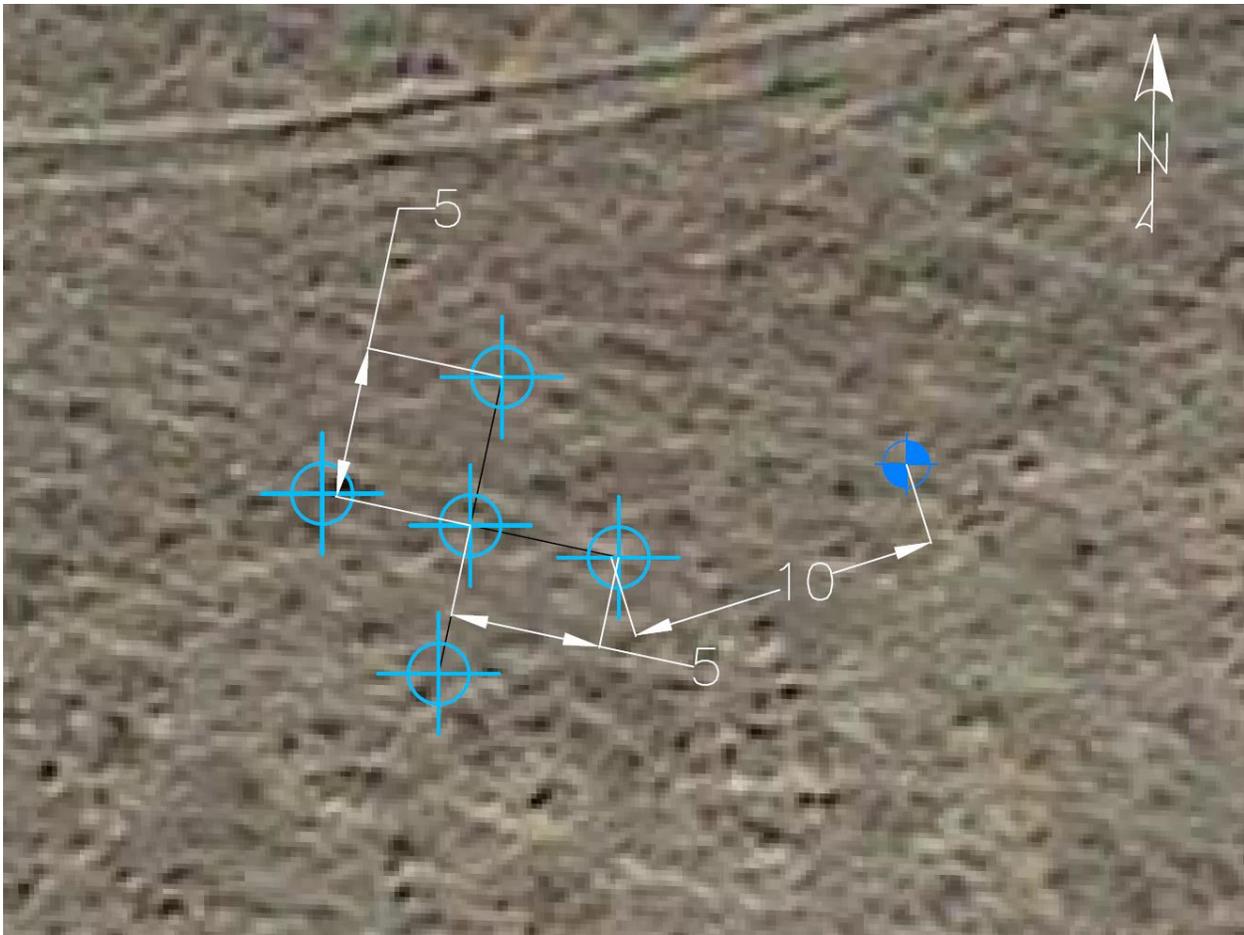


Figure B-4: East River site map, light blue symbols are staffs, dark blue are cameras



Figure B-5: Red Mountain pass site map, light blue symbols are staffs, dark blue are cameras

Appendix C: Full Sensitivity Analysis Results

Sensitivity analysis results not shown in the results section with results that can be well visualized.

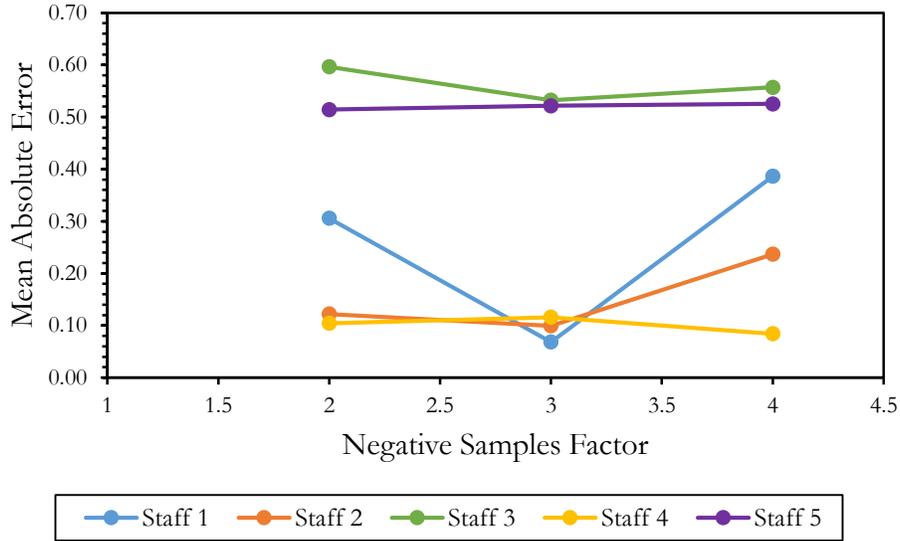


Figure C-1: How many times more negative samples are used in the calibration process than positive samples

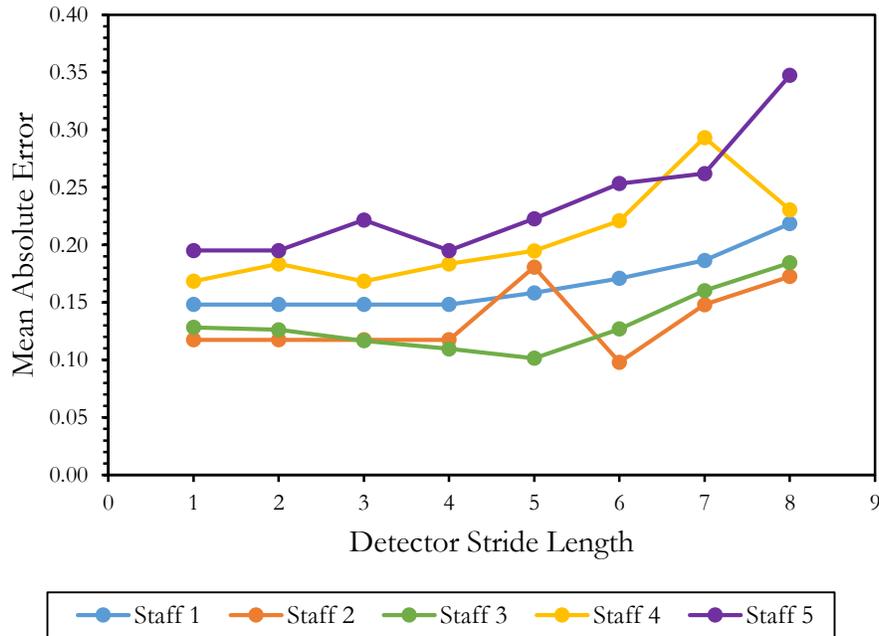


Figure C-2: Window Stride, how many pixels the searching window moves between analyzing groups of pixels

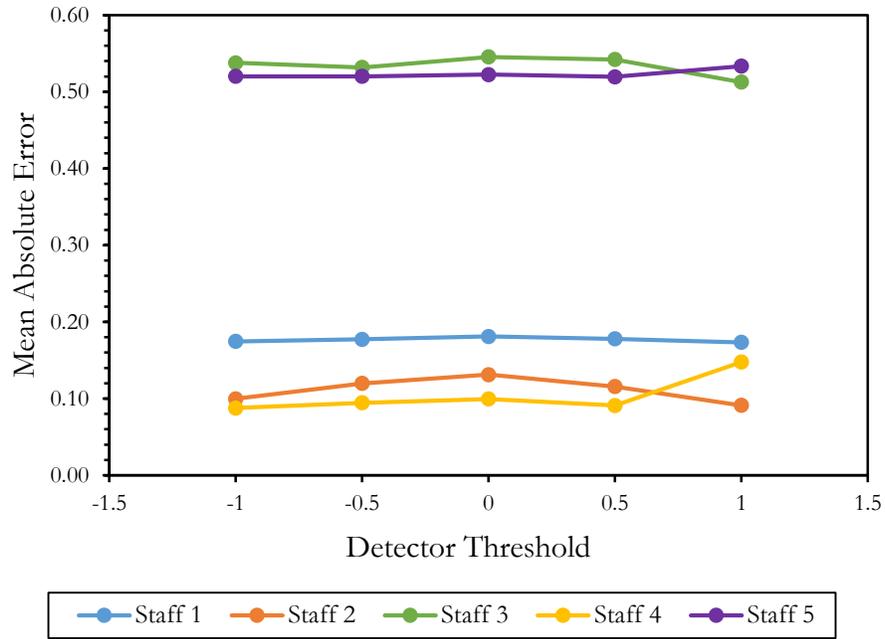


Figure C-3: Threshold, a parameter that controls how many positive returns are rejected within the detection protocol

Appendix D: SNOTEL Depth Data WY 2009 – WY 2018

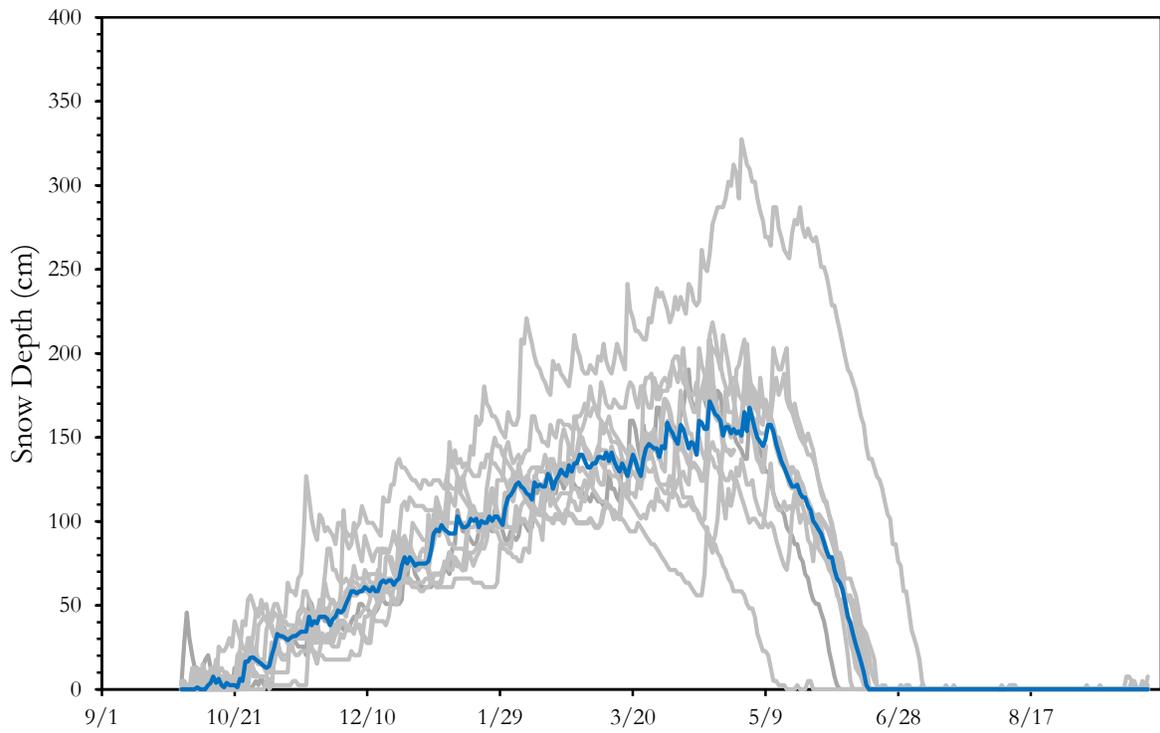


Figure D-1: Joe Wright SNOTEL depth data for WY 2009 – 2018 with the median snow depth for the period in blue

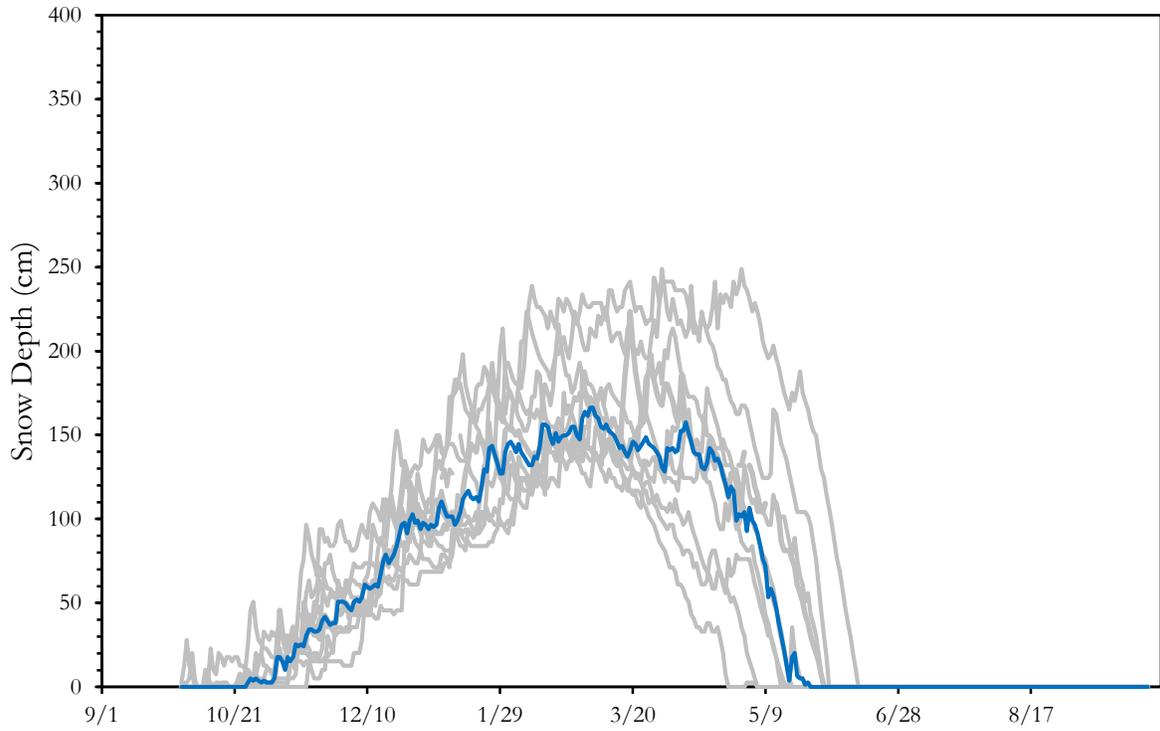


Figure D-2: Columbine SNOTEL depth data for WY 2009 – 2018 with the median snow depth for the period in blue

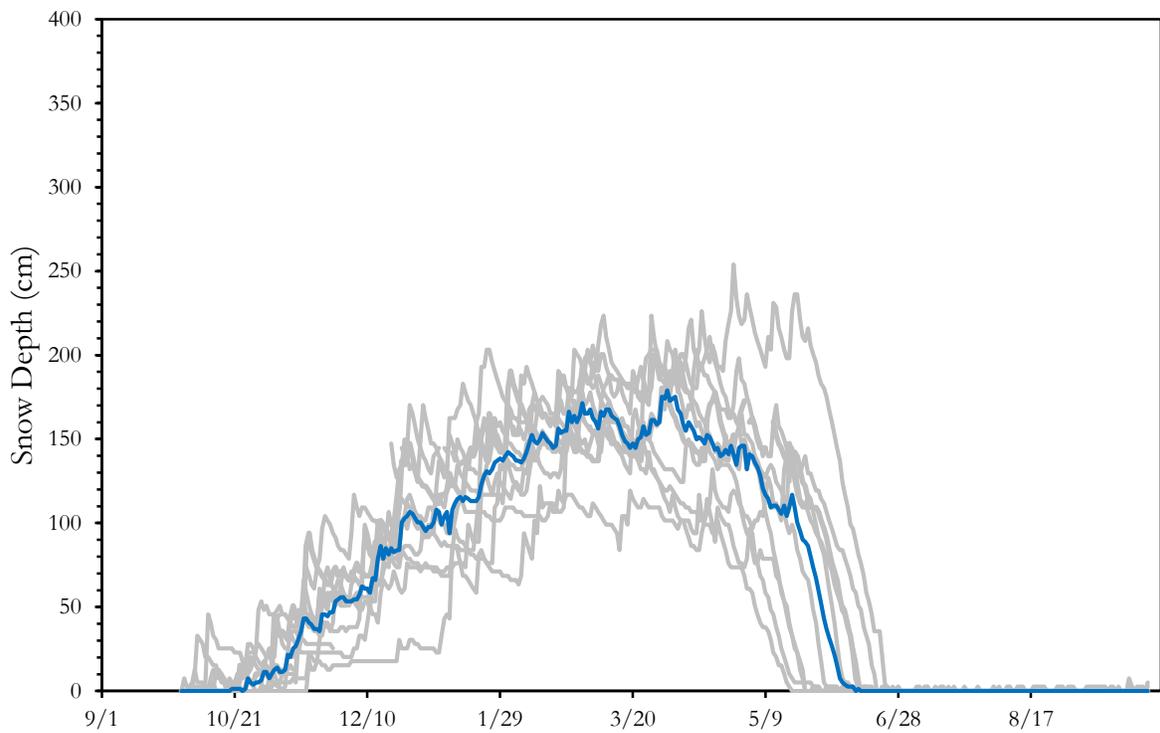


Figure D-3: Red Mountain Pass SNOTEL depth data for WY 2009 – 2018 with the median snow depth for the period in blue