

ASSESSING THE LIMITATIONS AND CAPABILITIES OF LIDAR
AND LANDSAT 8 TO ESTIMATE THE ABOVEGROUND VEGETATION
BIOMASS AND COVER IN A RANGELAND ECOSYSTEM
USING A MACHINE LEARNING ALGORITHM

by

Shital Dhakal

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Hydrologic Sciences

Boise State University

May 2016

© 2016

Shital Dhakal

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Shital Dhakal

Thesis Title: Assessing the Limitation and Capabilities of Lidar and Landsat 8
to Estimate the Aboveground Vegetation Biomass and Cover
in a Rangeland Ecosystem Using a Machine Learning Algorithm

Date of Final Oral Examination: 28 October 2015

The following individuals read and discussed the thesis submitted by student Shital Dhakal, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Nancy F. Glenn, Ph.D.	Chair, Supervisory Committee
Alejandro N. Flores, Ph.D.	Member, Supervisory Committee
Douglas J. Shinneman, Ph.D.	Member, Supervisory Committee
Aihua Li, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Nancy F. Glenn, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude and immeasurable appreciation to my supervisor, Dr. Nancy Glenn, for her thoughtful encouragement, careful mentorship, and most importantly her trust and confidence in my work. Without her infectious enthusiasm and unflappable determination, this thesis would not have taken shape.

I am also grateful to my advisory committee - Dr. Aihua Li, Dr. Alejandro N. Flores, and Dr. Douglas J. Shinneman - for their guidance and support throughout this research. Heartiest thanks to Dr. Li for being a friend of mine, providing me encouragement at every step without judging me. I should also thank Dr. Rupesh Shrestha and Lucas Spaete for being there whenever I was in need.

Special shout out goes to Abdolhamid Dashtiahangar, Ann Marie Raymondi, Erin Murray, Kyle Anderson, Nayani Thanuja Illangakoon, Peter Olsoy, and Sikha Neupane for supporting me in different phases of this research. I must have gained 3 credits worth of knowledge from each one of you. Thanks to Boise Center Aerospace Laboratory (BCAL) and Lab for Ecohydrology and Alternative Futuring (LEAF) for being a great support system.

Lots of love and appreciation to my parents, family, and friends for their moral support and undying faith in me. Special thanks from the bottom of my heart to Trishna Koirala, Sakar Panta, Prakash Subedi, Yudishthir Bhetwal, Ujjwal Acharya, and Anantarupa das for making sure I get enough distraction and don't become a workaholic.

I also appreciate everyone who contributed in making my two years stay in Boise a pleasant experience.

Finally, I thank the USGS (Joint Fire Sciences Program Award, *Quantifying and predicting fuels and the effects of reduction treatments along successional and invasion gradients in sagebrush habitats*) and NOAA OAR Earth Systems Research Laboratory/Physical Sciences Division (Award NA10OAR4680240) for funding this research.

ABSTRACT

Remote sensing based quantification of semiarid rangeland vegetation provides the large scale observations required for monitoring native plant distribution, estimating fuel loads, modeling climate and hydrological dynamics, and measuring carbon storage. Fine scale 3-dimensional vertical structural information from airborne lidar and improved signal to noise ratio and radiometric resolution of recent satellite imagery provide opportunities for refined measurements of vegetation structure.

In this study, we leverage a large number of time series Landsat 8 vegetation indices and lidar point cloud - based vegetation metrics with ground validation for scaling aboveground shrub and herb biomass and cover from small scale plot to large, regional scales in the Morley Nelson Snake River Birds of Prey National Conservation Area (NCA), Idaho. The Landsat vegetation indices were trained and linked to *in-situ* measurements ($n = 141$) with the random forest regression to impute vegetation biomass and cover across the NCA. We also validated our model with an independent dataset ($n = 44$), explaining up to 63% and 53% of variation in shrub cover and biomass, respectively. Forty six of the *in-situ* plots were used in a model to compare the performance of lidar and Landsat data in estimating vegetation characteristics. Our results demonstrate that Landsat performs better in estimating both herb ($R^2 \sim 0.60$) and shrub cover ($R^2 \sim 0.75$) whereas lidar performs better in estimating shrub and total biomass ($R^2 \sim 0.75$ and 0.68 , respectively). Using the lidar only model, we demonstrate that lidar metrics based on shrub height have a strong correlation with field-measured shrub biomass ($R^2 \sim 0.76$).

We also compare processing the lidar data with raster-based and point cloud-based approaches. The results are scale-dependent, with improved results of biomass estimation at coarser scales with point cloud processing. Overall, the results of this study indicate that Landsat and lidar can be efficiently utilized independently and together to estimate biomass and cover of vegetation in this semi-arid rangeland environment.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvi
1 CHAPTER ONE: INTRODUCTION AND BACKGROUND	1
1.1 Statement of Problem.....	1
1.2 Use of Remote Sensing Techniques	4
1.3 Use of Statistical Models	7
1.4 Conclusion	8
1.5 Thesis Organization	9
2 CHAPTER TWO: MANUSCRIPT ONE - AIRBORNE LIDAR BASED ESTIMATION AND SCALING OF SEMIARID BIOMASS USING RANDOM FOREST VARIABLE SELECTION.....	10
Abstract.....	10
2.1 Introduction.....	11
2.2 Study Area and Data	16
2.2.1 Study Area.....	16
2.2.2 Field Sampling	17
2.2.3 Airborne Lidar Data Acquisitions	20

2.2.4	Lidar Data Processing	20
2.3	Statistical Analysis.....	23
2.3.1	Regression Analysis	23
2.3.2	Random Forest (RF) Regression for Variable Selection.....	24
2.3.3	Nearest Neighbor (NN) Imputation.....	25
2.4	Results.....	26
2.4.1	Regression Analysis	26
2.4.2	Random Forest Raster Analysis	28
2.4.3	Random Forest Point Cloud Analysis	30
2.4.4	Imputation	32
2.5	Discussion	34
2.6	Conclusion	41
2.7	References.....	41
3	CHAPTER THREE: MANSUCRIPT TWO - IMPUTATION OF VEGETATION BIOMASS AND COVER FOR LARGE SCALE MAPPING IN A SEMI-ARID RANGELAND USING AIRBORNE LIDAR, LANDSAT 8 AND A MACHINE LEARNING ALGORITHM.....	55
	Abstract.....	55
3.1	Introduction.....	56
3.2	Study Area and Data	62
3.2.1	Study Site	62
3.2.2	Acquiring Landsat 8/OLI and Ancillary Data.....	63
3.2.3	Field Sampling	63
3.2.4	Airborne Lidar Data Acquisitions.....	67
3.3	Methodology.....	67

3.3.1	Spectral Vegetation Indices (VI).....	67
3.3.2	Lidar Data and Derivative Processing.....	69
3.3.3	Modeling Approach.....	71
3.4	Results.....	73
3.4.1	Calibration and Validation of OLI	73
3.4.2	Biomass and Cover Maps.....	75
3.4.3	Comparison of Landsat OLI Performance with Lidar	80
3.5	Discussion.....	81
3.5.1	Selection and Performance of Predictor Variables	81
3.5.2	Necessity of Data Fusion.....	84
3.5.3	Relation Between Cover and Biomass	85
3.6	Conclusion	87
3.7	References.....	89
4	CHAPTER FOUR: CONCLUSIONS.....	105
5	REFERENCES	108
6	APPENDIX.....	118
6.1	An Application: Effect of Wildfire on Biomass And Cover.....	118
6.1.1	Fire Data.....	118
6.1.2	Relation with Fire Frequency	118
6.1.3	Implication of Fire Events.....	121
6.2	Stepwise Regression Between Biomass and Lidar Metrics	123
6.3	R Code for Imputation	127
6.4	MATLAB Code for Calculating Vegetation Metrics	129

6.5 Abbreviation of the Vegetative Indices Used 135

LIST OF TABLES

Table 1.1	Statistics of vegetation cover and biomass from the field sites. The results are from $n=414$ subplots (1 m^2) nested inside $n=46$ plots.	18
Table 1.2	Lidar metrics and their descriptions used in the analysis.	22
Table 1.3	Results of random forest regression using raster data processing for total and shrub biomass at different resolutions. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.	29
Table 1.4	Results of random forest regression for total biomass and shrub biomass at 70m x 70m plot level. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.	30
Table 1.5	Results of random forest regression using point cloud processing for total biomass and shrub biomass. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.	31
Table 1.6	Results of random forest regression for herbaceous biomass. Results for raster processing and point cloud processing are also shown. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.	32
Table 3.1	Summary of vegetation cover and biomass based on 1 ha field data ($n=141$).	65
Table 3.2	Overview of the spectral indices from Landsat 8 OLI used in this study.	68
Table 3.3	Lidar metrics and their descriptions used in the analysis.	70
Table 3.4	Random forest results for calibration and validation model of OLI. The RMSE values of cover predictions are in units of % and biomass	

	predictions are in g/m^2 . The variables are labeled by respective date and metric from Table 2. 2.	74
Table 3.5	Comparison of RF model ($n=46$) performance with Landsat 8 OLI and lidar. RMSE is in the units of g / m^2 for biomass predictions and % for cover predictions. The OLI metrics are labeled by date and metric from Table 2. 2. The lidar metric are labeled from Table 2. 3.	80
Table A.1	Abbreviation of the VI used.....	135

LIST OF FIGURES

Figure 1.1.	The Morley Nelson Snake River Birds of Prey National Conservation Area (NCA), located in southwest Idaho, United States. This study took place in the northwestern portion of the NCA.....	17
Figure 1.2	Frequency distribution of the in-situ shrub and herbaceous biomass (n=46) with bins of 100 g/ m ²	19
Figure 1.3	Study area within the NCA and (a) lidar coverage and distribution of sampled field plots over the study site and (b) schematic of the field sampling procedure.....	20
Figure 1.4	Comparison between observed total AGB and total AGB predicted from regression analysis.	27
Figure 1.5	Comparison between observed shrub biomass and shrub biomass predicted from regression analysis.....	28
Figure 1.6	Imputed total AGB map at 1 m resolution of a selected area indicated by the brown frame.	33
Figure 1.7	Comparison of imputed shrub and herbaceous biomass maps for a selected area indicated by red star. Rectangular box refers to area in Fig. 1. 6.	34
Figure 1.8	Simple linear regression between total AGB and Standard Deviation of Height	35
Figure 1.9	Relationship between pixel size and R ² and RMSE in raster data processing. As the pixel resolution decreases, the R ² decreases and RMSE increases. The figure represents the results from total AGB only but shrub biomass follows a similar trend.....	38
Figure 1.10	Relation between pixel size and R ² and RMSE in point cloud processing. Comparatively, the cell size has little effect on the R ² and RMSE. The figure represents the results from total AGB only but shrub biomass follows a similar trend.	38
Figure 1.11	Schematic showing the difference in extraction of vegetation characteristics due to different raster pixel size. The red cell represents	

	the field plot of interest. The black cells in the left figure show coarser resolution pixels and the black cells in the right represent finer resolution pixels.	39
Figure 3.1	Location of field plots and lidar coverage in the study site.	62
Figure 3.2	Frequency distribution of the in-situ (a) biomass and (b) percent cover of herb and shrub (n=144).....	66
Figure 3.3	Summary distribution of (a) biomass and (b) percent cover in the NCA after imputation.	76
Figure 3.4	Imputed map at Landsat resolution of a) herb biomass, b) shrub biomass, c) total biomass, d) herb cover and e) shrub cover.	79
Figure 3.5	Relationship between cover and biomass in shrub and herb from (a) in-situ measurements and (b) imputation for all n=144 plots. There is a significant relationship between shrub biomass and cover in the field measurements. Imputed estimation showed that the correlation between shrub biomass and cover decreased while correlation between herb biomass and cover increased relative to field measurements.....	87
Figure A.1	Graph showing change in a) average biomass and b) average percent cover with respect to fire frequency in the study site. The biomass and cover value are computed from the respective imputed map.....	119
Figure A.2	Frequency distribution of plots with respect to fire events.....	120
Figure A.3	Graph showing change in a) field average biomass and b) field average percent cover with respect to fire frequency. None of the field plots burned six times.	121
Figure A.4	Comparison between observed AGB and AGB predicted from stepwise regression.	124
Figure A.5	Comparison between observed shrub biomass and shrub biomass predicted from stepwise regression.....	125
Figure A.6	Residual analysis of stepwise regression for AGB.....	125
Figure A.7	Residual analysis of stepwise regression for shrub biomass	126

LIST OF ABBREVIATIONS

ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AGB	Above Ground Biomass
ALS	Airborne Laser Scanning
CART	Classification And Regression Tree
FHD	Foliage Height Diversity
LIDAR	Light Detection And Ranging
GDEM	Global Digital Elevation Model
GNSS	Global Navigation Satellite System
METI	Ministry of Economy, Trade and Industry (Japan)
OLI	Operational Land Imager
OOB	Out of Bag (Error)
RF	Random forest

1 CHAPTER ONE: INTRODUCTION AND BACKGROUND

1.1 Statement of Problem

The sagebrush-steppe represents one of the largest and most imperiled ecosystems in the North American continent (Miller et al. 2011; Barbour and Billings 2000). Arid and semi-arid ecosystems cover approximately one-third of the Earth's land surface and millions of square kilometers in the American Intermountain West (Schlesinger et al. 1990). However, the combination of increased fire frequency in lower elevation rangelands due to the spread of invasive exotic species like cheatgrass (*Bromus tectorum*) and medusahead (*Taeniatherum caputmedusae*) and juniper encroachment in higher elevation rangelands has resulted in a dramatic decrease in sagebrush (*Artemisia tridentata*) presence (Anderson and Inouye 2001; Knick 1999; Miller et al. 2011). Also, human activity such as urban and increased agricultural development, off-road vehicle activity and poorly managed livestock grazing have caused a large decrease in sagebrush-dominated rangelands. Only <10% of historic sagebrush-steppe ecosystem is estimated to be unaltered by human activities in the United States (West 1999).

Native shrubs like sagebrush are one of the most important plants on western rangelands from an ecological point-of-view. They are home to and provide food for imperiled animals like greater sage grouse (*Centrocercus urophasianus*) and pygmy rabbits (*Brachylagus idahoensis*) (Storch 2007; Shipley et al. 2006) . Sagebrush provides habitat for nearly 100 species of birds, hosts of invertebrates, reptiles and small mammals (Connelly et al. 2000). Thermal and security cover is also provided by sagebrush for

wildlife like pheasants, chukar, sharp tailed grouse and sage grouse (Eberhardt et al. 1984). A change in the sagebrush distribution will likely cause a decline in the population of these species, some of which are already highly imperiled.

Distribution of sagebrush across rangelands is also important for continuation of the hydrologic and carbon cycle (Ursino 2007; Pierson et al. 2003). Mature and big sagebrush plants have a two-part root system, a deep tap root and a shallow diffuse root system. The tap root system brings deep soil moisture and nutrients to the soil surface by “hydraulic lift” which is available for roots of other understory plants (Cardon et al. 2013). Disturbance such as overgrazing, fire or invasion of non-native plants results in the decrease of aboveground biomass which ultimately mirrors in a decrease in root biomass. The decrease in sagebrush biomass and cover also results in less organic matter which is an important factor in aggregate formation and stability (Heitschmidt and Stuth 1991). This decreases water holding capacity of soil and infiltration and increases the surface flow. Sagebrush also plays a crucial role in the hydrological cycle of water-limited regions (Wilcox 2010). Evapotranspiration is a major component of soil water content in rangelands and about 96% of incoming precipitation has been shown to be returned to the atmosphere by vegetation such as sagebrush (Branson et al. 1976). Several studies (Angell et al. 2001; Shrestha and Stahl 2008) have also shown the critical role that sagebrush plays in the terrestrial carbon storage.

Quantification of vegetation characteristics in rangelands is essential for the management, conservation and restoration of native plant communities. Studies have shown that aboveground biomass and percent cover of vegetation in rangelands are important for modeling vegetation dynamics, estimating pre-fire and post-fire fuel loads,

measuring carbon storage, assessing habitat quality and managing changes in native species (Polley et al. 2007; Storch 2007; Rengsirikul et al. 2011; Angell et al. 2001; Shrestha and Stahl 2008). Biomass and cover are also strong indicators of ecosystem structure and productivity that informs a range of applications such as assessing forage potential, species dominance, and wildlife habitat conditions (Polley et al. 2007; Rengsirikul et al. 2011; Chen et al. 2012). Vegetation structure is also important for the study of functional plant biology and growth analysis, and the calculation of net primary production and growth rates (Golzarian et al. 2011).

Various methods are available for in-situ measurement of biomass (Sala and Lauenroth 1982; Clark et al. 2008; Bonham 2013) but almost all are either destructive, labor intensive, expensive or only practical for small areas. Common methods include harvesting (Sims and Singh 1978, Sala and Lauenroth 1982), clip-and-weigh (Bonham 2013), visual estimation (Waite 1994), and point-intercept sampling (Clark et al. 2008). The destructive biomass measurement method is considered most accurate because it involves clipping, oven-drying, and weighing the plant material; however this method is also expensive and labor intensive (Bonham 2013). Point-intercept sampling is a non-destructive method that is relatively inexpensive, but it involves taking multiple individual field measurements for each species of interest to statistically represent its biomass (Clark et al. 2008). Percent vegetation cover is generally measured by visual estimation, line intercept sampling or through photographic methods, but these methods can be prone to observer bias and may produce unreliable results (Luscier et al. 2006). Moreover, while these field-based methods may be accurate and effective at a plot level, they are not easily scaled to the larger landscape.

1.2 Use of Remote Sensing Techniques

New advancements in remote sensing technology have brought promising results to quantify not only biomass and cover but other vegetation characteristics of rangelands (e.g., shrub height, leaf area index, fuel load) (Hudak et al. 2009). Remote sensing technologies offer potential solutions for extending biomass collected *in-situ* to a range of spatial scales in a cost-efficient manner (Mitchell et al. 2011). Passive spectral remote sensing techniques employ visible and near infrared (NIR) regions from aerial and satellite optical sensors for assessing spectral characteristics of vegetation.

Photosynthetically active vegetation typically reflects in the green and NIR, absorbs in the red and blue wavelengths and exhibits strong absorption properties in wavelengths where atmospheric water is present. Spectral remote sensing takes these variations into account and measures vegetation characteristics using a combination of spectral bands called Vegetation Indices (VI). These VI, including the Normalized Difference Vegetation Index (NDVI), Ratio Vegetation Index (RVI), Green Vegetation Index (GVI), Perpendicular Vegetation Index (PVI), Soil Adjusted Vegetation Index (SAVI), Transformed Soil Adjusted Vegetation Index (TSAVI), are known to provide meaningful information linked to plant cover, health, water content, environmental stress, and other characteristics (Zandler et al. 2015; Basso et al. 2004; Richardson and Everitt 1992). Multi-temporal monitoring of large scale landscape change of vegetation is possible using spectral remote sensing because of the availability of imagery archives dating back several decades. It is challenging, however, to quantify vegetation in semiarid environments because the vegetation is typically low in stature, spectrally indeterminate, and generally has low density or plant surface area (Mitchell and Glenn 2009). This

causes the signals coming from vegetation to be mixed with those coming from bare soil. Landsat 8's Operational Land Imager (OLI) has a pushbroom configuration generating 16-bit images with at least an eight fold increase in signal-to-noise ratio than previous Landsat missions, which may have the potential to improve detection of vegetation structure in high soil mixing environments (Roy et al. 2014).

Light Detection and Ranging, or lidar, can also provide practical, economical and reliable estimates of biomass and cover. Lidar is an active remote sensing technology with a proven ability to map aboveground biomass in forested ecosystems (He et al. 2013, Chen et al. 2012). Its capability of separating vegetation from ground is unique in comparison to optical remote sensing instruments (Campbell and Wynne 2011). Lidar is an active form of remote sensing that uses a series of short pulses of a narrow beam of coherent light, typically the infrared wavelength, from its sensor. The laser hits an object or surface at an acute angle and a portion of energy is returned to the sensor. The time delay and angle of the backscattered energy is used to measure the distance between the sensor and the reflecting surface. Multiecho sensors can detect several returns from a single pulse including first and last returns, which typically represent near the top of the canopy and the underlying ground surface, respectively (Jones and Vaughan 2010). These assembled returns, called the lidar point cloud, can be sorted, filtered and processed to provide information for ground and non-ground targets. In Airborne Laser Scanning (ALS, hereafter lidar), pulses of light (produced by a laser) are emitted from an instrument mounted in an aircraft and directed to the ground in a scanning pattern (Farid et al. 2008).

Lidar technology has proven useful due to its ability to obtain range and orientation information by capturing three-dimensional (3-D) data (Su and Bork 2007). The 3-D characteristic makes lidar a powerful tool to study vegetation characteristics. For more than a decade, lidar has been successfully used to measure forest volume, height and biomass (Lefsky et al. 2002; Zimble et al. 2003; Andersen et al. 2005; Hall et al. 2005). Remote sensing vegetation characteristics of shrubs in rangelands using lidar has also been of interest (Ritchie et al. 2006; Streutker and Glenn 2006; Su and Bork 2007; Glenn et al. 2011). However, the application of lidar in shrublands is marred by the low vegetation height and sparsely distributed vegetation across a uniform surface (Estornell et al. 2012). Many shrub-steppe communities have irregular morphology and are less than 2 m tall. Hence laser pulses that hit within shrub canopies can be misclassified as ground rather than canopy (Riaño et al. 2007). This causes accuracy problems for defining individual elements and estimating vegetation heights.

The terrestrial form of lidar is Terrestrial Laser Scanning (TLS), which operates at close range and has a higher point density in comparison to airborne or satellite based lidar. Although useful for measuring small areas, TLS point clouds can suffer from occlusion effects, which occur when laser beams are reflected from foreground objects such as stems and canopies, and objects behind are completely or partially missed . Although airborne lidar point density is lower (usually ~ 10 points / m^2) than TLS density (≥ 500 points/ m^2), airborne lidar is not as limited as TLS to small geographic areas (typically few thousand square meters) (Vierling et al. 2012). Thus, because the geographic coverage of lidar data is typically much larger than TLS, it is a complementary tool to TLS to scale vegetation characteristics from a small plot to a

regional level. Another benefit of airborne lidar is that unlike spectral remote sensing, the spectral signal mixing effect from soil is minimized (although the lidar beam may hit portions of branches). Despite the previous research and applications of lidar in rangeland ecosystems (e.g. Vierling et al. 2012; Li et al. 2015), using lidar to estimate biomass has been largely unexplored.

1.3 Use of Statistical Models

There is a strong link between plant height and other biophysical characteristics including cover, biomass, density and canopy volume (Dubayah et al. 2000). Three dimensional lidar point clouds can be used to model plant height to exploit these relationships (Bork and Su 2007). Studies have also demonstrated empirical correlation of spectral indices from Landsat 8 with vegetation attributes (Li et al. 2013; Ding et al. 2014). These remote sensing variables can be correlated with the biophysical vegetation characteristics in the field using a number of statistical methods such as regression analysis (Laurin et al. 2014; García-Gutiérrez et al. 2011), Hierarchical Bayesian modeling (Wilson et al. 2011), random forest (Hudak et al. 2008) and Artificial Neural Networks (Debouk et al. 2013). It is not uncommon to see a large number of predictor variables with a relatively modest number of ground-truth observations in remote sensing applications. This can potentially lead to over-fitting of models via high-dimensional data problems (Zandler et al. 2015). Regression analysis is widely adopted to relate field and remotely sensed data. However, in regression identifying suitable variables for meaningful correlation is critical as some variables are weakly related to the ground data and strongly related to each other (Mutanga et al. 2012; Montgomery et al. 2012; Fernandes and Leblanc 2005; Ahmed et al. 2015). Furthermore, regression analysis

involves assumptions of normality and homogeneity, which might not be satisfied with remote sensing datasets (Montgomery et al. 2012).

Recently, random forest (RF) has gained considerable attention in the field of remote sensing due to classification and computational accuracy, and the capability to provide a measure of variable importance (Mitchell et al. 2013; Guan et al. 2012; Pal 2005). This thesis study has a large number of predictors with a comparatively small sample size. Random forest is preferred over other methods in this study as the RF models result in smaller prediction variance and bias and better model performance (Mitchell et al. 2015). Random forest is a machine learning algorithm that uses a tree-based classifier technique developed by (Breiman 2001) that addresses the limitation of classification and regression trees (CART) by using a large number of decision trees. Random forest is an iterative classification tree statistical approach where bootstrap samples are drawn to construct multiple ‘trees’; each grown with a randomized subset of predictors (Breiman 2001). These ‘trees’ cast a unit vote for the most popular class to classify an input vector (Pal 2005; Breiman 1999). Thus, with RF we can estimate the best ‘predictors’ among the variables developed from a lidar point cloud and Landsat spectral data to quantify biomass and cover in a rangeland.

1.4 Conclusion

This study assesses the capability of Landsat 8 and lidar to quantify biomass and cover of native and non-native shrubs and herbs at a large-landscape scale. A reliable and cost effective method to estimate biomass and cover would provide important baseline vegetation data for monitoring of and managing the sagebrush-steppe. This baseline data could potentially be used to guide restoration of native rangeland ecosystems, evaluate

wildlife habitat, classify livestock grazing resources, and identify areas for applying fuels reduction and fire management strategies. This study is intended to develop methods to scale vegetation biomass and cover from fine to coarse scales using lidar and spectral data, and to assess the accuracy of these methods. Subsequent chapters will describe the process in detail and illustrate results.

1.5 Thesis Organization

This thesis consists of an introductory overview in Chapter 1 and two separate, independent and self-contained research manuscripts. Chapter 2 contains the first manuscript which builds on scaling biomass of rangeland vegetation from the plot to larger scales covered by lidar data available to the study. Lidar metrics are used as a proxy to estimate biomass and random forest is used to develop the model. Chapter 3 contains the second manuscript in which vegetation characteristics are scaled to a regional level. Landsat 8 is used to develop metrics that can be used in a random forest model to estimate biomass and cover. We also compare the performance of Landsat 8 to that of lidar, in order to better understand the capabilities and limitations of both of these remote sensors. Chapter 4 contains concluding remarks on the significance and potential of the work. The thesis ends with an appendix of raw reference data from the analysis of this research that can be useful for future studies and research.

2 CHAPTER TWO: MANUSCRIPT ONE - AIRBORNE LIDAR BASED
ESTIMATION AND SCALING OF SEMIARID BIOMASS USING RANDOM
FOREST VARIABLE SELECTION

Abstract

Quantifying aboveground total biomass of the sagebrush-steppe ecosystem can provide valuable information for a host of applications, including modeling climate and hydrological dynamics, estimating fuel loads, measuring carbon storage, assessing habitat quality and managing environmental changes. Various forms of remote sensing data have been used for biomass estimation modeling, typically across medium to broad landscape scales. In this study, we used airborne Light Detection and Ranging (lidar) data to estimate aboveground biomass in a sagebrush-steppe landscape that has largely been degraded by past land use and invaded by non-native annual grasses and forbs. We incorporated vegetation vertical structure information obtained from lidar data with ground-measured-validation data, allowing us to scale shrub and grass biomass measurements obtained from small field sites (1-100m plots) to a larger landscape scale (~ 75,164 hectares). Various vegetation metrics and statistics derived from the airborne lidar were trained and linked with the aboveground biomass data measured in the field using random forest (RF) regression. Our results demonstrated that lidar-derived metrics based on vegetation height had a strong correlation with the field-measured biomass ($R^2 \sim 0.74$). These developed relationships were then used to scale biomass estimates to the larger study area using imputation techniques. We also compared raster processing

techniques with point cloud processing and demonstrated that point cloud processing of lidar data significantly improved estimation of biomass at coarser scales. The RF method used in the study was well suited to determine the most important metrics for estimating biomass in rangelands with sparse vegetation cover.

2.1 Introduction

Aboveground biomass ('AGB' or 'biomass' hereafter) is a strong indicator of ecosystem structure and productivity that informs a range of applications such as forage potential, species dominance and wildlife habitat analysis (Polley et al. 2007; Rengsirikul et al. 2011; Chen et al. 2012). Although the AGB per unit area is low, dry rangelands cover one fifth of the earth's land area and thus play a significant role as a carbon sink and provider of essential ecosystem services (Perez-Quezada et al. 2011; Zandler et al. 2015). Accurate estimation of AGB in rangeland ecosystems is important for modeling vegetation dynamics, estimating fuel loads, measuring carbon storage, assessing habitat quality and monitoring changes in native species (Storch 2007; Rengsirikul et al. 2011; Angell et al. 2001; Shrestha and Stahl 2008). Hence, estimation of AGB can be used by resource managers to develop effective monitoring, conservation and restoration strategies in rangelands and to ensure their sustainability (Guo 2007; Brown and Archer 1999; Pieper 1988).

The sagebrush-steppe once extended across hundreds of millions of hectares in western US rangelands and is now one of the most imperiled ecosystems in the continent (Barbour and Billings 2000; Miller et al. 2011). Several sagebrush species (e.g. *Artemisia tridentate*, *Artemisia arbuscula*) and short bunchgrass species (e.g. *Festuca idahoensis*) are common natives of the sagebrush-steppe ecosystem. Factors such as invasion of

nonnative species, wildfire, overgrazing, urbanization and climate change are responsible for the degradation of these rangelands (Hemstrom et al. 2002; Miller and Rose 2006). In particular, increased fire frequency due to the spread of invasive nonnative species, such as cheatgrass (*Bromus tectorum*), medusahead (*Taeniatherum caputmedusae*) and encroachment of juniper (*Juniperus* spp.) has resulted in a dramatic decrease in sagebrush presence (Knick 1999; Wisdom et al. 2005; Miller et al. 2011). These changes resulted in a decline of already imperiled animals like greater sage grouse (*Centrocercus urophasianus*) and pygmy rabbits (*Brachylagus idahoensis*) which depend on sagebrush for habitat, food and shelter (Storch 2007; Shipley et al. 2006; Connelly et al. 2000). Degradation of sagebrush also affects the hydrologic cycle in dry lands due to less organic matter (Heitschmidt and Stuth 1991), decreased evapotranspiration (Branson et al. 1976; Wilcox 2010) and absence of “hydraulic lift” (Cardon et al. 2013).

Various direct and indirect methods are available for in-situ measurements of AGB of sagebrush and other shrubs (Sala and Lauenroth 1982; Clark et al. 2008; Bonham 2013). Some of the most common methods include harvesting (Sims et al. 1978, Sala and Lauenroth 1982), clip-and-weigh (Bonham 2013), visual estimation (Waite 1994), and point-intercept sampling (Clark et. al 2008). A destructive biomass measurement is considered most accurate because it involves clipping, oven-drying, and weighing the plant material; however this method is also expensive and labor intensive (Bonham 2013). Point-intercept sampling is a non-destructive alternative method that is relatively inexpensive, but it involves taking multiple individual measurements for each species of interest to statistically represent its biomass (Canfield 1941; Clark et al. 2008). These field-based methods are accurate and effective, but are not easily scaled across the

landscape and are time consuming. Moreover, these field methods are unlikely to be affordable to conduct and are unlikely to capture important changes in rangelands that are occurring at a swift pace across large areas (Vierling et al. 2012; Bonham 2013). Hence, there is a need for new techniques to estimate AGB accurately over a large area while being repeatable, automated and cost effective.

Remote sensing has often been cited as a tool for facilitating advances in understanding rangeland ecosystems and facilitating their management (Tueller 1992). Retrieving AGB from remote sensing data is advantageous compared to the labor and time intensive field methods because it can provide multi-scale contiguous estimates, ideally suited for modeling over broad scales and time (Li et al. 2015). In particular, Light Detection and Ranging (lidar) holds promise of effectively studying vegetation characteristics due to its ability to obtain range and orientation information by capturing three-dimensional data (Su and Bork 2007). Optical remote sensing also has been extensively used to estimate AGB but the use of vegetation indices and/or leaf area index (LAI) in rangelands are strongly affected by soil background, as well as small stature and sparse arrangements of shrubs creating a mixed pixel effect (Glenn et al. 2005; Chopping et al. 2008; Zandler et al. 2015). This is a major limitation for optical remote sensing, but a minor limitation in lidar. Thus, lidar has a significant advantage over other remote sensing based biomass estimation techniques (Campbell and Wynne 2011).

For more than a decade, lidar has been successfully used in forest applications such as measuring volume, height and AGB (Lefsky et al. 2002; Hall et al. 2005; Ku et al. 2012; Lin et al. 2012; Zheng et al. 2013). Lidar also has been increasingly used to study vegetation characteristics of shrubs (e.g., shrub height, canopy cover, leaf area index) in

rangelands (Ritchie et al. 2006; Streutker and Glenn 2006; Su and Bork 2007; Glenn et al. 2011; Hudak et al. 2009). There is a strong link between shrub height and other biophysical characteristics (e.g. cover, AGB, canopy volume) (Bork and Su 2007) emphasizing the importance of metrics developed from three dimensional point clouds. However, limited lidar research has been focused on estimating and scaling AGB in semiarid rangeland ecosystems. Many species and vegetation in shrub-steppe have irregular morphology, low stature (less than 2m tall), low density and small canopy surface area. These characteristics increase the probability of the laser pulse to be misclassified as ground rather than canopy (Riaño et al. 2007; Mitchell and Glenn 2009). This problem can be addressed to a certain extent by supplementary ground measurements, increases in lidar point density and attention to canopy penetration (Glenn et al. 2011). Hence, more research and exploration will allow novel approaches to successfully quantify AGB from 3D point clouds in dry rangelands.

Metrics derived from lidar (e.g. mean height, variance of height, canopy relief ratio etc.) can be correlated with biophysical vegetation characteristics in the field using statistical methods such as regression analysis (Laurin et al. 2014; García-Gutiérrez et al. 2011), Hierarchical Bayesian (Wilson et al. 2011), random forest (Hudak et al. 2008a), and Artificial Neural Networks (Debouk et al. 2013). Traditionally, regression analysis has been widely adopted for predicting AGB (Baskerville 1972; García-Gutiérrez et al. 2011; Hudak et al. 2006; Zolkos et al. 2013) and more generally for use with predicting modeling to relate field and remotely sensed data (Lefsky et al. 2002; Fernandes and Leblanc 2005; Berterretche et al. 2005; Haack and Rafter 2010). Regression analysis provides the unbiased minimum squared error estimate of response variables using a

linear combination of regressors, but commonly involves a set of assumptions such as non-multicollinearity, normality, homogeneity and independence of residuals (Montgomery et al. 2012; Fernandes and Leblanc 2005). Machine learning techniques are increasingly used in non-linear relational models and high dimensional data sets to reduce the effects of the correlation assumption for regression (Breidenbach et al. 2010; Vauhkonen et al. 2010; Gleason and Im 2012). Recently, random forest (RF) has gained considerable attention in the field of remote sensing due to the classification and computational accuracy, and the capability of providing a measure of variable importance (Mitchell et al. 2013; Guan et al. 2012; Pal 2005). RF is a machine learning algorithm that addresses the limitations of Classification and Regression Trees (CART) by bootstrapping samples to iteratively construct a large number of decision trees each grown with a randomized subset of predictors (Breiman 2001). These ‘trees’ cast a unit vote for the most popular class to classify an input vector in RF classification (Pal 2005; Breiman 1999). In regression mode, RF takes the average of the ‘trees’ to make a prediction.

RF has been shown to be more accurate than simple regression techniques for biomass estimation (Gleason and Im 2012; Powell et al. 2010). It doesn’t require assumptions about the relationship between dependent and independent variables and is well suited for analyzing complex non-linear and possibly hierarchical interactions in large data sets like this study (Olden et al. 2008; Ahmed et al. 2015). RF grows a large numbers of trees which does not over fit the data, keeps bias low by random predictor selection and thus can provide better models for prediction (Prasad et al. 2006). Thus, RF

can select the best predictors among many variables developed from a lidar point cloud for biomass estimation.

The objective of this study was to estimate AGB in a shrub-steppe rangeland using airborne lidar-derived vegetation metrics and to extend these estimates to a larger coverage of lidar. We derived 35 metrics from both lidar point clouds and rasterized products at four different resolutions, as a proxy for the estimation of AGB. The result suggests that there is a strong relation between AGB and height-based vegetation metrics, and these metrics can explain up to 76% of the variability in field biomass measurements. We also show that there is no substantial difference between biomass estimation at different resolutions when using point-clouds, whereas resolution has a significant impact on biomass estimation when using rasterized images.

2.2 Study Area and Data

2.2.1 Study Area

The study area is located within the Morley Nelson Snake River Birds of Prey National Conservation Area (NCA), a shrub-steppe rangeland once dominated by big sagebrush (*Artemisia tridentata*). The NCA encompasses about 242,800 hectares of the Snake River Plain ecoregion in southwestern Idaho, USA (**Fig. 1. 1**). It contains other native species including shadscale (*Altriplex confertifolia*), winterfat (*Ceratoides lanata*), budsage (*Artemisia spinescen*), and rabbitbrush (*Chrysothamnus viscidiflorus*) including rapidly invading annual nonnatives like cheatgrass (*Bromus tectorum*). The NCA receives 20 cm precipitation, 74 days with high temperature greater than 32° C, and 98 days with a low temperature below 0° C in an average year. The average annual maximum temperature is about 20° C and the minimum temperature is 6° C (WRCC,

2012). The native vegetation assemblage is composed of an understory of biological crusts and sparse native bunchgrass (*Festuca idahoensis*), overlain by an open canopy of shrubs ranging up to 1.5 m tall (Anderson 2014). Since 1980, over half of the NCA has burned resulting in a mosaic of plant communities, with compositions spanning a gradient between intact native shrublands, shrublands degraded by biological invasion and wildfire, and grasslands where native plants have been fully replaced by cheatgrass and other invasive annuals. Currently 37% or less of the NCA retains an intact native shrubland community (USDI, 2008).

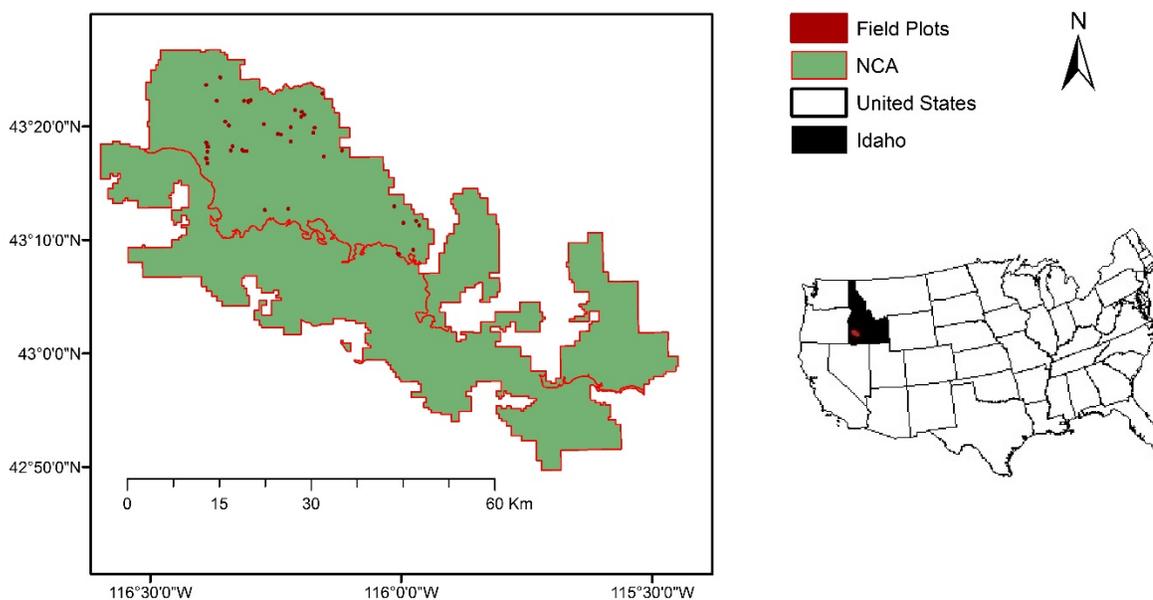


Figure 1.1. The Morley Nelson Snake River Birds of Prey National Conservation Area (NCA), located in southwest Idaho, United States. This study took place in the northwestern portion of the NCA.

2.2.2 Field Sampling

In the summer of 2012 and 2013, a total of forty six ($n = 46$), 100 m by 100 m (one hectare) field plots were established at locations throughout the northwestern NCA by the US Geological Survey, Forest and Rangeland Ecosystem Science Center (USGS FRESC)

(Shinneman et al. 2011). This area was chosen based on the availability of airborne lidar data (see below). A stratified random sampling approach was used to select the plots, and the corners of each plot were precisely located using a survey-grade GNSS (Global Navigation Satellite System). The sites were selected based on accessibility and the goal of capturing a range of plant community compositions. The sampling design for each 1-ha plot included a 3 by 3 grid of 9 subplots of 1 m² each, with 25 m spacing between subplots (**Fig. 1. 3**). Over the 46 plots, a total of 414 subplots were established.

Vegetation within the subplot was then destructively sampled and classified as either herbaceous or shrub. If shrubs were too bulky to be harvested efficiently, a portion was collected for reference and the number of equivalent portions remaining in the quadrat was estimated. The harvested vegetation was oven dried and weighed, and biomass across each 1-ha plot was then calculated as the average from the nine subplots for the herbaceous and shrub classes.

Herbaceous cover ranged from 0 to 100% and shrub cover from 0 to 87%. Across all 414 subplots, herbaceous had a mean biomass of ~ 144 g and shrub had a mean biomass of ~ 414 g. The summary is listed in **Table 1. 1**. The distribution of the field data was skewed towards smaller biomass values (**Fig. 1. 2**). Only two plots had more than 300 g/ m² herbaceous biomass.

Table 1.1 **Statistics of vegetation cover and biomass from the field sites. The results are from n=414 subplots (1 m²) nested inside n=46 plots.**

	Herbaceous Cover (%)	Shrub Cover (%)	Herbaceous Biomass (g/m ²)	Shrub Biomass (g/m ²)
Minimum	0	0	2	0
Maximum	100	87	1207	3301
Mean ± SE	39 ± 1.47	12 ± 0.85	144 ± 7	414 ± 20

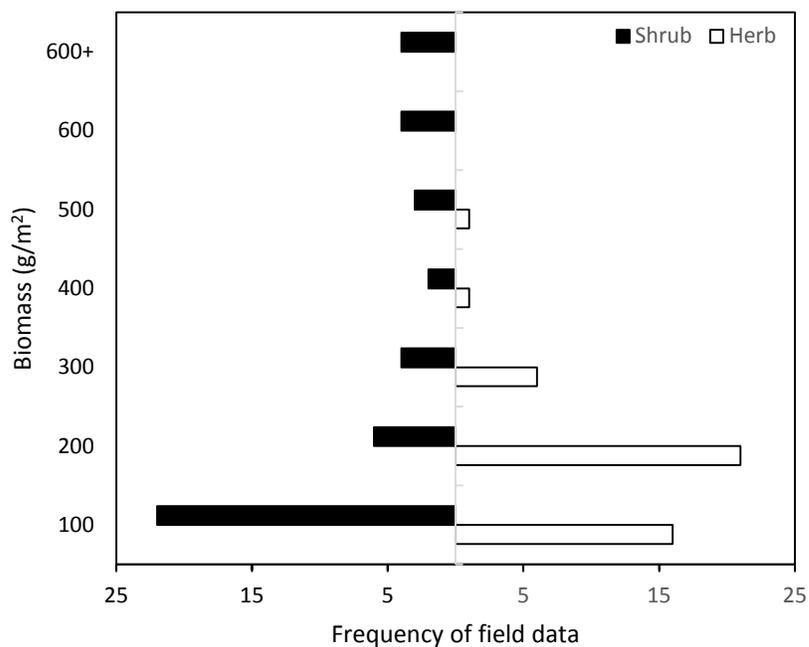
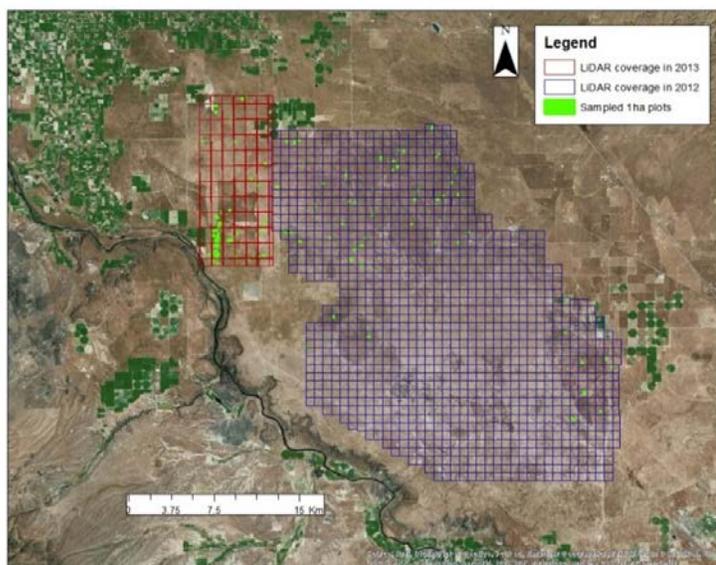


Figure 1.2 Frequency distribution of the in-situ shrub and herbaceous biomass (n=46) with bins of 100 g/ m².

a)



b)

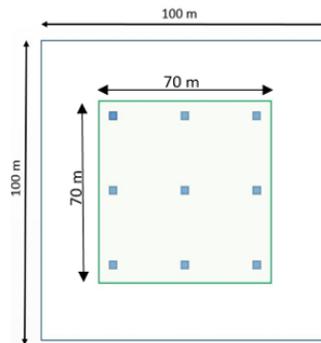


Figure 1.3 Study area within the NCA and (a) lidar coverage and distribution of sampled field plots over the study site and (b) schematic of the field sampling procedure.

2.2.3 Airborne Lidar Data Acquisitions

The discrete small footprint lidar data were collected over 65,194 hectares in 2012 and 9,970 hectares in 2013, with a ALS60 system (Leica Geosystems, Heerbrugg, Switzerland) operated by Watershed Sciences (Corvallis/Portland, OR), with a point density of ~ 8 points per m^2 . The 46 field plots were spatially nested in the lidar footprints. The lidar system was set to acquire $\geq 148,000$ laser pulses per second and was flown at 1,500 meters above ground level, with a scan angle of $48^\circ (\pm 12^\circ)$ from nadir (field of view). An opposing flight line side-lap of $\geq 50\%$ (i.e. 100% overlap) was maintained to increase point density. The absolute vertical accuracy (RMSE_z) was about 0.03 m and the relative accuracy was about 0.024 m. The vertical accuracy was primarily assessed from ground check points on open, bare earth surfaces with level slope ($< 20^\circ$) by the vendor.

2.2.4 Lidar Data Processing

Lidar point cloud data were buffered and height filtered using the ‘BCAL Lidar Tools’ developed for semiarid vegetation (<http://bcal.boisestate.edu/tools/lidar>; Streutker

and Glenn 2006). The height filtering classifies lidar points into ground and vegetation points. The filtering was performed using 5 m canopy spacing, a 5 cm ground threshold, nearest neighbor interpolation and 40 iterations. Two groups of metrics were calculated from resulting vegetation points: metrics based on numerical values (e.g. elevation, canopy height) and metrics based on the density of points (e.g. canopy density). We calculated 33 metrics previously used by Evans et al. 2009 and two additional metrics based on Foliage Height Diversity (FHD) (MacArthur and MacArthur 1961; Maltamo et al. 2014). The FHD measures the foliage arrangement of the plant in the vertical direction and increases by additional layers or by a greater evenness of foliage cover among each layer (Sasaki et al. 2012). All 35 vegetation metrics developed for this study are listed in **Table 1. 2**. In order to explore the effect of rasterization of the point cloud on vegetation metric calculations, we conducted two separate analyses of the 35 metrics, for each plot. The first was to average the metrics extracted from the rasterized vegetation products of the plot and the second was to average the metrics directly from the point cloud of the same plot, with no rasterization. Both rasterized and non-rasterized (or point cloud based) multiband vegetation metrics were produced in four resolutions: 1 m, 7 m, 30 m and 100 m pixel sizes, capturing the spectrum of resolution popular in spectral remote sensing. The Matlab software (Matlab version R2014a, The MathWorks, Inc., Natick Massachusetts, United States) was used to extract the vegetation metrics from the point cloud. These metrics were analyzed at the 100 m by 100 m (1-ha) plot scale and at the 70m by 70m scale (**Fig. 1. 3**). The 70m by 70 m scale was used to reduce the spatial dimensions to encompass the field plot design with the purpose of providing a comparison to the 1-ha plot size by disregarding the non-contributing portion of the plot.

Table 1.2 Lidar metrics and their descriptions used in the analysis.

Lidar Metrics	Description
Minimum Height (H_{\min})	The minimum of all height points within each pixel
Maximum Height (H_{\max})	The maximum of all height points within each pixel
Height Range (H_{range})	The difference of maximum and minimum of all height points within each pixel
Mean Height (H_{mean})	The average of all height points within each pixel
Median Absolute Deviation from Median Height (H_{MAD})	The MAD value of all height points within each pixel $H_{\text{MAD}} = 1.4826 \times \text{median}(\text{height} - \text{median height})$
Mean Absolute Deviation from Mean Height (H_{AAD})	The AAD value of all height points within each pixel $H_{\text{AAD}} = \text{mean}(\text{height} - \text{mean height})$
Height Variance (H_{var})	The variance of all height points within each pixel
Height St. Deviation (H_{std})	The standard deviation of all height points within each pixel This is also called 'absolute vegetation roughness'
Height Skewness (H_{skew})	The skewness of all height points within each pixel
Height Kurtosis (H_{kurt})	The kurtosis of all height points within each pixel
Interquartile Range (H_{IQR}) of Height	The IQR of all height points within each pixel $H_{\text{IQR}} = Q_{75} - Q_{25}$, where Q_x is x^{th} percentile
Height Coefficient of Variation (H_{CV})	The coefficient of variation of all height points within each pixel
Height Percentiles (H_5 , H_{10} , H_{25} etc.)	The 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of all height points within each pixel
Number of Lidar Returns	The total number of all points within each pixel
Number of Lidar Vegetation Returns (nV)	The total number of all the points within each pixel that are above the specified crown threshold value (CT)
Number of Lidar Ground Returns (nG)	The total number of all the points within each pixel that are below the specified ground threshold value (GT)

Total Vegetation Density (Veg_density)	The percent ratio of vegetation returns and ground returns within each pixel. Density = $nV/nG*100$
Vegetation Cover (Veg_cov)	The percent ratio of vegetation returns (nV) and total returns within each pixel
Percentage of Ground Return (pG)	Percent of points within each pixel that are below the specified Ground Threshold
Percent of Vegetation in Height Range (pH ₁ , pH _{2.5} , pH ₁₀ etc.)	Percent of vegetation in height ranges 0-1m, 1-2.5m, 2.5-10m, 10-20m, 20-30m, and >30m within each pixel Percent of Vegetation = Number of vegetation returns in the range/Total vegetation returns
Canopy Relief Ratio (CRR)	Canopy relief ratio of points within each pixel. Canopy relief ratio = $((H_{\text{mean}} - H_{\text{min}}))/(H_{\text{max}} - H_{\text{min}})$
Texture of Heights (H _{text})	Texture of height of points within each pixel. Texture = St. Dev. (Height > Ground Threshold and Height < Crown Threshold)
Foliage Height Diversity (FHD _{all})- All points	Foliage arrangement in the vertical direction. $FHD_{\text{all}} = -\sum p_i \ln p_i$ where p_i is the proportion of horizontal foliage coverage in the i th layer to the sum of the foliage coverage of all the layers
Foliage Height Diversity- Points above ground threshold (FHD _{GT})	FHD calculated only from the points above the ground threshold.

2.3 Statistical Analysis

2.3.1 Regression Analysis

Regression analysis was performed to model the relationship between lidar derived metrics and field AGB at 1 m raster resolution. Total and shrub biomass were considered dependent variables and the 35 lidar derived vegetation metrics were considered independent variables. The common problem with linear regression and its

use in biomass estimation is multicollinearity within the independent variable which might lead to the violation of basic assumptions (Baskerville 1972; Routledge 1990; Lu and Chen 2012). Hence, we used the regression approach adopted by Lefsky et al. (2002), which selected the two most important independent variables that were not collinear to each other using the Pearson's correlation coefficient and thus not violating the basic assumptions of regression.

2.3.2 Random Forest (RF) Regression for Variable Selection

To further assess the relationship between field level biomass with vegetation metrics developed from lidar, we also used the non-parametric machine learning approach, random forest. This study has 'broad data' i.e. many predictors with comparatively low sample size. Random forest is preferred over other methods here as its results have smaller prediction variance and bias and better model performance (Prasad et al. 2006; Mutanga et al. 2012; Mitchell et al. 2015). We used SPM Suite (Salford Predictive Modeler Software Suite version 7, Salford Systems, San Diego, CA) for the implementation of the RF algorithm.

Each RF regression run generated 2000 trees and the maximum number of variables considered per node was kept equal to the square root of the number of variables for the run (Breiman 1999; Breiman 2001). All 35 predictor variables (**Table 1. 2**) were used to perform the initial RF run and each of them were ranked based on their predictive power. The predictive power of the variable or variable ranking was performed by a 'Standard Method': in each tree in the forest, a variable was tested by first scrambling its values and then measuring the decline of accuracy in the model. This means, if a variable substituted with incorrect values can predict the target accurately, then the variable has no relevance

to predicting the outcome and hence is assigned a low score (SPM user guide, 2013). For the best variable selection, we used the backward feature elimination method where lowest performing variables were iteratively removed until the best model was obtained. The best models for total AGB, shrub biomass and herb biomass were determined based on the highest coefficient of determination (R^2), lowest root-mean-square error (RMSE) and maintaining model parsimony (number of predictor variables were kept as low as possible). The variable selection was done not only to reduce the explanatory variables but also to understand which explanatory variables are most suitable to estimate biomass (Ismail et al. 2010). The analyses were performed for all four resolutions for both raster and point cloud i.e. 1m, 7 m, 30 m and 100 m. We also performed a separate analysis for two different dimensions of raster plots across a 1 m scale: 100 m by 100 m (1ha) and 70 m by 70 m.

2.3.3 Nearest Neighbor (NN) Imputation

The best variables selected in RF were used in a Nearest Neighbor imputation with 1 m pixels in the R statistical computing environment (R Development Core Team 2013). In the NN imputation, the estimates for the attributes of interest (e. g. biomass) are produced as weighted averages of the attributes of the reference observation. The reference observations are similar in terms of a distance metric calculated in the predictor space formed by the independent variable (best variables selected by RF) (Vauhkonen et al. 2010; Hudak et al. 2008b). NN imputation methods can use different distance metrics to determine the similarity between target and reference records, including Euclidean, Mahalanobis, Minkowski, fuzzy etc. (Eskelson et al. 2009). However the reference data should cover the entire phenomenon of interest or field site to make an accurate

imputation. Hence, the process was applied to the available lidar coverage to get a contiguous map of biomass. A R package, *yaimpute*, was used which has a built-in function to calculate NN distances based on the RF proximity matrix (Crookston and Finley 2008; Hudak et al. 2008a). A detail explanation of imputation, its types and its fundamental difference with interpolation can be found in Hudak et al. 2008a, 2008b.

2.4 Results

2.4.1 Regression Analysis

The Pearson's correlation analysis identified the metric 'Standard Deviation of Height' (or Absolute Vegetation Roughness) as the variable with highest correlation with total AGB (Pearson's correlation $r = 0.85$) and shrub biomass (Pearson's correlation $r = 0.84$). A regression analysis of total AGB with Standard Deviation of Height provided us with the following equation with an R^2 of 0.72 and P values < 0.05 :

$$\text{Total AGB} = 12374.67 \times \text{Standard Deviation of Height} - 142.058$$

Analysis of the residuals obtained from the above equation was correlated with the remaining of the 34 metrics and 'Skewness of Height' was found to have the highest correlation (Pearson's correlation $r = 0.39$). Hence 'Skewness of Height' was added to the equation resulting in an R^2 of 0.78, P value < 0.05 and F statistic < 0.05 :

$$\text{Total AGB} = 10230 \times \text{Standard Deviation of Height} + 386 \times \text{Skewness of Height} - 226.416$$

Applying the same methodology to the shrub biomass (SB), provided the following model with an R^2 of 0.76, P value < 0.05 and F statistic < 0.05 :

$$\text{SB} = 25655.23 \times \text{Standard Deviation of Height} - 19052.4 \times \text{Mean Absolute Deviation from Median Height} - 169.626$$

The graphs presenting the relationship between observed and predicted total AGB and shrub biomass are shown in **figure 1.4** and **figure 1.5** respectively. The negative shrub biomass shown in **figure 1.5** may be explained by artifacts from the lidar height filtering contributed by herbaceous cover. The herbaceous cover is sensitive to lidar but not taken into account by the shrub biomass model.

A leave-one-out cross-validation performed using the ‘boot’ package in R statistical software (R Development Core Team 2013) yielded a RMSE of approximately 139 g for total AGB and 128 g for shrub biomass. The ‘k’ parameter was 46 (i.e. number of rows) to make the k-fold validation into a leave-one-out cross validation.

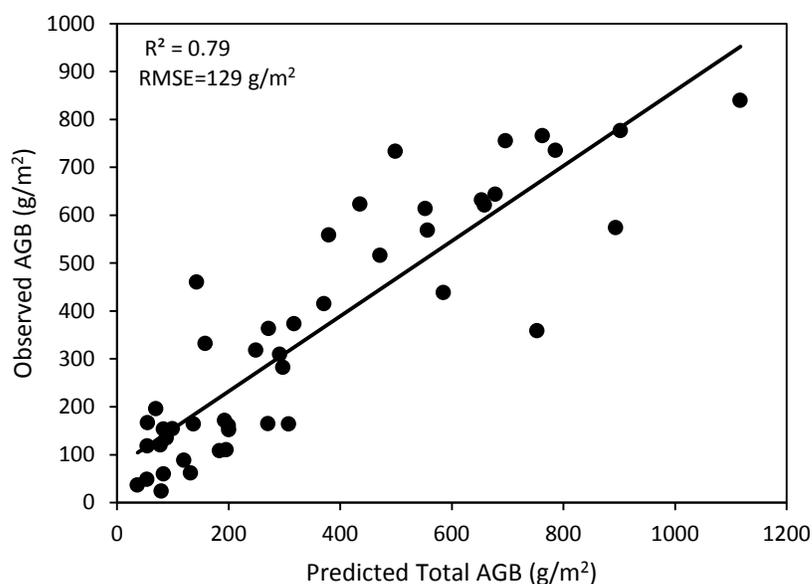


Figure 1.4 Comparison between observed total AGB and total AGB predicted from regression analysis.

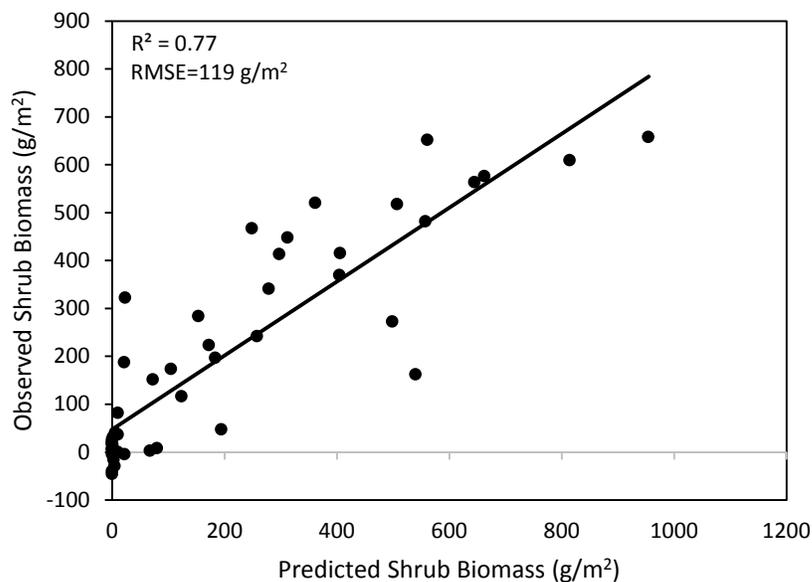


Figure 1.5 Comparison between observed shrub biomass and shrub biomass predicted from regression analysis.

2.4.2 Random Forest Raster Analysis

Lidar derived metrics were found to have a strong relationship with *in-situ* total AGB and shrub biomass using RF regression model. Lidar metrics, including ‘Mean Absolute Deviation (AAD) of Height’ and ‘Standard Deviation of Height’ from the 1 m raster image predicted the total biomass with an R^2 of 0.73 and RMSE of 146 g per m^2 while shrub biomass was predicted with an R^2 of 0.76 and RMSE of 125 g per m^2 (**Table 1.3**).

However, as the raster resolution decreased, the prediction capability of lidar metrics was also reduced with R^2 of 0.70 at 7 m raster, 0.58 at 30 m raster and 0.42 at 100 m for total AGB. Prediction of shrub biomass using raster analysis was also found to be scale dependent.

Table 1.3 Results of random forest regression using raster data processing for total and shrub biomass at different resolutions. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.

	Area(ha)	Resolution(m)	R^2	RMSE(g/m ²)	Best Predictors
Total biomass		1	0.74	141	H_{std} , H_{AAD} , H_{90} , H_{Skew} , H_{var} , H_{text}
		7	0.70	152	H_{text} , FHD_{GT} , H_{95} , H_{AAD}
		30	0.58	180	FHD_{GT} , nV , H_{AAD} , H_5
		100	0.52	188	FHD_{GT} , nV , H_{16} , H_{AAD}
Shrub biomass		1	0.76	152	H_{std} , H_{AAD} , H_{CV} , H_{range} , FHD_{all}
		7	0.67	143	H_{text} , FHD_{GT} , H_{AAD}
		30	0.50	176	FHD_{GT} , H_{AAD} , H_{CV}
		100	0.4	184	H_{text} , H_{50} , pG , nG

We also considered a reduced plot size of 70 m by 70 m to test the effect on biomass estimation. In comparison to the 100 m plot, the reduced plot size didn't improve the estimation of biomass: total AGB was predicted with an R^2 of 0.68 (RMSE=156 g) and shrub biomass with R^2 of 0.75 (RMSE = 126 g). Hence for further analyses, we only considered the plot size of 100 m by 100 m. The results are presented in **Table 1. 4**.

Table 1.4 Results of random forest regression for total biomass and shrub biomass at 70m x 70m plot level. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.

	Area(m ²)	Resolution(m)	R^2	RMSE(g/m ²)	Best Predictors
Total biomass	70 x 70	1	0.68	156	FHD _{all} , H _{std} , H _{AAD} , H _{range} , H _{Skew}
Shrub biomass	70 x 70	1	0.75	126	H _{std} , H _{range} , FHD _{all} , H _{CV}

2.4.3 Random Forest Point Cloud Analysis

We analyzed point clouds at 1 m, 7 m, 30 m and 100 m scales using similar RF regression for this purpose. Unlike the raster processing, the increase in pixel size didn't affect the total AGB prediction capability of the point clouds. The total AGB estimation ability of the RF model from point clouds was not statistically different from raster processing at 1m but was better than predictions using rasters at 7 m, 30 m or 100 m resolution. The results are summarized in **Table 1. 5**.

Table 1.5 Results of random forest regression using point cloud processing for total biomass and shrub biomass. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.

	Area(ha)	Resolution(m)	R^2	RMSE(g/m ²)	Best Predictors
Total biomass	1	1	0.71	147	H_{MAD} , H_{Skew} , H_{IQR} , H_{AAD} , H_{std} , H_{kurt} , H_{90} , H_{CV}
	1	7	0.71	148	H_{text} , H_{IQR}
	1	30	0.70	151	H_{AAD} , H_{95} , H_{IQR} , pH_1 , pG
	1	100	0.67	160	H_{90} , H_{95} , H_{text} , $veg_density$
Shrub biomass	1	1	0.73	129	H_{IQR} , H_{std} , H_{MAD} , H_{CV}
	1	7	0.72	132	H_{text} , H_{90} , H_{IQR} , H_{CV}
	1	30	0.65	146	H_{90} , H_{IQR} , H_{text} , pH_1
	1	100	0.64	151	H_{95} , H_{text} , pH_1 , G_{IQR} , FHD_{GT}

In contrast to shrub and total biomass, herbaceous biomass was poorly predicted by lidar metrics. This was expected as herbaceous vegetation types are short in stature and hence differentiating ground from herbaceous returns in lidar is difficult. The results were consistent across all scales and all processing approaches and hence only results from 1m raster and point cloud are listed in **Table 1. 6**.

Table 1.6 Results of random forest regression for herbaceous biomass. Results for raster processing and point cloud processing are also shown. R^2 and RMSE values are estimated using “out-of-bag” testing. The R^2 value of random forests regression model is the percent variance computed as the ratio of Mean Square Error and Variance of target response subtracted from 1.

	Area(ha)	Source	Resolution(m)	R^2	RMSE(g/m ²)	Best Predictors
Herbaceous biomass	1	Raster	1	0.2	6.86	H _{Skew} , H _{text}
	1	Point Cloud	1	0.19	7.54	H _{CV} , H _{text} , H _{Skew}

2.4.4 Imputation

A spatially explicit contiguous 1 m aboveground biomass map for the entire lidar coverage (~75,164 hectares) was produced by imputation using predictors associated with the 1m raster at 1 ha. A cross validation for the RF regression was not performed as Out-of-Bag (OOB) error acts as an internal cross-validation and is capable of providing an unbiased estimate of error (Naidoo et al. 2012; Prasad et al. 2006; Prinzie and Van den Poel 2008). While the results from the linear regression were similar to the RF regression, we chose to use the RF results in the final imputation. This was based on the RF versatility in handling complex relationships between variables in large data sets and its invulnerability from normality, homogeneity and correlation assumptions (Ismail et al. 2010; Mutanga et al. 2012; Vincenzi et al. 2011).

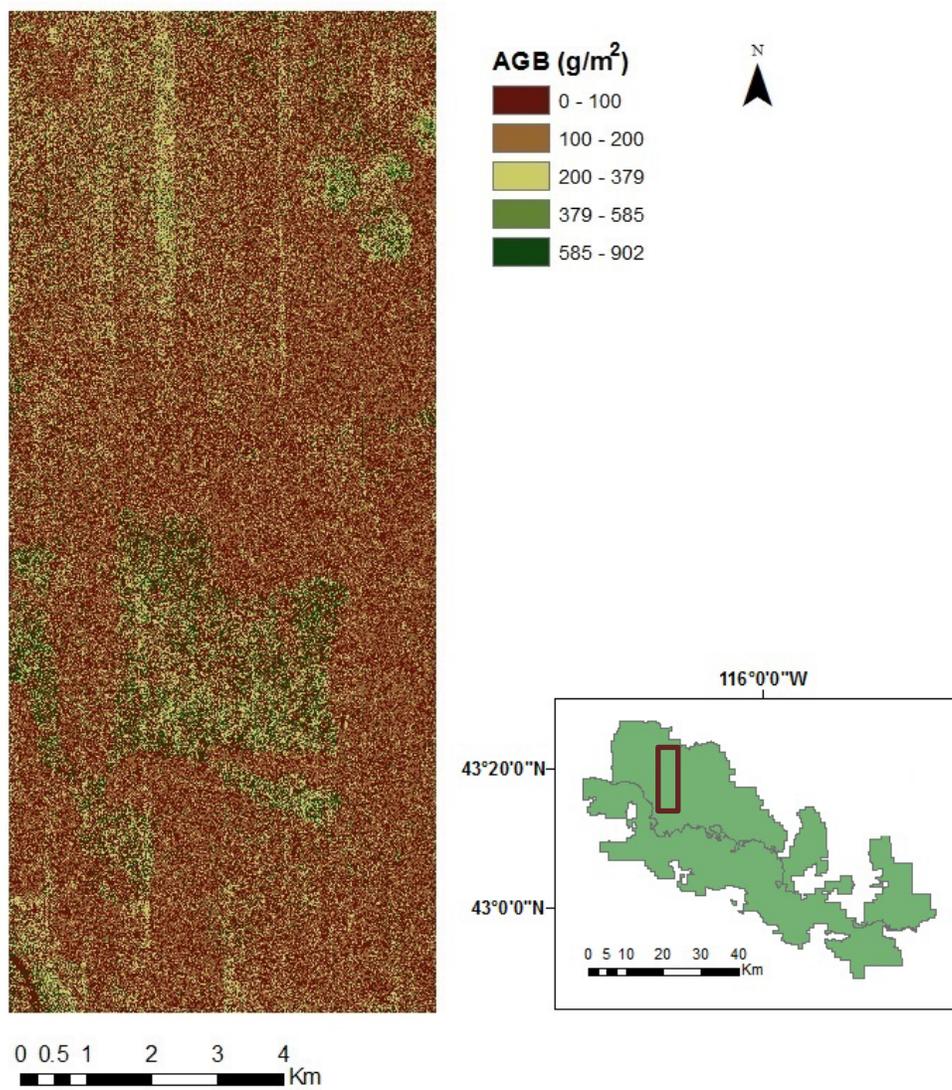


Figure 1.6 Imputed total AGB map at 1 m resolution of a selected area indicated by the brown frame.

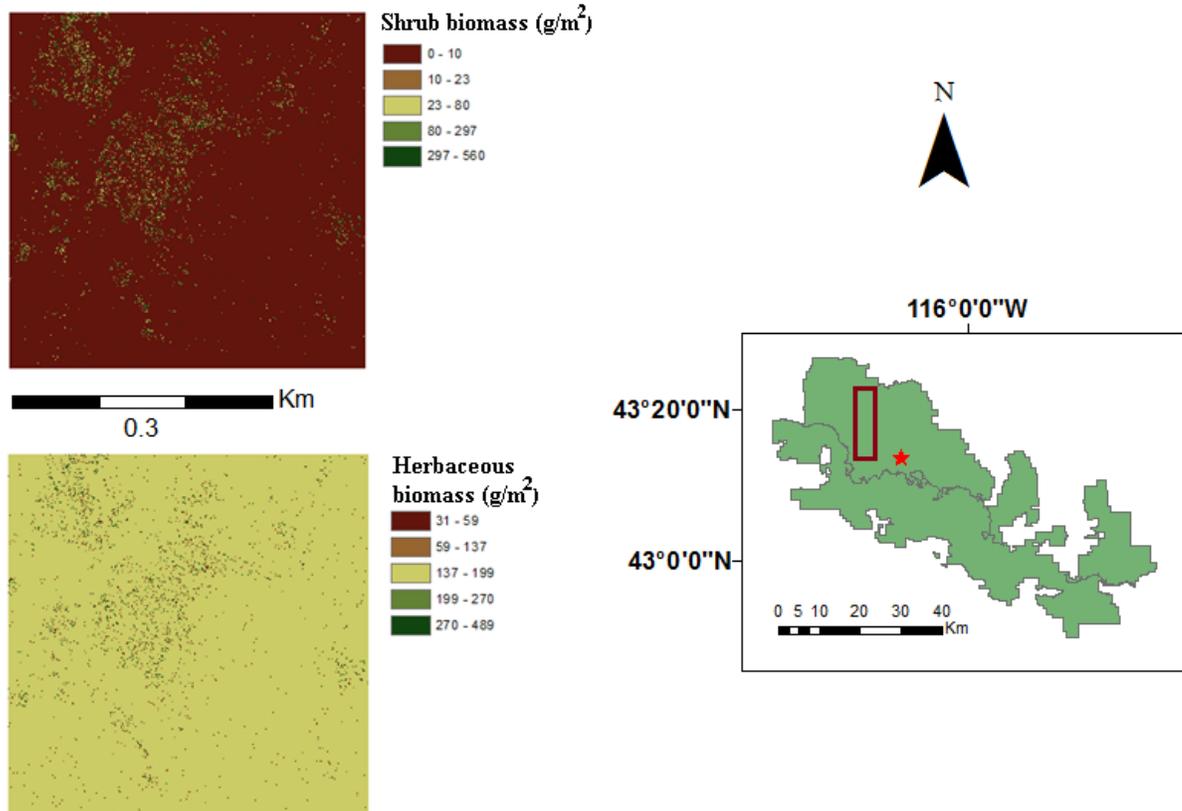


Figure 1.7 Comparison of imputed shrub and herbaceous biomass maps for a selected area indicated by red star. Rectangular box refers to area in Fig. 1. 6.

2.5 Discussion

Both the RF and regression models showed a high correlation of lidar derived shrub height metrics with total AGB. Interestingly, in nearly all RF models with high R^2 and low RMSE, ‘Standard Deviation of Height’ along with ‘Mean Absolute Deviation (AAD) of height’ and ‘Median Absolute Deviation (MAD) of height’ scored higher among other predictors of total AGB and shrub biomass. For instance, when used in combination, Standard Deviation of Height and AAD at the 1 m scale explained about 76% of variability in total AGB. ‘Standard Deviation of Height’ alone estimated total AGB with R^2 of 0.73 using simple linear regression (**Fig. 1. 8**).

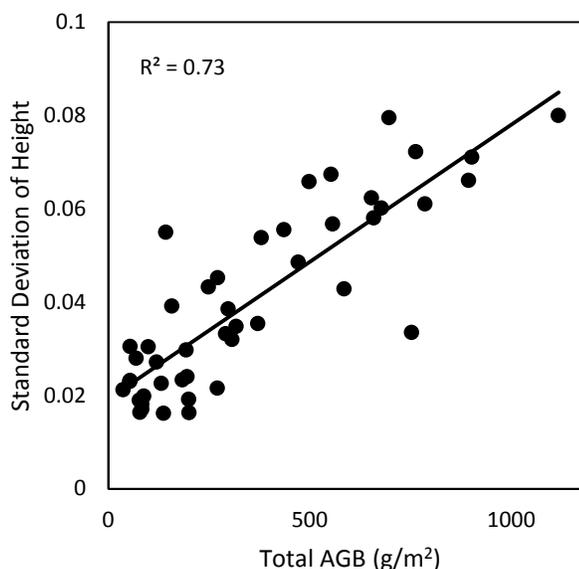


Figure 1.8 Simple linear regression between total AGB and Standard Deviation of Height

Standard deviation of vegetation height measures the roughness of vegetation by calculating the deviation of all height points within each pixel. Contrary to other variables that only take into account true height, this metric represents the variability in vegetation height. Our field site is heterogeneous in terms of shrub and herbaceous height and the patches of sagebrush are distributed randomly. In related lidar biomass studies, Olsoy et al. (2014) and Li et al. (2015) showed that the volume and percent vegetation cover, respectively, are important predictors of total AGB in semiarid rangeland. Li et al. (2015) demonstrated that airborne lidar derived percent vegetation cover can explain about 87% and 95% of variation in reference biomass at 5 m and 30 m resolutions, respectively. Hence, a metric able to capture volume of the shrub (by combining shrub height as well as width) as well as percent vegetation cover may be even more efficient in estimating biomass in rangeland. Further analysis of these metrics are warranted to better estimate the total AGB with lidar measurement.

Rangeland lidar studies have generally been limited to assessing shrub height, canopy cover, volume, species detection and fire severity (Ritchie et al. 2006; Sankey and Bond 2011; Wang and Glenn 2009a). Some studies (Streutker and Glenn 2006; Glenn et al. 2011; Spaete et al. 2011) have shown consistent underestimation of vegetation height by 30-50% using airborne lidar. The underestimation is attributed to the low probability of the laser hitting the top of the canopy (Mitchell et al. 2011). In addition, erroneous modeling of ground elevations due to shrubs being close to the ground is possible (Vierling et al. 2012). Glenn et al. (2011) suggest using mean point densities > 4 points per m^2 to accurately model biomass. In this study, we explained 76% of the variability in shrub biomass using an average point density of 8 points per m^2 . The 25% error may be credited to the uncertainties associated with sparse vegetation distribution, misclassification of canopy as ground and underestimation of the vegetation height (Riaño et al. 2007; Streutker and Glenn 2006).

While Terrestrial Laser Scanning (TLS) has also been employed in many studies, volume and height based metrics have been found to better estimate shrub biomass in dry land. For example, Olsoy et al. (2014) and Greaves et al. (2015) used TLS point clouds to develop a 3-D convex hull and voxel to successfully estimate individual sagebrush biomass and Arctic deciduous shrubs, respectively. Ku et al. (2012) demonstrated strong correlation of point cloud height variables with biomass at the plot scale using regression analysis. The height variables used in their study were maximum height, minimum height, standard deviation of height and 25th, 50th, 75th, 90th and 95th percentile of height; a subset of metrics used in this study. Vierling et al. (2012) also successfully demonstrated shrub characterization in rangelands using point clouds by retrieving shrub

heights with a strong correlation ($p < 0.01$) and R^2 of 0.94. Estornell et al. (2012) used airborne lidar and multispectral imagery to satisfactorily estimate biomass and volume in a Mediterranean shrubland. They used median, standard deviation and percentile of height derived from lidar as the best predictors, explaining up to 78% and 84% of variability for biomass and volume respectively. Our finding is similar to these studies as we found lidar metrics, especially height related metrics as the best predictor of shrub biomass.

The raster data models are most commonly used to represent lidar data as they are relatively easy to store and easier to process than point clouds. When point data are resampled into a grid space, the 3D data are converted to a 2D grid, causing a loss in detail (Yunfei et al. 2008). Raster image creation is based on aggregation of the irregularly distributed points returned value in grid cells. For cells that contain no points, interpolation is performed. El-Ashmawy and Shaker (2014) compared raster and point cloud classification in two areas in British Columbia. They found that the overall accuracy of point cloud classification was slightly better than the overall accuracy of raster classification. In this study, processing of the point cloud data significantly improved the estimation of total AGB and shrub biomass at coarser scales (7 m, 30 m and 100 m) in comparison to the raster image processing (**Table 1. 3 and 1. 5**). As shown in **figure 1. 9**, for raster processing, the R^2 decreases and RMSE increases as the resolution coarsens. However, the resolution of point cloud processing had comparatively negligible effect on the total AGB estimation (**fig 1. 10**). There is almost no loss of detail while extracting or averaging information from the original point cloud. This is because the point cloud metrics are not produced from interpolation of values.

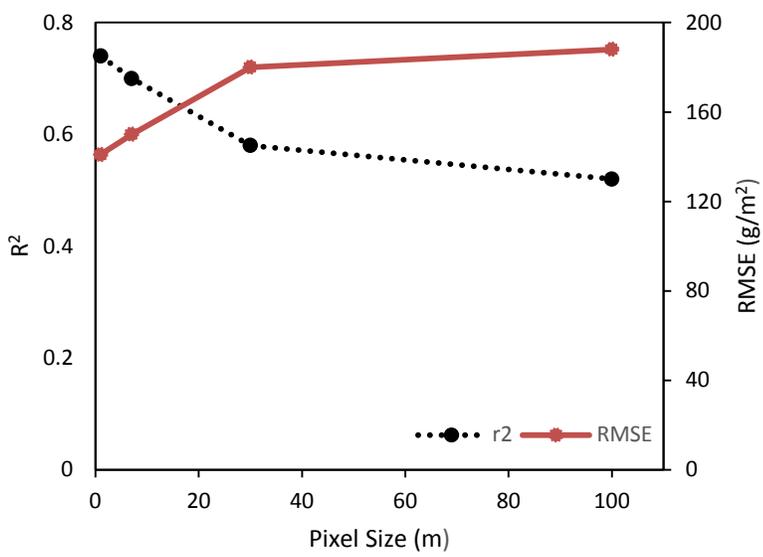


Figure 1.9 Relationship between pixel size and R² and RMSE in raster data processing. As the pixel resolution decreases, the R² decreases and RMSE increases. The figure represents the results from total AGB only but shrub biomass follows a similar trend.

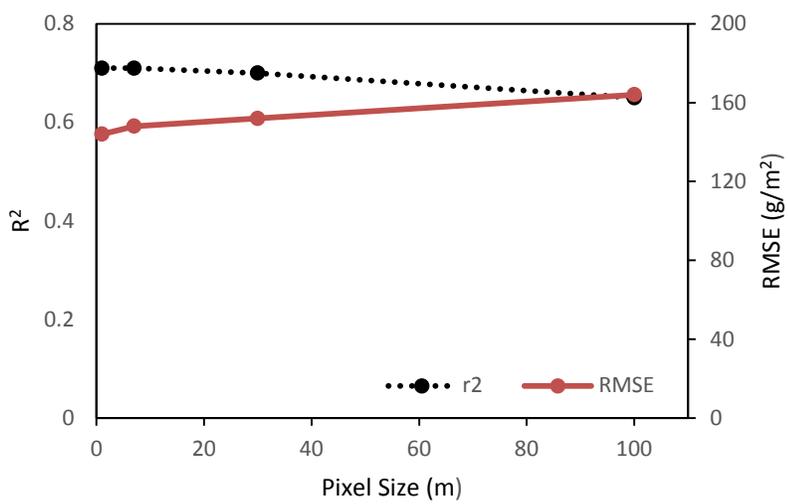


Figure 1.10 Relation between pixel size and R² and RMSE in point cloud processing. Comparatively, the cell size has little effect on the R² and RMSE. The figure represents the results from total AGB only but shrub biomass follows a similar trend.

For 1 m resolution, however, the point cloud processing was not significantly different than raster data processing. The increased performance of finer raster cell sizes

may be credited to what we can term as the ‘edge effect’ and ‘boundary shift effect’.

When the fine scale rasters are created, the rasterization process takes into account less point clouds that are outside of the pixel boundary but are in close proximity.

Furthermore, while extracting information, the pixel in a raster generally will not coincide with the actual plot in the field. But extraction of information from a finer resolution is comparatively less influenced by the values from adjoining peripherals pixels than coarser pixels (**fig 1. 11**). This decreases error in the finer resolution and helps finer pixel to represent the reality more closely. The results indicate that the rasterization method preserves most of the 3D point cloud vegetation characteristics at fine resolutions. However, rasterization is not an ideal approach at coarser scales such as 100 m.

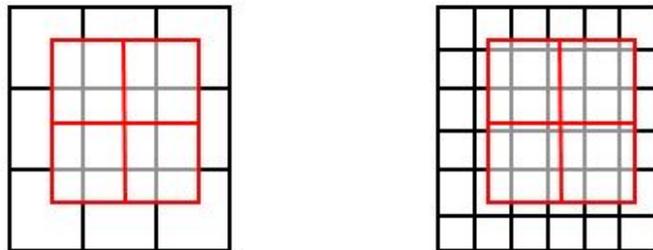


Figure 1.11 Schematic showing the difference in extraction of vegetation characteristics due to different raster pixel size. The red cell represents the field plot of interest. The black cells in the left figure show coarser resolution pixels and the black cells in the right represent finer resolution pixels.

The results from point cloud processing may be further improved by taking into account the intensity (sometimes called amplitude) in addition to the 3D information of the laser returns. Backscattered laser intensity in lidar is determined by the reflectance characteristics of objects in the near infrared spectra which can be used to identify land-

cover classes (Lang and McCarty 2009; Yoon et al. 2008). Since, lidar intensity and structural information can be used to identify same feature, the land-cover classes can be used for better qualitative and quantitative visual analysis of vegetation (Wang and Glenn 2009b).

As shown in the schematic in **figure 1. 3**, the averaging of biomass in subplots to get the *in-situ* plot level biomass leaves non-contributing space in the peripheral part of the plot. The reduced 70 m by 70 m plots were used to test the effect of these non-contributing portions of the plot. These reduced spatial dimensions more closely matched the field plot design and hence examined the effect of plot size on biomass estimates. However, as shown in **Table 1. 4**, the use of the reduced plot size did not improve results. One reason may be that the size of the field plot used in this study was not big enough to capture the individual heterogeneity in the distribution of vegetation at a large scale. Thus the biomass distribution inside each plot was almost the same, the average biomass and hence the estimated biomass was identical between the reduced 70 m by 70 m plot and original 100 m by 100 m plot.

The use of RF imputation methods places a high requirement on *in-situ* data as the biases in the *in-situ* data are represented in the final results (Vauhkonen et al. 2010). This is likely the reason behind the appearance of the long linear features of relatively high biomass in the resulting imputation map shown in **figure 1. 6**. Although the aggregation of biomass inside 9 subplots of 1 m² area might have been representative of herbaceous and small shrubs in a 1 ha plot, substantial error in the field data may have been introduced because of large shrubs close to the subplot edge. Moreover, an attempt of estimating biomass in rangeland without delineating species level classification can have

a major disadvantage when vegetation have similar structural arrangements but different AGB (Anderson 2014). For example, bluegrass (*Poa secunda*) which is a perennial can be incorrectly identified as cheatgrass which is an annual non-native herbaceous grass.

2.6 Conclusion

This study explored the use of lidar to map aboveground biomass in a semi-arid rangeland. This will allow a better management of native shrubs and their obligates, projection of wildfire and accurate inventory of vegetation in large landscape. Using the RF model and linear regression, we successfully established that lidar derived metrics can be used as a proxy to estimate biomass and can be further imputed to produce a contiguous map covering a large area. We also demonstrated the advantages of point cloud processing over raster processing at coarser scales. While this method can be reproduced to estimate AGB of sagebrush dominated rangelands, a different set of lidar metrics may prove to be better predictors in other rangeland ecosystems, depending on the geography, density, stature and distribution of shrubs. Furthermore, the estimation process may be further improved by including more representative *in-situ* data, considering the effect of topographic and elevation factors and synergetic use of other remote sensing data such as multispectral and hyperspectral data.

2.7 References

- Ahmed, Oumer S., Steven E. Franklin, Michael A. Wulder, and Joanne C. White. 2015. "Characterizing Stand-Level Forest Canopy Cover and Height Using Landsat Time Series, Samples of Airborne LiDAR, and the Random Forest Algorithm." *ISPRS Journal of Photogrammetry and Remote Sensing* 101 (March): 89–101. doi:10.1016/j.isprsjprs.2014.11.007.
- Anderson, Kyle. 2014. "Vegetation Measurement in Sagebrush Steppe Using Terrestrial Laser Scanning." Idaho State University.

- Angell, Raymond F, Tony Svejcar, Jon Bates, Nicanor Z Saliendra, and Douglas A Johnson. 2001. "Bowen Ratio and Closed Chamber Carbon Dioxide Flux Measurements over Sagebrush Steppe Vegetation." *Agricultural and Forest Meteorology* 108 (2): 153–61. doi:10.1016/S0168-1923(01)00227-1.
- Barbour, Michael G., and William Dwight Billings. 2000. *North American Terrestrial Vegetation*. Cambridge University Press.
- Baskerville, G. L. 1972. "Use of Logarithmic Regression in the Estimation of Plant Biomass." *Canadian Journal of Forest Research* 2 (1): 49–53. doi:10.1139/x72-009.
- Berterretche, M., Hudak, A. T., Cohen, W. B., Maierasperger, T. K., Gower, S. T., & Dungan, J. (2005). Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment*, 96(1), 49-61.
- Bonham, Charles D. 2013. *Measurements for Terrestrial Vegetation*. John Wiley & Sons.
- Bork, Edward W., and Jason G. Su. 2007. "Integrating LIDAR Data and Multispectral Imagery for Enhanced Classification of Rangeland Vegetation: A Meta Analysis." *Remote Sensing of Environment* 111 (1): 11–24. doi:10.1016/j.rse.2007.03.011.
- Branson, F. A., Reuben F. Miller, and I. S. McQueen. 1976. "Moisture Relationships in Twelve Northern Desert Shrub Communities Near Grand Junction, Colorado." *Ecology* 57 (6): 1104–24. doi:10.2307/1935039.
- Breidenbach, Johannes, Erik Næsset, Vegard Lien, Terje Gobakken, and Svein Solberg. 2010. "Prediction of Species Specific Forest Inventory Attributes Using a Nonparametric Semi-Individual Tree Crown Approach Based on Fused Airborne Laser Scanning and Multispectral Data." *Remote Sensing of Environment* 114 (4): 911–24. doi:10.1016/j.rse.2009.12.004.
- Breiman, Leo. 1999. "Random Forests--Random Features." Statistics Technical Reports. University of California at Berkeley, Berkeley, California: Statistics Department, University of California, Berkeley. <http://stat-reports.lib.berkeley.edu/accessPages/567.html>.

- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.
doi:10.1023/A:1010933404324.
- Brown, Joel R., and Steve Archer. 1999. "Shrub Invasion of Grassland: Recruitment Is Continuous and Not Regulated by Herbaceous Biomass or Density." *Ecology* 80 (7): 2385–96. doi:10.1890/0012-9658(1999)080[2385:SIOGRI]2.0.CO;2.
- Campbell, James B., and Randolph H. Wynne. 2011. *Introduction to Remote Sensing, Fifth Edition*. 5th edition. New York: The Guilford Press.
- Canfield, R. H. 1941. "Application of the Line Interception Method in Sampling Range Vegetation." *Journal of Forestry* 39 (4): 388–94.
- Cardon, Zoe G., John M. Stark, Patrick M. Herron, and Jed A. Rasmussen. 2013. "Sagebrush Carrying out Hydraulic Lift Enhances Surface Soil Nitrogen Cycling and Nitrogen Uptake into Inflorescences." *Proceedings of the National Academy of Sciences* 110 (47): 18988–93. doi:10.1073/pnas.1311314110.
- Chen, Qi, Gaia Vaglio Laurin, John J. Battles, and David Saah. 2012. "Integration of Airborne Lidar and Vegetation Types Derived from Aerial Photography for Mapping Aboveground Live Biomass." *Remote Sensing of Environment* 121 (June): 108–17. doi:10.1016/j.rse.2012.01.021.
- Chopping, M., Su, L., Rango, A., Martonchik, J. V., Peters, D. P., & Laliberte, A. 2008. Remote sensing of woody shrub cover in desert grasslands using MISR with a geometric-optical canopy reflectance model. *Remote Sensing of Environment*, 112(1), 19-34.
- Clark, Patrick E., Stuart P. Hardegree, Corey A. Moffet, and Fredrick B. Pierson. 2008. "Point Sampling to Stratify Biomass Variability in Sagebrush Steppe Vegetation." *Rangeland Ecology & Management* 61 (6): 614–22. doi:10.2111/07-147.1.
- Connelly, John W., Michael A. Schroeder, Alan R. Sands, and Clait E. Braun. 2000. "Guidelines to Manage Sage Grouse Populations and Their Habitats." *Wildlife Society Bulletin* 28 (4): 967–85.
- Crookston, Nicholas L., and Andrew O. Finley. 2008. *yaImpute: An R Package for kNN Imputation*. <http://www.treesearch.fs.fed.us/pubs/29365>.

- Debouk, Haifa, Ramon Riera-Tatché, and Cristina Vega-García. 2013. "Assessing Post-Fire Regeneration in a Mediterranean Mixed Forest Using Lidar Data and Artificial Neural Networks." *Photogrammetric Engineering & Remote Sensing* 79 (12): 1121–30. doi:10.14358/PERS.79.12.1121.
- El-Ashmawy, N., and A. Shaker. 2014. "Raster Vs. Point Cloud LiDAR Data Classification." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 7 (September): 79–83. doi:10.5194/isprsarchives-XL-7-79-2014.
- Eskelson, Bianca N. I., Hailemariam Temesgen, Valerie Lemay, Tara M. Barrett, Nicholas L. Crookston, and Andrew T. ; Hudak. 2009. "The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases." *Scandinavian Journal of Forest Research* 24: 235–46.
- Fernandes, Richard, and Sylvain G. Leblanc. 2005. "Parametric (modified Least Squares) and Non-Parametric (Theil–Sen) Linear Regressions for Predicting Biophysical Parameters in the Presence of Measurement Errors." *Remote Sensing of Environment* 95 (3): 303–16. doi:10.1016/j.rse.2005.01.005.
- García-Gutiérrez, Jorge, Eduardo González-Ferreiro, Daniel Mateos-García, Jose C. Riquelme-Santos, and David Miranda. 2011. "A Comparative Study between Two Regression Methods on LiDAR Data: A Case Study." In *Hybrid Artificial Intelligent Systems*, edited by Emilio Corchado, Marek Kurzyński, and Michał Woźniak, 311–18. Lecture Notes in Computer Science 6679. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-21222-2_38.
- Gleason, Colin J., and Jungho Im. 2012. "Forest Biomass Estimation from Airborne LiDAR Data Using Machine Learning Approaches." *Remote Sensing of Environment* 125 (October): 80–91. doi:10.1016/j.rse.2012.07.006.
- Glenn, N. F., Mundt, J. T., Weber, K. T., Prather, T. S., Lass, L. W., & Pettingill, J. 2005. Hyperspectral data processing for repeat detection of small infestations of leafy spurge. *Remote Sensing of Environment*, 95(3), 399-412.

- Glenn, N. F., L. P. Spaete, T. T. Sankey, D. R. Derryberry, S. P. Hardegree, and J. J. Mitchell. 2011. "Errors in LiDAR-Derived Shrub Height and Crown Area on Sloped Terrain." *Journal of Arid Environments* 75 (4): 377–82.
doi:10.1016/j.jaridenv.2010.11.005.
- Greaves, Heather E., Lee A. Vierling, Jan U. H. Eitel, Natalie T. Boelman, Troy S. Magney, Case M. Prager, and Kevin L. Griffin. 2015. "Estimating Aboveground Biomass and Leaf Area of Low-Stature Arctic Shrubs with Terrestrial LiDAR." *Remote Sensing of Environment* 164 (July): 26–35. doi:10.1016/j.rse.2015.02.023.
- Guan, H., Yu, J. and Luo, L., 2012. Random Forests-Based Feature Selection for Land-Use Classification Using LIDAR Data and Orthoimagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39, p.B7.
- Guo, Qinfeng. 2007. "The Diversity–biomass–productivity Relationships in Grassland Management and Restoration." *Basic and Applied Ecology* 8 (3): 199–208.
doi:10.1016/j.baae.2006.02.005.
- Haack, B., & Rafter, A. 2010. Regression estimation techniques with remote sensing: a review and case study. *Geocarto International*, 25(1), 71-82.
- Hall, S. A., I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker. 2005. "Estimating Stand Structure Using Discrete-Return Lidar: An Example from Low Density, Fire Prone Ponderosa Pine Forests." *Forest Ecology and Management* 208 (1–3): 189–209. doi:10.1016/j.foreco.2004.12.001.
- Hemstrom, M. A., Wisdom, M. J., Hann, W. J., Rowland, M. M., Wales, B. C., & Gravenmier, R. A. (2002). Sagebrush-Steppe Vegetation Dynamics and Restoration Potential in the Interior Columbia Basin, USA. *Conservation Biology*, 16(5), 1243-1255.
- Heitschmidt, Rodney K., and Jerry W. Stuth, eds. 1991. *Grazing Management: An Ecological Perspective*. Portland, Or: Timber Press, Incorporated.
- Hudak, Andrew T., Nicholas L. Crookston, Jeffrey S. Evans, Michael K. Falkowski, Alistair M. S. Smith, Paul E. Gessler, and Penelope Morgan. 2006. *Regression*

Modeling and Mapping of Coniferous Forest Basal Area and Tree Density from Discrete-Return Lidar and Multispectral Data.

<http://www.treesearch.fs.fed.us/pubs/24612>.

- Hudak, Andrew T., Nicholas L. Crookston, Jeffrey S. Evans, David E. Hall, and Michael J. Falkowski. 2008a. "Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data." *Remote Sensing of Environment, Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue*, 112 (5): 2232–45. doi:10.1016/j.rse.2007.10.009.
- Hudak, Andrew T., Jeffrey S. Evans, Nicholas L. Crookston, Michael J. Falkowski, Brant K. Steigers, Rob Taylor, and Halli Hemingway. 2008b. "Aggregating Pixel-Level Basal Area Predictions Derived from LiDAR Data to Industrial Forest Stands in North-Central Idaho." In , 133–45. <http://www.treesearch.fs.fed.us/pubs/30976>.
- Hudak, A., Evans, J., & Smith, A. 2009. LiDAR Utility for Natural Resource Managers. *USDA Forest Service / UNL Faculty Publications*. Retrieved from <http://digitalcommons.unl.edu/usdafsfacpub/202>.
- Ismail, Riyadh, Onesimo Mutanga, and Lalit Kumar. 2010. "Modeling the Potential Distribution of Pine Forests Susceptible to Sirex Noctilio Infestations in Mpumalanga, South Africa." *Transactions in GIS* 14 (5): 709–26. doi:10.1111/j.1467-9671.2010.01229.x.
- Knick, Steven T. 1999. "Requiem for a Sagebrush Ecosystem?" <https://research.wsulibs.wsu.edu:8443/xmlui/handle/2376/1169>.
- Ku, N. W., S. C. Popescu, R. J. Ansley, H. L. Perotto-Baldivieso, and A. M. Filippi. 2012. "Assessment of Available Rangeland Woody Plant Biomass with a Terrestrial Lidar System." *Special Issue: Advances in Terrestrial Lidar Techniques and Applications*. 78 (4): 349–61.
- Laurin, Gaia, Qi Chen, Jeremy A. Lindsell, David A. Coomes, Fabio Del Frate, Leila Guerriero, Francesco Pirotti, and Riccardo Valentini. 2014. "Above Ground Biomass Estimation in an African Tropical Forest with Lidar and Hyperspectral

- Data.” *ISPRS Journal of Photogrammetry and Remote Sensing* 89 (March): 49–58. doi:10.1016/j.isprsjprs.2014.01.001.
- Lefsky, M. A., W. B. Cohen, Geoffrey G. Parker, and D. J. Harding. 2002. “Lidar Remote Sensing for Ecosystem Studies.”
<https://repository.si.edu/handle/10088/2924>.
- Lefsky, Michael A., Warren B. Cohen, David J. Harding, Geoffrey G. Parker, Steven A. Acker, and S. Thomas Gower. 2002. “Lidar Remote Sensing of above-Ground Biomass in Three Biomes.” *Global Ecology and Biogeography* 11 (5): 393–99. doi:10.1046/j.1466-822x.2002.00303.x.
- Li, Aihua, Nancy F. Glenn, Peter J. Olsoy, Jessica J. Mitchell, and Rupesh Shrestha. 2015. “Aboveground Biomass Estimates of Sagebrush Using Terrestrial and Airborne LiDAR Data in a Dryland Ecosystem.” *Agricultural and Forest Meteorology* 213 (November): 138–47. doi:10.1016/j.agrformet.2015.06.005.
- Li, Peng, Luguang Jiang, and Zhiming Feng. 2013. “Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors.” *Remote Sensing* 6 (1): 310–29. doi:10.3390/rs6010310.
- Lin, Yi, Juha Hyypä, Antero Kukko, Anttoni Jaakkola, and Harri Kaartinen. 2012. “Tree Height Growth Measurement with Single-Scan Airborne, Static Terrestrial and Mobile Laser Scanning.” *Sensors* 12 (9): 12798–813. doi:10.3390/s120912798.
- Liu, Jian, Ming Dong, Shi Li Miao, Zhen Yu Li, Ming Hua Song, and Ren Qing Wang. 2006. “Invasive Alien Plants in China: Role of Clonality and Geographical Origin.” *Biological Invasions* 8 (7): 1461–70. doi:10.1007/s10530-005-5838-x.
- MacArthur, R. H., & MacArthur, J. W. 1961. On Bird Species Diversity. *Ecology*, 42(3), 594–598. <http://doi.org/10.2307/1932254>
- Maltamo, M., Næsset, E., & Vauhkonen, J. (2014). *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. Springer Science & Business Media.

- Miller, R. F., and J. A. Rose. 2006. "Fire History and Western Juniper Encroachment in Sagebrush Steppe." *Journal of Range Management Archives* 52 (6): 550–59.
- Miller, Richard F., Steven T. Knick, David A. Pyke, Cara W. Meinke, Steven E. Hanser, Michael J. Wisdom, and Ann L. Hild. 2011. "Characteristics of Sagebrush Habitats and Limitations to Long-Term Conservation." *Studies in Avian Biology* 38: 145–84.
- Mitchell, J J, and N F Glenn. 2009. "Subpixel Abundance Estimates in Mixture-Tuned Matched Filtering Classifications of Leafy Spurge (*Euphorbia Esula* L.)." *International Journal of Remote Sensing* 30 (23): 6099–6119.
- Mitchell, J J, N Glenn, T Sankey, D Derryberry, M Anderson, and R Hruska. 2011. "Small-Footprint Lidar Estimations of Sagebrush Canopy Characteristics." *Photogrammetric Engineering & Remote Sensing* 77 (5).
- Mitchell, J.J., Shrestha, R., Moore-Ellison, C.A. and Glenn, N.F., 2013. Single and multi-date Landsat classifications of basalt to support soil survey efforts. *Remote Sensing*, 5(10), pp.4857-4876.
- Mitchell, Jessica J., Rupesh Shrestha, Lucas P. Spaete, and Nancy F. Glenn. 2015. "Combining Airborne Hyperspectral and LiDAR Data across Local Sites for Upscaling Shrubland Structural Information: Lessons for HypsIRI." *Remote Sensing of Environment*, Special Issue on the Hyperspectral Infrared Imager (HypsIRI), 167 (September): 98–110. doi:10.1016/j.rse.2015.04.015.
- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Mutanga, Onisimo, Elhadi Adam, and Moses Azong Cho. 2012. "High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random Forest Regression Algorithm." *International Journal of Applied Earth Observation and Geoinformation* 18 (August): 399–406. doi:10.1016/j.jag.2012.03.012.
- Naidoo, L., M. A. Cho, R. Mathieu, and G. Asner. 2012. "Classification of Savanna Tree Species, in the Greater Kruger National Park Region, by Integrating

- Hyperspectral and LiDAR Data in a Random Forest Data Mining Environment.” *ISPRS Journal of Photogrammetry and Remote Sensing* 69 (April): 167–79. doi:10.1016/j.isprsjprs.2012.03.005.
- Olden, Julian D., Lawler, and N. LeRoy Poff. 2008. “Machine Learning Methods Without Tears: A Primer for Ecologists.” *The Quarterly Review of Biology* 83 (2): 171–93. doi:10.1086/587826.
- Olsoy, Peter J., Nancy F. Glenn, Patrick E. Clark, and DeWayne R. Derryberry. 2014. “Aboveground Total and Green Biomass of Dryland Shrub Derived from Terrestrial Laser Scanning.” *ISPRS Journal of Photogrammetry and Remote Sensing* 88 (February): 166–73. doi:10.1016/j.isprsjprs.2013.12.006.
- Pal, M. 2005. “Random Forest Classifier for Remote Sensing Classification.” *International Journal of Remote Sensing* 26 (1): 217–22. doi:10.1080/01431160412331269698.
- Perez-Quezada, J. F., C. A. Delpiano, K. A. Snyder, D. A. Johnson, and N. Franck. 2011. “Carbon Pools in an Arid Shrubland in Chile under Natural and Afforested Conditions.” *Journal of Arid Environments* 75 (1): 29–37. doi:10.1016/j.jaridenv.2010.08.003.
- Pieper, Rex D. 1988. “Rangeland Vegetation Productivity and Biomass.” In *Vegetation Science Applications for Rangeland Analysis and Management*, edited by P. T. Tueller, 449–67. Handbook of Vegetation Science 14. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-009-3085-8_18.
- Polley, H. Wayne, Brian J. Wilsey, and Justin D. Derner. 2007. “Dominant Species Constrain Effects of Species Diversity on Temporal Variability in Biomass Production of Tallgrass Prairie.” *Oikos* 116 (12): 2044–52. doi:10.1111/j.2007.0030-1299.16080.x.
- Powell, Scott L., Warren B. Cohen, Sean P. Healey, Robert E. Kennedy, Gretchen G. Moisen, Kenneth B. Pierce, and Janet L. Ohmann. 2010. “Quantification of Live Aboveground Forest Biomass Dynamics with Landsat Time-Series and Field

- Inventory Data: A Comparison of Empirical Modeling Approaches.” *Remote Sensing of Environment* 114 (5): 1053–68. doi:10.1016/j.rse.2009.12.018.
- Prasad, Anantha M., Louis R. Iverson, and Andy Liaw. 2006. “Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction.” *Ecosystems* 9 (2): 181–99. doi:10.1007/s10021-005-0054-1.
- Prinzie, Anita, and Dirk Van den Poel. 2008. “Random Forests for Multiclass Classification: Random MultiNomial Logit.” *Expert Systems with Applications* 34 (3): 1721–32. doi:10.1016/j.eswa.2007.01.029.
- Rengsirikul, Kannika, Assadaporn Kanjanakuha, Yasuyuki Ishii, Kunn Kangvansaichol, Prapa Sripichitt, Vittaya Punsuvon, Pilanee Vaithanomsat, Ganda Nakamanee, and Sayan Tudsri. 2011. “Potential Forage and Biomass Production of Newly Introduced Varieties of *Leucaena* (*Leucaena Leucocephala* (Lam.) de Wit.) in Thailand.” *Grassland Science* 57 (2): 94–100. doi:10.1111/j.1744-697X.2011.00213.x.
- Riaño, D., E. Chuvieco, S. L. Ustin, J. Salas, J. R. Rodríguez-Pérez, L. M. Ribeiro, D. X. Viegas, J. M. Moreno, and H. Fernández. 2007. “Estimation of Shrub Height for Fuel-Type Mapping Combining Airborne LiDAR and Simultaneous Color Infrared Ortho Imaging.” *International Journal of Wildland Fire* 16 (3): 341–48.
- Ritchie, J. C., J. H. Everitt, D. E. Escobar, T. J. Jackson, and M. R. Davis. 2006. “Airborne Laser Measurements of Rangeland Canopy Cover and Distribution.” *Journal of Range Management Archives* 45 (2): 189–93.
- Routledge, R. D. 1990. “When Stepwise Regression Fails: Correlated Variables Some of Which Are Redundant.” *International Journal of Mathematical Education in Science and Technology* 21 (3): 403–10. doi:10.1080/0020739900210308.
- Sala, O. E., and W. K. Lauenroth. 1982. “Small Rainfall Events: An Ecological Role in Semiarid Regions.” *Oecologia* 53 (3): 301–4. doi:10.1007/BF00389004.
- Sankey, Temuulen Tsagaan, and Pamela Bond. 2011. “LiDAR-Based Classification of Sagebrush Community Types.” *Rangeland Ecology & Management* 64 (1): 92–98. doi:10.2111/REM-D-10-00019.1.

- Sasaki, Takeshi, Junichi Imanishi, Wataru Fukui, Fukiko Tokunaga, and Yukihiro Morimoto. 2012. "Fine-Scale Replication and Quantitative Assessment of Forest Vertical Structure Using LiDAR for Forest Avian Habitat Characterization." *Forest Science and Technology* 8 (3): 145–53. doi:10.1080/21580103.2012.704969.
- Shinneman, D.J., R. Arkle, D. Pilliod, N.F. Glenn. 2011. Quantifying and Predicting Fuels and the Effects of Reduction Treatments Along Successional and Invasion Gradients in Sagebrush Habitats. Retrieved from: http://www.firescience.gov/JFSP_funded_project_detail.cfm?jdbid=%24%26ZO9V0%20%20%0A
- Shiple, Lisa A., Tara B. Davila, Nicole J. Thines, and Becky A. Elias. 2006. "Nutritional Requirements and Diet Choices of the Pygmy Rabbit (*Brachylagus Idahoensis*): A Sagebrush Specialist." *Journal of Chemical Ecology* 32 (11): 2455–74. doi:10.1007/s10886-006-9156-2.
- Shrestha, Gyami, and Peter D. Stahl. 2008. "Carbon Accumulation and Storage in Semi-Arid Sagebrush Steppe: Effects of Long-Term Grazing Exclusion." *Agriculture, Ecosystems & Environment* 125 (1–4): 173–81. doi:10.1016/j.agee.2007.12.007.
- Spaete, Lucas P., Nancy F. Glenn, Dewayne R. Derryberry, Temuulen T. Sankey, Jessica J. Mitchell, and Stuart P. Hardegree. 2011. "Vegetation and Slope Effects on Accuracy of a LiDAR-Derived DEM in the Sagebrush Steppe." *Remote Sensing Letters* 2 (4): 317–26. doi:10.1080/01431161.2010.515267.
- Stokes, C. J., A. Ash, and S. W. Howden. 2008. "Climate Change Impacts on Australian Rangelands." *Rangelands Archives* 30 (3): 40–45.
- Storch, Ilse. 2007. "Conservation Status of Grouse Worldwide: An Update." *Wildlife Biology* 13 (sp1): 5–12. doi:10.2981/0909-6396(2007)13[5:CSOGWA]2.0.CO;2.
- Streutker, David R., and Nancy F. Glenn. 2006. "LiDAR Measurement of Sagebrush Steppe Vegetation Heights." *Remote Sensing of Environment* 102 (1–2): 135–45. doi:10.1016/j.rse.2006.02.011.

- Su, J.G. and Bork, E.W., 2007. Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Applied Vegetation Science*, 10(3), pp.407-416.
- Tueller, Paul T. 1992. "Overview of Remote Sensing for Range Management." *Geocarto International* 7 (1): 5–10. doi:10.1080/10106049209354346.
- USDI Bureau of Land Management. (2008). *Snake River Birds of Prey National Conservation Area Proposed Resource Management Plan and Final Environmental Impact Statement*, Boise, ID
- Vauhkonen, Jari, Ilkka Korpela, Matti Maltamo, and Timo Tokola. 2010. "Imputation of Single-Tree Attributes Using Airborne Laser Scanning-Based Height, Intensity, and Alpha Shape Metrics." *Remote Sensing of Environment* 114 (6): 1263–76. doi:10.1016/j.rse.2010.01.016.
- Vierling, Lee A., Yanyin Xu, Jan U.H. Eitel, and John S. Oldow. 2012. "Shrub Characterization Using Terrestrial Laser Scanning and Implications for Airborne LiDAR Assessment." *Canadian Journal of Remote Sensing* 38 (06): 709–22. doi:10.5589/m12-057.
- Vincenzi, Simone, Matteo Zucchetta, Piero Franzoi, Michele Pellizzato, Fabio Pranovi, Giulio A. De Leo, and Patrizia Torricelli. 2011. "Application of a Random Forest Algorithm to Predict Spatial Distribution of the Potential Yield of Ruditapes Philippinarum in the Venice Lagoon, Italy." *Ecological Modelling* 222 (8): 1471–78. doi:10.1016/j.ecolmodel.2011.02.007.
- Waite, R.B., 1994. The application of visual estimation procedures for monitoring pasture yield and composition in exclosures and small plots. *Tropical Grasslands*, 28, pp.38-38.
- Wang, C., and N. F. Glenn. 2009a. "Estimation of Fire Severity Using Pre- and Post-Fire LiDAR Data in Sagebrush Steppe Rangelands." *International Journal of Wildland Fire* 18 (7): 848–56.

- Wang, Cheng, and N.F. Glenn. 2009b. "Integrating LiDAR Intensity and Elevation Data for Terrain Characterization in a Forested Area." *IEEE Geoscience and Remote Sensing Letters* 6 (3): 463–66. doi:10.1109/LGRS.2009.2016986.
- Western Region Climate Center. (WRCC) 2012. Swan Falls Power House, Idaho, Period of Record General Climate Summary. Retrieved from: <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?id8928>
- Wilcox, Bradford P. 2010. "Transformative Ecosystem Change and Ecohydrology: Ushering in a New Era for Watershed Management." *Ecohydrology* 3 (1): 126–30. doi:10.1002/eco.104.
- Wilson, Adam M., John A. Silander, Alan Gelfand, and Jonathan H. Glenn. 2011. "Scaling up: Linking Field Data and Remote Sensing with a Hierarchical Model." *International Journal of Geographical Information Science* 25 (3): 509–21. doi:10.1080/13658816.2010.522779.
- Wisdom, M.J., M.M. Rowland, and R.J. Tausch. 2005. "Effective Management Strategies for Sage-Grouse and Sagebrush: A Question of Triage?" In *Transactions of the 70th North American Wildlife and Natural Resources Conference*, edited by R.D. Sparrowe and L.H. Carpenter, 70:206. Arlington, Virginia, USA: Wildlife Management Institute. http://www.fs.fed.us/rm/pubs_other/rmrs_2005_wisdom_m001.pdf.
- Yunfei B, Guoping L., Chunxiang C, Xiaowen L, Hao Z, Qisheng H, Linyan B, Chaoyi C. 2008. Classification of lidar point cloud and generation of dtm from lidar height and intensity data in forested area. In: International society for photogrammetry and remote sensing congress, pp 313- 318.
- Yoon, Jong-Suk, Jung-Il Shin, and Kyu-Sung Lee. 2008. "Land Cover Characteristics of Airborne LiDAR Intensity Data: A Case Study." *IEEE Geoscience and Remote Sensing Letters* 5 (4): 801–5. doi:10.1109/LGRS.2008.2000754.
- Zandler, H., A. Brenning, and C. Samimi. 2015. "Quantifying Dwarf Shrub Biomass in an Arid Environment: Comparing Empirical Methods in a High Dimensional

Setting.” *Remote Sensing of Environment* 158 (March): 140–55.
doi:10.1016/j.rse.2014.11.007.

Zheng, Guang, L.M. Moskal, and Soo-Hyung Kim. 2013. “Retrieval of Effective Leaf Area Index in Heterogeneous Forests With Terrestrial Laser Scanning.” *IEEE Transactions on Geoscience and Remote Sensing* 51 (2): 777–86.
doi:10.1109/TGRS.2012.2205003.

Zolkos, S. G., S. J. Goetz, and R. Dubayah. 2013. “A Meta-Analysis of Terrestrial Aboveground Biomass Estimation Using Lidar Remote Sensing.” *Remote Sensing of Environment* 128 (January): 289–98. doi:10.1016/j.rse.2012.10.017.

3 CHAPTER THREE: MANSUCRIPT TWO - IMPUTATION OF VEGETATION
BIOMASS AND COVER FOR LARGE SCALE MAPPING IN A SEMI-ARID
RANGELAND USING AIRBORNE LIDAR, LANDSAT 8 AND A MACHINE
LEARNING ALGORITHM

Abstract

Remote sensing based quantification of semi-arid rangeland vegetation provides the large scale observations necessary for monitoring native plant distribution, estimating fuel loads and measuring carbon storage. Improved signal to noise ratio and radiometric resolution of recent satellite imagery and fine scale 3-dimensional information from lidar provides opportunities for refined measurements of vegetation structure. We leverage a large number of Landsat 8 and lidar-based metrics for prediction of biomass and cover in the semi-arid rangeland of the western United States. Time-series Landsat 8 images were used to develop 20 ratio-based vegetation indices. Similarly, 35 vegetation metrics, including metrics based on numerical values (e.g. elevation, canopy height) and on density of points (e.g. canopy density) were developed from airborne lidar point clouds. These vegetation indices were trained and linked to *in-situ* measurements ($n = 141$) with the Random forest regression to impute semi-arid vegetation biomass and cover, resulting in a large scale map of these vegetation characteristics. We also validated our model with an independent data-set ($n = 44$), explaining up to 63% of variation in cover and 53% in biomass of shrub. Forty six of the *in-situ* plots were also used in a model to compare performance of lidar and Landsat data. Our results demonstrate that Landsat performs

better in estimating vegetation cover ($R^2 = 0.75$) whereas lidar performs better in estimating shrub biomass ($R^2 = 0.75$). We also examined the effect of topographic variables (e.g. slope, aspect and elevation) on the biomass and cover estimates and found no significant relationship. This study demonstrates new opportunities of using Landsat 8 with established lidar approaches to better quantify vegetation in semiarid ecosystems.

3.1 Introduction

Native plants in arid and semi-arid regions such as in western North America, Africa, Asia and Australia, are threatened by exotic annuals and perennials (Purdie and Slatyer 1976; Hobbs and Atkins 1990; Anderson and Inouye 2001; Brooks et al. 2004; Liu et al. 2006; Hill et al. 2005; Didham et al. 2007; van Wilgen et al. 2008; Davies and Nafus 2013). In North America, sagebrush-steppe constitute the largest temperate semi-desert and yet are imperiled and critically endangered ecosystems (Barbour and Billings 2000; Miller et al. 2011). The dramatic decrease of sagebrush (*Artemisia tridentata*) can be largely attributed to increased fire frequency due to the spread of invasive nonnative species like cheatgrass (*Bromus tectorum*), medusahead (*Taeniatherum caputmedusae*) and juniper (*Juniperus* spp.) (Knick 1999; Wisdom et al. 2005; Miller and Rose 2006; Miller et al. 2011). The change in sagebrush results in a decline of already endangered animals such as the greater sage grouse and pygmy rabbits (Connelly et al. 2004; Shipley et al. 2006; Storch 2007), a decline of organic matter, evapotranspiration and hydraulic lift in soil (Branson et al. 1976; Heitschmidt and Stuth 1991; Wilcox 2010; Cardon et al. 2013) and a loss in carbon sequestration (Angell et al. 2001; Shrestha and Stahl 2008). Accurate and cost-effective characterization of the spatial distribution of native and non-native plants and quantification of their biomass in drylands are crucial for measurements

of carbon storage, estimation of fire fuel loads, developing restoration strategies and ensuring sustainability of native ecosystems.

Various methods are available for *in-situ* measurements of aboveground biomass and percent cover of vegetation in semi-arid ecosystems. Some of the most common methods include harvesting (Sims and Singh 1978, Sala and Lauenroth 1982), clip-and-weigh (Bonham 2013), visual estimation (Waite 1994), and point-intercept sampling (Clark et al. 2008). Remote sensing methods are complementary to these traditional methods as remote sensing can provide multi-scale contiguous estimates over space and through time (Wilson et al. 2011; Lu 2006; Zolkos et al. 2013). Multispectral, hyperspectral, radar and airborne lidar are used extensively in studies to develop attributes that can be then statistically related to vegetation characteristics (Lu and Chen 2012; Mitchard et al. 2012; Swatantran et al. 2011; Asner et al. 2010). Recent studies have also up-scaled vegetation characteristics to regional, national and even global scales using remote sensing data, resulting in high spatial coverage with low uncertainties (Lefsky 2010; Wilson et al. 2011; Mitchell et al. 2015; Li et al. 2015). Heterogeneity in cover types and land surfaces, however, challenges the upscaling when performed at very large scales (Hufkens et al. 2008; Wu and Li 2009).

Selection of suitable predictor variables derived from remote sensing data is a critical step for developing a vegetation biomass or percent cover model. In passive optical remote sensing, spectral indices must be able to empirically correlate with vegetation attributes and be able to differentiate vegetation from soil features (Todd et al. 1998). Previous studies have shown that there is a relationship between vegetation cover and biomass with vegetation indices (Purevdorj et al. 1998; Avitabile et al. 2012). It has

been widely demonstrated that vegetation indices are more sensitive than individual spectral bands to vegetation parameters (Jackson and Huete 1991; Huete et al. 2002; Price et al. 2002; Glenn et al. 2008; Yang et al. 2012). Robinson and Novelly (2009) point out three issues to be considered when choosing appropriate indices for arid and semi-arid rangelands with sparse vegetation and low leaf area index: i) low reflectivity of vegetation in the infrared portion of spectrum; ii) greater influence of background soil reflectance and; iii) limitation in performance of indices using the difference in red and near infrared bands due to the senescence of grasses and shrubs. Remote sensing indices used in characterizing vegetation in semi-arid rangelands can be characterized into three types. The first type of indices are 'Simple ratio based' that considers that all vegetation isolines converge at a single point (e.g. NDVI, SVI and VCI). The second type is 'soil adjusted vegetation indices' that account for soil brightness variations (e.g. SAVI, GSAVI and SATVI) and the third type is 'perpendicular vegetation indices' that consider vegetation isolines parallel to the soil line (e.g. BI, GVI and WI) (Zandler et al. 2015; Basso et al. 2004). Algorithms such as the Normalized Difference Vegetation Index (NDVI) have been widely accepted and empirically correlated to structural parameters such as Leaf Area Index (LAI) (Vazirabad et al. 2011). The tasseled cap brightness and wetness indices are responsive to variations in total reflectance and sensitive to soil-plant moisture respectively, which makes them useful for estimating biomass and cover (Crist 1983; Todd et al. 1998). Similarly, soil adjusted indices have been shown to be important in drylands as they account for soil-vegetation spectral mixing caused by sparse vegetation cover (Eitel et al. 2009; Veraverbeke et al. 2012). The inherent challenges in semi-arid ecosystems, including the short stature, sparse vegetation, considerable bare

ground and multiple species with similar characteristics, are not yet explored with data from the relatively new Landsat 8 Operational Land Imager (OLI) (Glenn et al. in press). As the latest member of the Landsat family of remote sensing satellites, Landsat 8 uses a push-broom scanner which receives stronger signals and improved signal-to-noise ratio performance (Li et al. 2013).

One of the major limitations of Landsat is that the data saturation in sites with high biomass and penetrable canopies in low cover areas generate large uncertainties (Lu 2005; Sivanpillai et al. 2009; Frank and Tweddale 2006). However, lidar can overcome this shortcoming because of its ability to obtain range and orientation information by capturing three-dimensional structural data (Su and Bork 2007). Lidar has been extensively used in forest applications for characterizing canopy structure or measuring volume, height and biomass (Lefsky et al. 2002; Zimble et al. 2003; Andersen et al. 2005; Hall et al. 2005; Lin et al. 2012). Vegetation characteristics of shrubs in rangelands using lidar has also been of interest to several studies (Ritchie et al. 2006; Streutker and Glenn 2006; Su and Bork 2007; Glenn et al. 2011). But studies using lidar are typically confined to the areal extent of lidar, which is generally smaller than an individual Landsat scene. Hence, there has also been effort on deriving metrics from both lidar and spectral remote sensing and relating this information with vegetation features using statistical models (Erdody and Moskal 2010; Hartfield et al. 2011; Swatantran et al. 2011; Estornell et al. 2012; Lu and Chen 2012; Naidoo et al. 2012). Ancillary topographic data, fire events and climatic factors have also been incorporated into the modeling process to understand their influence on biomass and percent cover (Watson et al. 1998; Liu et al. 2003; Osumi et al. 2003; Stokes et al. 2008; Balch et al. 2013). Specifically, topographic variables like

slope, aspect, elevation or texture have been shown to correlate with vegetation characteristics (Sternberg and Shoshany 2001; Powell et al. 2010; Avitabile et al. 2012).

In remote sensing applications, it is not uncommon to see a large number of predictor variables with a relatively modest number of ground-truth observations. This creates a high-dimensional problem which may lead to over-fitting of models (Zandler et al. 2015). Regression analysis is widely adopted to relate field and remotely sensed data but identifying suitable variables for meaningful correlation is critical in regression as some variables are weakly related to the ground data and strongly related to each other (Mutanga et al. 2012; Montgomery et al. 2012; Fernandes and Leblanc 2005; Ahmed et al. 2015). Furthermore, regression analysis involves assumptions of normality and homogeneity which may not always be satisfied with the available datasets (Montgomery et al. 2012).

Alternatively, non-parametric machine learning approaches like random forest (RF) are gaining considerable attention to select suitable variables and modeling due to their versatility and computational accuracy in high-dimensional, non-normal and non-linear remote sensing applications (Mitchell et al. 2013; Guan et al. 2012; Pal 2005). RF is an improvement over classification and regression trees (CART). RF bootstraps samples to iteratively construct a large number of decision trees each grown with a randomized subset of predictors (Breiman 2001). RF models can be used in either classification or regression approaches. These 'trees' cast a unit vote for the most popular class to classify an input vector in a RF classification (Pal 2005; Breiman 1999). In regression mode, RF takes the average of the 'trees' to make a prediction. As the response variables (biomass and cover) are numerical and continuous, we confine our

attention to regression random forest models (Vincenzi et al. 2011) rather than classification. The RF model doesn't require assumptions about the relationship between dependent and independent variables and is well suited for analyzing complex non-linear and possibly hierarchical interactions in large data sets (Olden et al. 2008; Ahmed et al. 2015). RF grows a large number of trees which does not over fit the data and the random predictor selection keeps biases low. RF also has an internal cross-validation mechanism and thus provides better models for prediction (Prasad et al. 2006; Naidoo et al. 2012).

In this study, we investigated the utility and limitation of Landsat and lidar derived metrics in mapping aboveground vegetation biomass and cover in a semi-arid rangeland in a National Conservation Area (NCA) located in western United States. Different random forest models were developed and the results were compared with *in-situ* data independently for Landsat and lidar. We used spectral indices as a proxy to successfully extend the biomass and cover model for the whole NCA. The study presented here reveals the capabilities and weaknesses of both Landsat and lidar and identifies the coupling of multiple sensors for improved modeling results.

3.2 Study Area and Data

3.2.1 Study Site

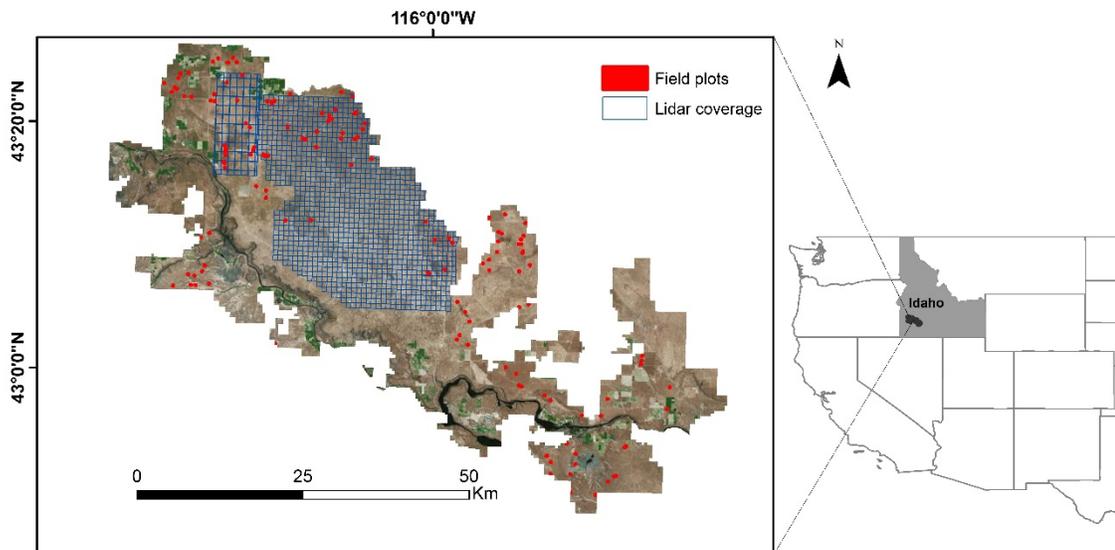


Figure 3.1 Location of field plots and lidar coverage in the study site.

The research was carried out in a semiarid shrub-steppe rangeland once dominated by big sagebrush located within the Morley Nelson Snake River Birds of Prey National Conservation Area (NCA). The NCA encompasses about 242,800 hectares of the Snake River Plain ecoregion in southwestern Idaho, USA (**Fig. 3. 1**). It also contains other native species including shadscale (*Altriplex confertifolia*), winterfat (*Ceratoides lanata*), budsage (*Artemisia spinescen*), rabbitbrush (*Chrysothamnus viscidiflorus*) and rapidly invading annual exotic like cheatgrass (*Bromus tectorum*). In an average year, the NCA receives 20 cm precipitation, 74 days with a high temperature greater than 32° C, and 98 days with a low temperature below 0° C (WRCC, 2012). The native vegetation assemblage is composed of an understory of biological crusts and sparse native bunchgrass (*Festuca idahoensis*), overlain by an open canopy of shrubs ranging up to 1.5 m tall (Anderson, 2014). Since 1980, over half of the NCA has burned resulting in a

mosaic of plant communities, with compositions spanning a gradient between intact native shrublands, shrublands degraded by biological invasion and wildfire, and grasslands where native plants have been fully replaced by cheatgrass and other invasive annuals. Currently 37% or less of the NCA retains an intact native shrubland community (USDI, 2008).

3.2.2 Acquiring Landsat 8/OLI and Ancillary Data

A time series of three surface reflectance Landsat 8 Operational Land Imager (OLI) images from April 11, June 30 and October 4 2013 in path/row 41/30 were acquired covering the field site. These specific dates were chosen to cover the growing season and correspond to 2013 field sampling efforts. Field sampling also occurred in 2012, but Landsat 8 data were not available for that period. The cloud cover over the 2013 images were < 1.5 %. Of the 11 bands of Landsat 8, 6 bands (blue, green, red, NIR, SWIR 1, and SWIR 2) of OLI were used to produce vegetation indices at 30m pixel scale for each image date.

Slope, aspect and elevation were calculated from an ASTER GDEM (product of METI and NASA) to include topographic attributes as possible proxies in the model and to account for site suitability and background soil conditions (Zandler et al. 2015).

3.2.3 Field Sampling

In the summer of 2012 and 2013, a total of 141 1-ha plots were established at locations throughout the western NCA by the United States Geological Survey, Forest and Rangeland Ecosystem Science Center (USGS FRESC) (Shinneman et al. 2011). A stratified random sampling approach was used to select the locations of these plots, and the corners of each plot were precisely located using a survey-grade Global Navigation

Satellite System (GNSS). The sites were selected considering the accessibility and to capture a variety of plant community compositions. The sampling design included a 3 x 3 grid of nine 1 m² subplots with 25 m spacing between them inside a 1-ha field plot, for a total of 1269 subplots. In our analysis, we modified the 1-ha plot to utilize the inner 70 m x 70 m area because the Landsat OLI sensor has a pixel size of 30m and an accuracy of 12 m resulting in a calculated minimum size of plot of 54 m (USGS, 2013). This also provided an additional 6 m to provide for GPS horizontal deviation and 10 m as a buffer distance to minimize side effects (Zandler et al. 2015).

Aboveground vegetation within the subplots were destructively sampled and classified as herbaceous or shrub. All forbs, herbs and grasses (native and non-native) were lumped into a single herbaceous category (hereafter referred to as herb). Litter was also included in the herb class as it was difficult to separate litter from grass in the field. If shrubs were too bulky to be harvested efficiently, a portion was collected for reference and weighing and the number of equivalent portions remaining in the quadrat was estimated (Anderson 2014). The harvested vegetation was oven dried, weighed and biomass at each plot was estimated as the average from the nine subplots by herb and shrub class. SamplePoint Software (Booth et al. 2006) was used for photo plot analysis and to produce cover estimates of herb and shrub for the 1 m subplots which were later averaged to produce average cover for the 70 m x 70 m plots. We included grass, forbs and litter in the herbaceous category as it was difficult to separate litter and grass in the field.

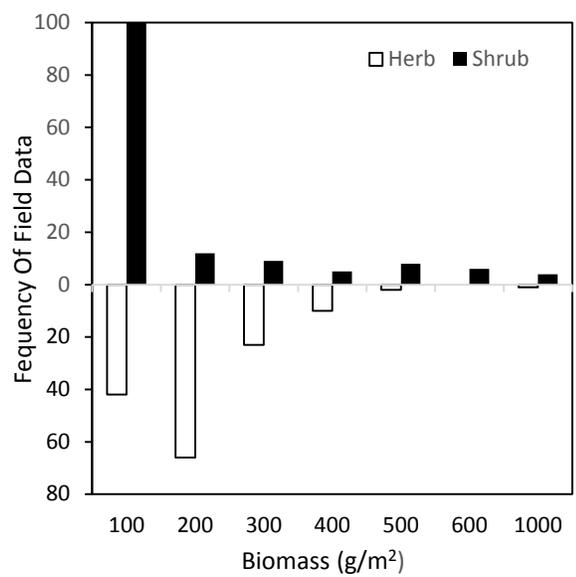
Based on the 141 plots, herbaceous cover ranged from 10 to 100% and shrub cover from 0 to 48%. Similarly, the herbaceous class had a mean biomass of ~ 157 g/m²

and shrub had a mean biomass of $\sim 118 \text{ g/m}^2$ (**Table 3. 1**). The distribution of the field data was skewed towards smaller biomass values (**Fig. 3. 2a**). With regards to cover, the distribution of shrub was skewed towards smaller values whereas herb was skewed towards higher values (**Fig. 3. 2b**). Almost one third of the plots didn't have any shrub cover.

Table 3.1 Summary of vegetation cover and biomass based on 1 ha field data (n=141).

	Herb Cover (%)	Shrub Cover (%)	Herb Biomass (g/m^2)	Shrub Biomass (g/m^2)
Minimum	10	0	5	0
Maximum	100	48	804	954
Mean \pm SE	72 ± 2	7 ± 1	157 ± 9	118 ± 16
Standard Deviation	23	11	104	197

(a)



(b)

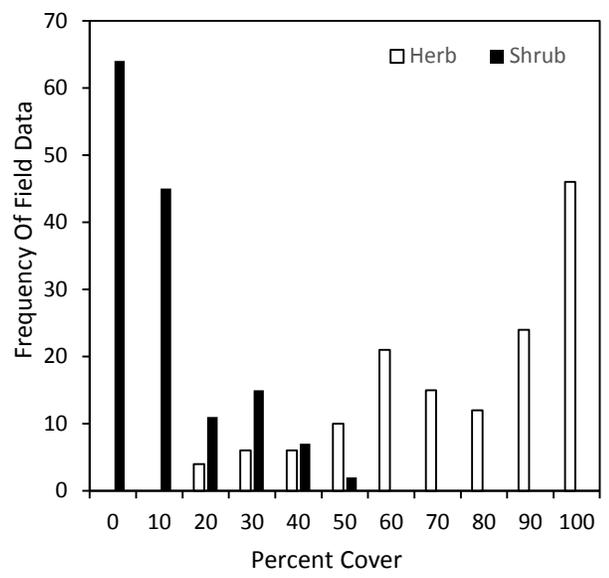


Figure 3.2 Frequency distribution of the in-situ (a) biomass and (b) percent cover of herb and shrub (n=144).

3.2.4 Airborne Lidar Data Acquisitions

Discrete small footprint lidar data were collected over 65,194 hectares in 2012 and 9,970 hectares in 2013, with a ALS60 system (Leica Geosystems, Heerbrugg, Switzerland) operated by Watershed Sciences (Corvallis/Portland, OR), with a point density of ~ 8 points per m^2 and a vertical accuracy of ~ 3 cm. The lidar system acquired approximately $\geq 148,000$ laser pulses per second and was flown at 1,500 meters above ground level, with a scan angle of $48^\circ (\pm 12^\circ)$ from nadir (field of view). An opposing flight line side-lap of $\geq 50\%$ (i.e. 100% overlap) was maintained to reduce shadowing and increase point density. The absolute vertical accuracy (RMSE_z) was about 0.03 m and the relative accuracy was about 0.024 m. The vertical accuracy was primarily assessed from ground check points on open, bare earth surface with slope, by the vendor. Of the 141 field plots, 46 plots were spatially nested within the lidar footprints.

3.3 Methodology

3.3.1 Spectral Vegetation Indices (VI)

In addition to the six spectral bands of the OLI sensor, we developed 20 spectral indices from the OLI spectral bands for each plot (**Table 3. 2**). For all soil adjusted indices, we used a 0.25 soil correction factor (Glenn et al. in press). The spectral bands and vegetation indices were used in our statistical model to relate to the field data.

Table 3.2 Overview of the spectral indices from Landsat 8 OLI used in this study.

Indices	Formula	Reference
GI/ SVI	NIR / Red	Sivanpillai et al. (2009) Phua and Saito (2003)
VCI	MIR / NIR	Sivanpillai et al. (2009)
NDVI	NIR - Red / NIR + Red	Tucker et al. (1983)
TVI	$\sqrt{(\text{NDVI}+0.5)}$	Phua and Saito (2003) Nellis and Briggs (1992)
DVI	SWIR1- SWIR2	Phua and Saito (2003)
MIRI	SWIR2 - Red/ SWIR2 + Red	Sivanpillai et al. (2009)
Green + Red	-	Wallace and Thomas (1998)
STVI-1	SWIR2 x Red/NIR	Thenkabail and Ward (1994)
MSI	SWIR/NIR	Hunt Jr. and Rock (1989)
MSI2	SWIR2/NIR	
NDWI	$(\text{NIR}-\text{SWIR})/(\text{NIR}+\text{SWIR})$	Gao (1996)
NDWI2	$(\text{NIR}-\text{SWIR2})/(\text{NIR}+\text{SWIR2})$	
NBR	$(\text{SWIR1}-\text{SWIR2})/(\text{SWIR1}+\text{SWIR2})$	Escuin et al. (2008)
GVI	$-0.2941 * \text{Blue} - 0.243 * \text{Green} - 0.5424 * \text{Red} + 0.7276 * \text{NIR} - 0.0713 * \text{SWIR1} - 0.1608 * \text{SWIR2}$	Kauth and Thomas (1976) Baig et al. (2014)
BI	$0.3029 * \text{Blue} + 0.2786 * \text{Green} + 0.4733 * \text{Red} + 0.5599 * \text{NIR} + 0.508 * \text{SWIR1} + 0.1872 * \text{SWIR2}$	Kauth and Thomas (1976)

		Baig et al. (2014)
WI	$0.1511 * \text{Blue} + 0.1973 * \text{Green} + 0.3283 * \text{Red} + 0.3407 * \text{NIR} - 0.7117 * \text{SWIR1} - 0.4559 * \text{SWIR2}$	Kauth and Thomas (1976) Baig et al. (2014)
SAVI	$[(\text{NIR1} - \text{Red}) / (\text{NIR1} + \text{Red} + \text{L})] \times (1 + \text{L})$ L=Soil Correction factor; L=0.25	Huete (1988)
GSAVI	$[(\text{NIR} - \text{green}) / (\text{NIR} + \text{Green} + \text{L})] \times (1 + \text{L})$ L=0.25	Sripada et al. (2006)
MSAVI	$[2 \times \text{NIR} + 1 - \sqrt{((2 \times \text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{Red}))}] / 2$	Qi et al. (1994)
SATVI	$[(\text{NIR1} - \text{Red}) / (\text{NIR1} + \text{Red} + \text{L})] \times (1 + \text{L}) - \text{SWIR1} / 2$; L=0.25	Marsett and Jiaguo (2006)

3.3.2 Lidar Data and Derivative Processing

Lidar point cloud data were buffered and height filtered using the ‘BCAL Lidar Tools’ developed for semiarid vegetation (<http://bcal.boisestate.edu/tools/lidar>; Streutker and Glenn 2006.. Height filtering classifies lidar points into ground and vegetation points. The filtering was performed using a 5 m canopy spacing, a 5 cm ground threshold, nearest neighbor interpolation and 40 iterations. The resulting vegetation points were rasterized into multiband vegetation metrics at one meter raster resolution, calculated directly on the binned point clouds (**Table 3. 3**). These 35 metrics were developed with the intention to use them as a proxy for vegetation characteristics such as density, biomass, volume, cover etc. A description of lidar metrics and their use in biomass modeling can be found in (Dhakal et al. 2016).

Table 3.3 Lidar metrics and their descriptions used in the analysis.

Lidar Metrics	Description
Minimum Height (H_{\min})	The minimum of all height points within each pixel
Maximum Height (H_{\max})	The maximum of all height points within each pixel
Height Range (H_{range})	The difference of maximum and minimum of all height points within each pixel
Mean Height (H_{mean})	The average of all height points within each pixel
Median Absolute Deviation from Median Height (H_{MAD})	The MAD value of all height points within each pixel $H_{\text{MAD}} = 1.4826 \times \text{median}(\text{height} - \text{median height})$
Mean Absolute Deviation from Mean Height (H_{AAD})	The AAD value of all height points within each pixel $H_{\text{AAD}} = \text{mean}(\text{height} - \text{mean height})$
Height Variance (H_{var})	The variance of all height points within each pixel
Height St. Deviation (H_{std})	The standard deviation of all height points within each pixel This is also called 'absolute vegetation roughness'
Height Skewness (H_{skew})	The skewness of all height points within each pixel
Height Kurtosis (H_{kurt})	The kurtosis of all height points within each pixel
Interquartile Range (H_{IQR}) of Height	The IQR of all height points within each pixel $H_{\text{IQR}} = Q_{75} - Q_{25}$, where Q_x is x^{th} percentile
Height Coefficient of Variation (H_{CV})	The coefficient of variation of all height points within each pixel
Height Percentiles (H_5 , H_{10} , H_{25} etc.)	The 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of all height points within each pixel
Number of Lidar Returns	The total number of all points within each pixel
Number of Lidar Vegetation Returns (nV)	The total number of all the points within each pixel that are above the specified crown threshold value (CT)
Number of Lidar Ground Returns (nG)	The total number of all the points within each pixel that are below the specified ground threshold value (GT)

Total Vegetation Density (Veg_density)	The percent ratio of vegetation returns and ground returns within each pixel. Density = $nV/nG*100$
Vegetation Cover (Veg_cov)	The percent ratio of vegetation returns (nV) and total returns within each pixel
Percentage of Ground Return (pG)	Percent of points within each pixel that are below the specified Ground Threshold
Percent of Vegetation in Height Range (pH ₁ , pH _{2.5} , pH ₁₀ etc.)	Percent of vegetation in height ranges 0-1m, 1-2.5m, 2.5-10m, 10-20m, 20-30m, and >30m within each pixel Percent of Vegetation = Number of vegetation returns in the range/Total vegetation returns
Canopy Relief Ratio (CRR)	Canopy relief ratio of points within each pixel Canopy relief ratio = $((H_{\text{mean}} - H_{\text{min}})) / ((H_{\text{max}} - H_{\text{min}}))$
Texture of Heights (H _{text})	Texture of height of points within each pixel. Texture = St. Dev. (Height > Ground Threshold and Height < Crown Threshold)
Foliage Height Diversity (FHD _{all})- All points	Foliage arrangement in the vertical direction $FHD_{\text{all}} = - \sum p_i * \ln p_i$ where p_i is the proportion of horizontal foliage coverage in the i th layer to the sum of the foliage coverage of all the layers
Foliage Height Diversity- Points above ground threshold (FHD _{GT})	FHD calculated only from the points above the ground threshold

3.3.3 Modeling Approach

For the Landsat analyses we randomly selected 97 field plots for calibration and used the remaining 44 plots for independent validation. We first modeled the relationship between biomass and cover with remote sensing variables (spectral bands, spectral indices and topographic data) using a RF regression. The biomass was subcategorized into total, shrub and herb biomass; cover was categorized into shrub and herb cover and

individual models were developed for each category. We used the SPM Suite (Salford Predictive Modeler Software Suite version 7, Salford Systems, San Diego, CA) for the implementation of the RF algorithm. Each RF regression run generated 2000 trees and the maximum number of variables considered per node was kept equal to the square root of the number of variables for the run (Breiman 1999; Breiman 2001). All predictor variables were used to perform the initial RF run and each of them were ranked based on their predictive power using the ‘Standard Method’. In each ‘tree’ in the ‘forest’, a variable was tested by first scrambling (substituting) its values and then measuring the decline of accuracy in the model caused by the substitution. This means if a variable substituted with incorrect values can predict the target accurately, then the variable has no relevance to predicting the outcome and hence is assigned a low score (SPM user guide, 2013). A backward feature elimination method was used for the best variable selection method, where lowest performing variables were iteratively removed until the best model was obtained. The best models for all five categories were determined based on the highest coefficient of determination (R^2), lowest root-mean-square error (RMSE) and maintaining model parsimony (number of predictor variables were kept as low as possible). The variable selection was done not only to reduce the explanatory variables but also to understand the most suitable explanatory variables to estimate biomass and cover (Ismail et al. 2010).

The best variables selected for five categories (shrub biomass, herb biomass, total biomass, shrub cover and herb cover) were then used in Nearest Neighborhood (NN) Imputation across 30 m pixels to produce a contiguous map for each category separately. This imputed map was validated using the remaining independent plots ($n = 44$). The

imputation was performed in the R statistical computing environment (R Development Core Team 2013) using the *yaimpute* package (Crookston and Finley 2008; Hudak et al. 2008). In NN imputation, the estimates for the attributes of interest (e.g. biomass, cover) are produced as weighted averages of the attributes of the reference observation (e.g. spectral information). But unlike interpolation, imputation reference records should cover the entire phenomenon of interest or field site to make an accurate estimation. NN imputation methods can use different distance metrics to determine the similarity between target and reference records, including Eculidean, Mahalanobis, Minkowski, fuzzy etc. (Eskelson et al. 2009). For this study, we used a proximity matrix obtained from the RF method (Crookston and Finley 2008).

In order to compare the performance of Landsat 8 OLI with lidar, we ran two additional RF analyses with a) OLI and topographic metrics and; b) lidar metrics. These analyses were performed over the area of the lidar footprints and corresponding 46 field plots. The analyses were performed independently for five categories: shrub cover, herb cover, total biomass, shrub biomass and herb biomass and the best variables for each category were determined.

3.4 Results

3.4.1 Calibration and Validation of OLI

The RF model indicated that the reflectance bands and indices developed from OLI were better in predicting vegetation cover than biomass. The model explained 63% and 69% variance in shrub and herb cover with a RMSE of 7% and 13% respectively (**Table 3. 4**). However, shrub and herb biomass were predicted with a R^2 of 0.60 and 0.49, and RMSE of 126 g/m² and 65 g/m², respectively. Regarding the variable

importance, the spectral information and indices acquired from images in summer and fall were more dominant than those acquired in spring. Soil adjusted indices and tasseled cap indices were among the important variables in each RF run. However, we did not find any statistically significant relationship of topographic variables with biomass or cover in the study site.

In order to assess the reliability of the model, the performance was tested against the independent dataset that was excluded from the model development and included randomly chosen plots ($n=44$). The validation model had a comparatively lower explanatory power than the calibration model. The results indicated our model explained 63% of variation in herb cover and only 30% of herb biomass, 44% of shrub cover, 53% of shrub biomass and 37% of total biomass (**Table 3. 4**).

Table 3.4 Random forest results for calibration and validation model of OLI. The RMSE values of cover predictions are in units of % and biomass predictions are in g/m^2 . The variables are labeled by respective date and metric from Table 2. 2.

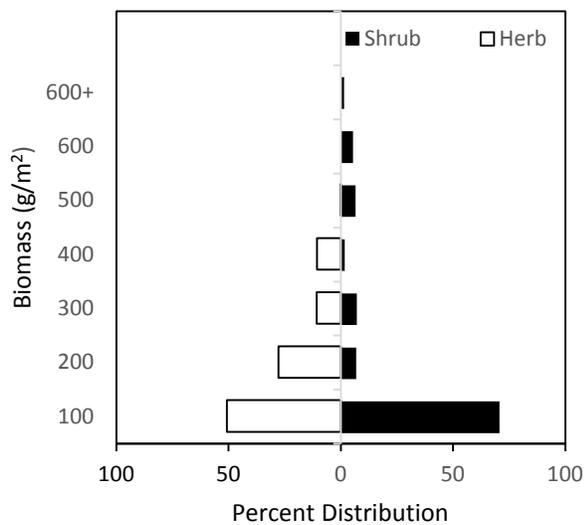
	Calibration ($n=97$)			Validation ($n=44$)	
	R^2	RMSE (g/m^2)	Variables	R^2	RMSE (g/m^2)
Shrub Cover	0.63	7	June30 SATVI, June30 GVI, Oct4 SAVI, Oct4 MSAVI, June30 NBR	0.44	8
Herbaceous Cover	0.69	13	Oct4 MIRI, June30 MSI, Oct4 Green, June30 SATVI	0.63	16
Total Biomass	0.54	147	June30 SATVI, June30 GVI, Oct4 NIR, Oct4	0.37	158

			MSAVI, Oct4 Red, Oct4 SAVI		
Shrub Biomass	0.60	126	June30 SATVI, June30 GVI, Oct4 MSAVI, June30 NDWI, Oct4 GSAVI	0.53	128
Herb Biomass	0.49	65	Apr11 NIR, June30 MSI, June30 NIR, Oct4 NIR	0.3	64

3.4.2 Biomass and Cover Maps

We scaled our RF models of biomass and cover to create spatially explicit contiguous maps of biomass and percent cover (**Fig. 3. 4**). The imputed map covered the entire NCA (~ 75,164 hectares) using imputation. The imputed maps are relevant for the year and resolution in which we created the models (2013, 30 m pixel resolution). Based on the imputation, the NCA contained ~ 344925 kg of herbaceous biomass and ~ 313420 kg of shrub biomass in 2013.

(a)



(b)

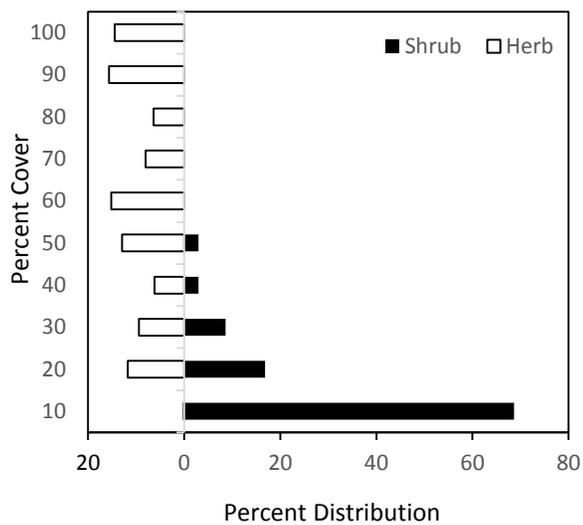
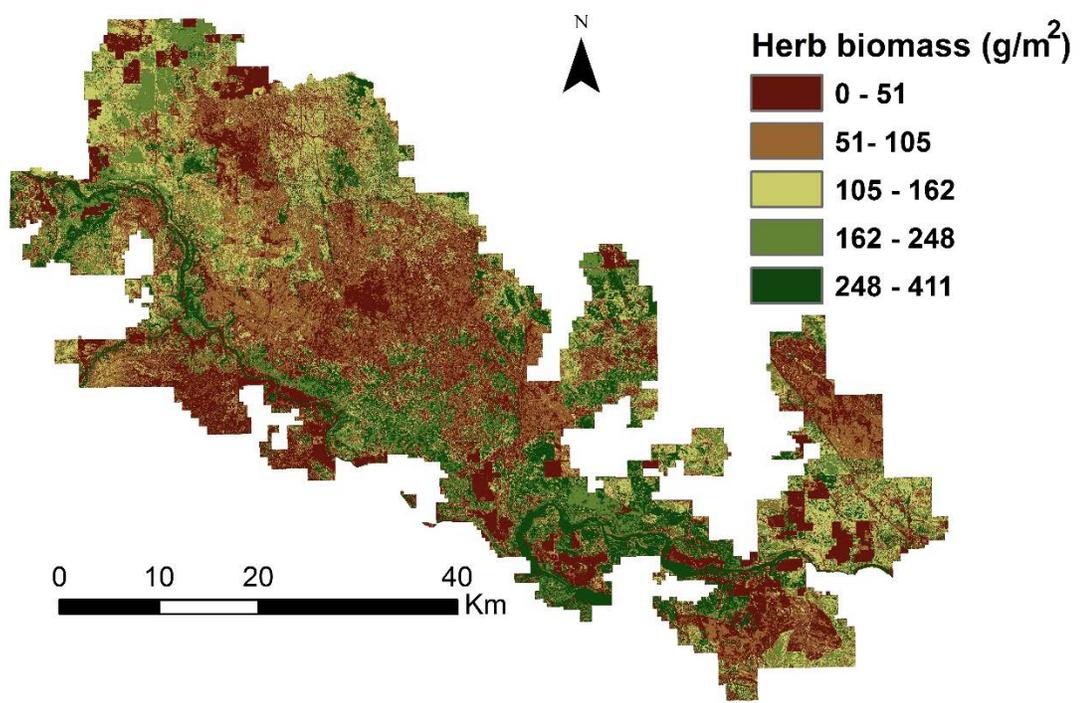


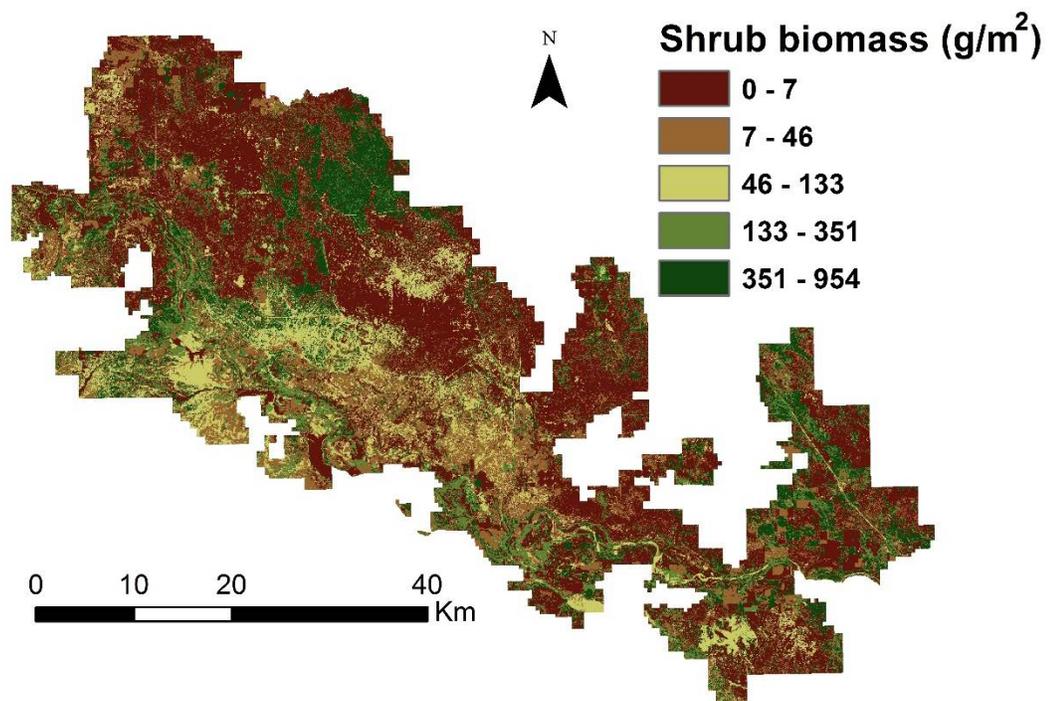
Figure 3.3 Summary distribution of (a) biomass and (b) percent cover in the NCA after imputation.

The frequency distribution of the estimated biomass and cover (**Fig. 3. 3**) is similar with *in-situ* training data (**Fig. 3. 2**). The imputed maps demonstrated that more than 70% of NCA has shrub biomass less than 100 g/m² and 50% of land with less than 100 g/m² of herb biomass in 2013. Similarly, more than 68% of NCA has shrub cover of 10% or less. However, herb cover was distributed more evenly throughout the NCA in 2013 (**Fig. 3. 3**).

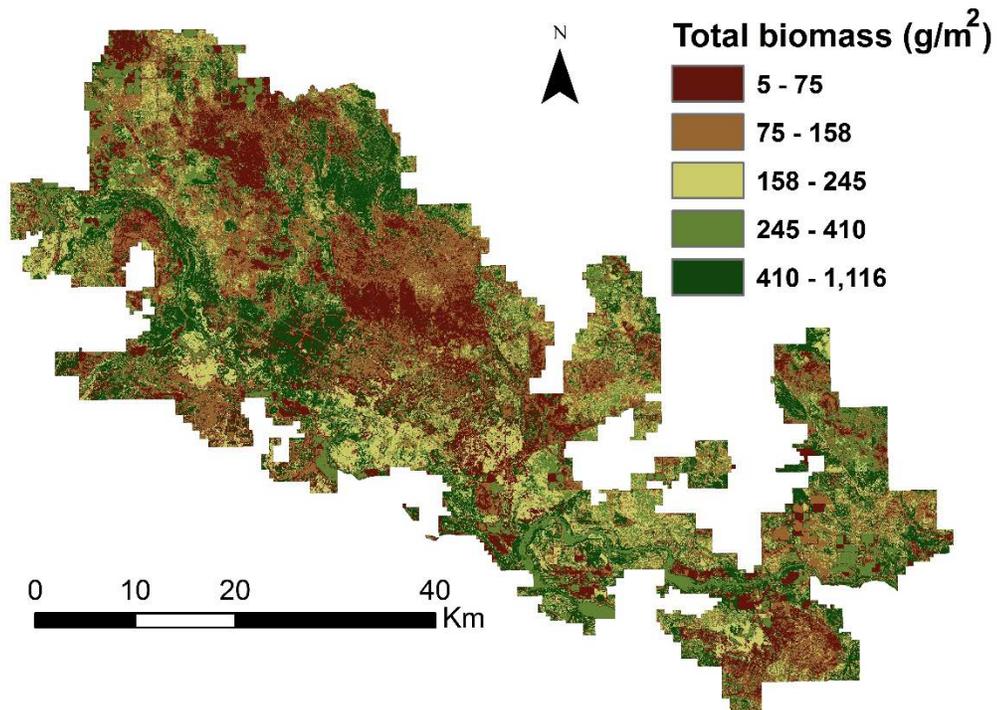
(a)



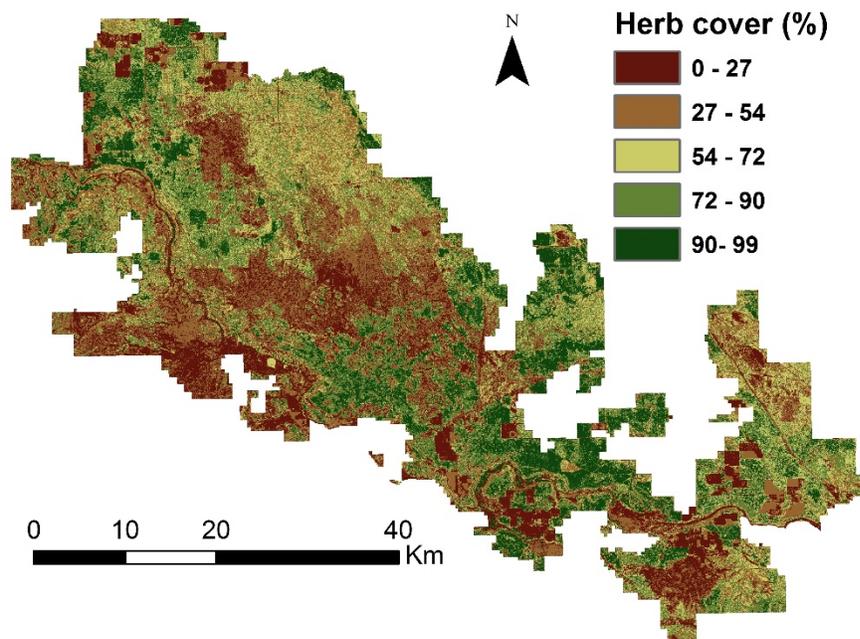
(b)



(c)



(d)



(e)

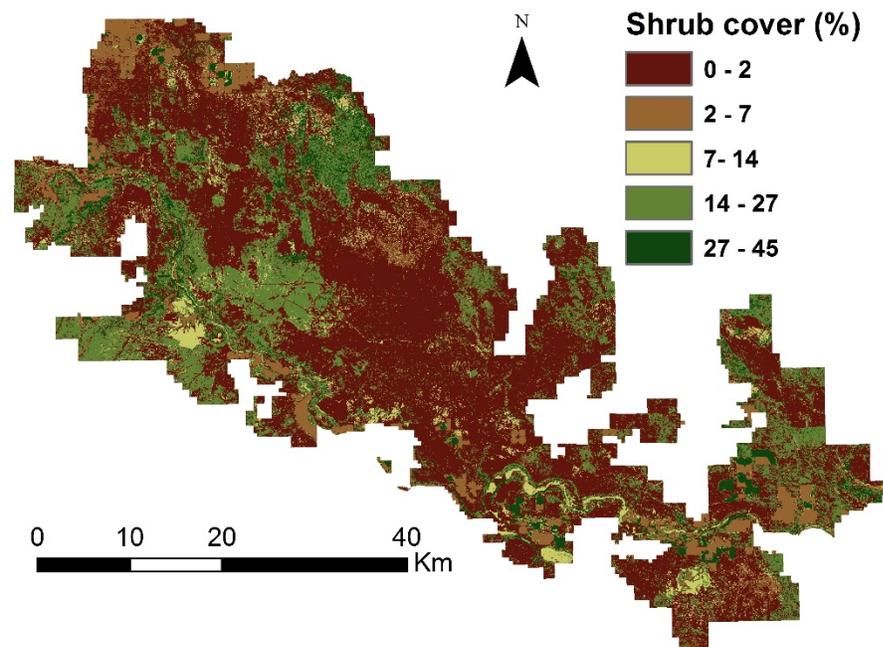


Figure 3.4 Imputed map at Landsat resolution of a) herb biomass, b) shrub biomass, c) total biomass, d) herb cover, and e) shrub cover.

3.4.3 Comparison of Landsat OLI Performance with Lidar

A comparison of the RF models from lidar and Landsat OLI demonstrated the ability of lidar to comparatively better predict vegetation characteristics of shrubs with a high R^2 and low RMSE. The lidar model explained about 75% variance in shrub biomass and cover with a RMSE of 126 g/m² and 6.7%, respectively. The Landsat model performed similarly for shrub cover with a R^2 of 0.75 and RMSE of 6.5%. For shrub biomass, however, the Landsat model explained 61% of variance and RMSE of 151 g/m². Based on the explained variance and RMSE, lidar metrics were better in predicting shrub biomass than OLI variables. Similarly, lidar explained 68% of variability while Landsat explained 57% of variability of total biomass. However, the lidar model was not able to predict the biomass and cover of herbaceous plants satisfactorily ($R^2 < 0.22$), whereas OLI variables predicted herb cover and biomass with a R^2 of more than 0.57 (**Table 3. 5**).

Table 3.5 Comparison of RF model ($n=46$) performance with Landsat 8 OLI and lidar. RMSE is in the units of g / m² for biomass predictions and % for cover predictions. The OLI metrics are labeled by date and metric from Table 2. 2. The lidar metric are labeled from Table 2. 3.

	Landsat8 OLI			Lidar		
	R^2	RMSE	Variables	R^2	RMSE	Variables
Shrub Cover	0.75	6.5	June30 SATVI, Oct4 GSAVI, Apr11 DVI	0.74	6.7	H_{range} , FHD_{All}
Herbaceous Cover	0.6	12.5	Apr11 DVI, June30 MSI, June30 VCI, June30 SATVI, Oct4 BI	0.21	17.5	H_{range} , H_{IQR}

Total Biomass	0.57	177	Oct4 BI, Oct4 NIR, Elevation, Oct4 SWIR, Oct4 MSAVI	0.68	156	FHD _{All} , H _{std} , AAD, H _{range} , H _{Skew}
Shrub Biomass	0.61	151	June30 DVI, Oct4 BI, Elevation, Apr11 DVI	0.75	126	H _{std} , H _{range} , FHD _{All} , H _{CV}
Herb Biomass	0.57	57	June30 GSAVI, June30 MSAVI, June30 SWIR, Oct4 GVI	0.12	83	H ₁₀ , H _{Skew} , CRR, AAD

3.5 Discussion

3.5.1 Selection and Performance of Predictor Variables

The results suggest that spectral indices developed from OLI alone are able to explain up to 69% of variability in herbaceous cover and up to 60% in shrub biomass in the total study area. Our results also demonstrate that corrected or adjusted indices were chosen over simple unadjusted indices in the analyses. For example, soil adjusted variables (e.g. SATVI, SAVI, GSAVI) were important predictors in all RF models. Similarly, GVI developed from a tasseled cap transformation of OLI was also an important predictor. The GVI measures the scattering of infrared radiation resulting from the cellular structure of green vegetation and absorption of visible radiation by plant pigments (Baig et al. 2014). Results in **Table 3. 4** also demonstrate that near infrared (NIR) and indices derived from SWIR (e.g. MIRI, MSI), acquired mostly in summer and

fall, have a strong relationship with herbaceous characteristics. This can be explained by the senescence of herbs, such as cheatgrass, in our field site during that time of year. The SWIR bands are sensitive to water content, and hence reflection generally increases as the vegetation dries. Similarly, NIR is sensitive to healthy plant pigments and can indicate unhealthy vegetation with low reflectance values. These reflectance values might have helped in separation of high and low biomass data and aided for a better model prediction. Similar trends were also followed by the Landsat analysis with the smaller sample size (n=46).

Another noteworthy observation is the strong relationship of indices derived from multiple acquisitions dates of Landsat 8 with both biomass and cover estimations. The phenological characteristics and soil moisture uptake of predominant annual i.e. cheatgrass and perennial i.e. sagebrush are different and changes over season which makes it relatively easy to identify and monitor using temporal remote sensing data and methods. The seasonal phenological transitions captured in a multi-temporal dataset allowed vegetation from multiple functional groups to be unmixed from each other and from the background soil which might have led to better predictions in the model. This compares favorably with the work of Shoshany and Svoray (2002), Clinton et al. (2010), and Sant et al. (2014). Using a subset of the sample size of this study, Glenn et al. (in press) also demonstrated the use of multi-temporal spectral data for successful discrimination of vegetation cover types relative to the unchanging soil background.

The model developed using lidar metrics showed strong correlation with vegetation characteristics of shrub but lower agreement with that of the herb class, which is not counter intuitive. The results are in par with the findings of Dhakal et al. (2016)

who concluded that due to the very short stature of herbaceous vegetation, it is challenging to differentiate ground from herbaceous returns using airborne lidar. The OLI spectral information was able to provide fairly good results for estimating both herb cover ($R^2 \sim 0.70$) and biomass ($R^2 \sim 0.50$). Importantly, optical data were found to be more correlated to vegetation density than to its vertical structure, demonstrated by high correlation of optical metrics with shrub and herb cover compared to their biomass. This was in par with the findings of Avitabile et al. (2012).

While our study didn't find topographic variables (slope, aspect and elevation) as important predictors of biomass and cover, several other studies have established the relationship between topographic variables and vegetation characteristics citing water availability, evapotranspiration and wind or grazing pattern as the influencing factors (Sternberg and Shoshany 2001; Osumi et al. 2003; Powell et al. 2010; Zandler et al. 2015). The results from our study simply infer that though topographic variables might be closely related to many essential biophysical factors controlling vegetation, they did not contain any additional information than OLI variables for cover and biomass estimations (Dirnböck and Dullinger 2004; Avitabile et al. 2012). Also, topography at our study site did not contain significant variation, which might have limited the influence of the topographic variables in the models. Sivanpillai et al. (2009) also evaluated topographic ancillary data to estimate sagebrush cover but did not find any statistically significant explanatory value above and beyond Landsat spectral information. They surmise that the highly disturbed nature of the study area may have limited the explanatory power of the topographic variables. A further comprehensive study is required to ascertain the role of

topographic variables in quantifying vegetation characteristics in the light of availability of new and improved remote sensing data.

3.5.2 Necessity of Data Fusion

An interesting observation in this study was the out-performance of OLI variables in estimating herb characteristics and out-performance of lidar metrics in estimating shrub characteristics, relative to each sensor. While lidar obtains its ability of estimating vegetation structure from capturing the structural information, OLI uses spectral information to record biophysical and biochemical attributes of vegetation. Lidar has been shown to have an advantage over passive spectral remote sensing methods in providing vertical distribution of canopy elements both in rangeland and forested ecosystems (Bork and Su 2007; Zolkos et al. 2013). Unlike spectral indices, lidar is less prone to saturation effects even at higher biomass yet may not differentiate between stress and healthy vegetation (Huete et al. 1997; Lefsky et al. 2002; Hyde et al. 2006; Rosenqvist et al. 2003). Lidar is less prone to the spectral soil-vegetation mixing problem but expensive for wall-to-wall coverage and known to underestimate the true height of vegetation (Streutker and Glenn 2006). Hence, lidar and spectral data are complementary in the semi-arid rangeland studied here.

Zolkos et al. (2013) were able to estimate biomass to within ~ 10% of field measurements using a fusion of lidar and passive optical sensors in forested ecosystem. Glenn et al. (in press) used a synergy of Landsat 8 OLI and lidar in the same field site as this study, to show an improvement in the estimation of biomass and cover. The R^2 improved from 0.52-0.56 when using OLI or lidar-based metrics in isolation to 0.68 in the combined OLI-lidar model for biomass. Bork and Su (2007) combined multispectral

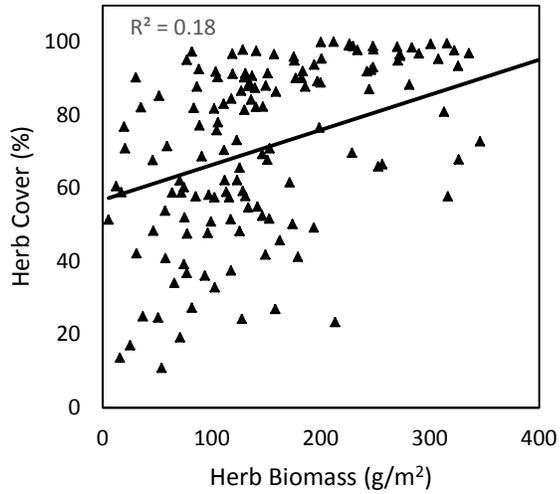
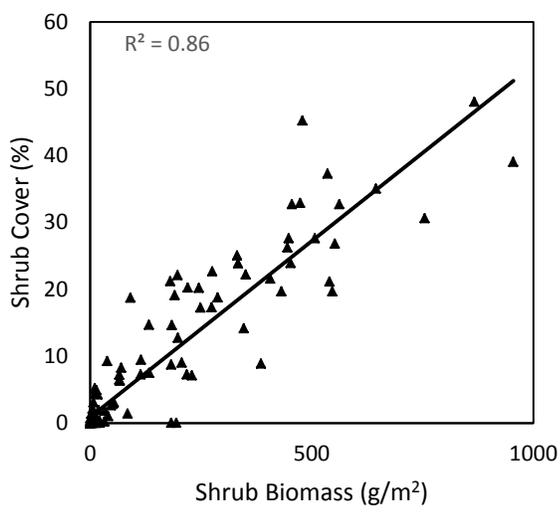
imagery and lidar for classification of rangeland vegetation with 83.9% of overall accuracy. Similarly, Mitchell et al. (2015) estimated cover using hyperspectral and discrete return lidar explaining 58% of variance in cover in a similar rangeland. Hence, future applications should take advantage of multi-sensor fusion where possible to reap the benefit of data integration in low stature ecosystem management.

3.5.3 Relation Between Cover and Biomass

We analyzed the relationship between cover and biomass for herbaceous and shrubs in the field as well as in the imputed Landsat-based map. A linear regression from *in-situ* measurements demonstrated a strong relation between shrub biomass and cover (**Fig. 3. 5**) but a weak correlation between herb biomass and cover. Our result implies that shrub cover can explain about 86% variance in biomass of shrub. This is consistent with the findings of Li et al. (2015) who found shrub biomass in a rangeland study site was highly related to percent vegetation cover ($R^2 = 0.87$). Shrubs are higher in stature than herbs and their heights are representative of both volume and biomass which in turn are highly related to percent cover. However, biomass prediction of herbs using cover measurements, is challenging because of the short stature. The herb class in our field site was composed of grasses and forbs, including annuals like cheatgrass (*Bromus tectorum*), Russian thistle/tumbleweed (*Salsola* spp.), medusahead (*Taeniatherum caputmedusae*) and perennials such as bunchgrass (*Festuca idahoensis*), intermediate wheatgrass (*Thinopyrum intermedium*), Russian wild rye (*Psathyrostachys juncea*), etc. Unlike perennials, annual herbs have a higher percent cover, higher density and occupy less volume, which will lead to an overestimation of biomass in remote sensing. However, analysis of the imputed map had a different result - the correlation between cover and

biomass increased for herb and decreased for shrub. These differences may be explained by the continuous distribution and higher density of herbs across the field site which was easily detected by the passive spectral sensors. On the other hand, heterogeneity in the distribution of shrub together with its low density made its detection more challenging at a large scale with Landsat.

(a) *In-situ* measurements



(b) Estimation from imputation

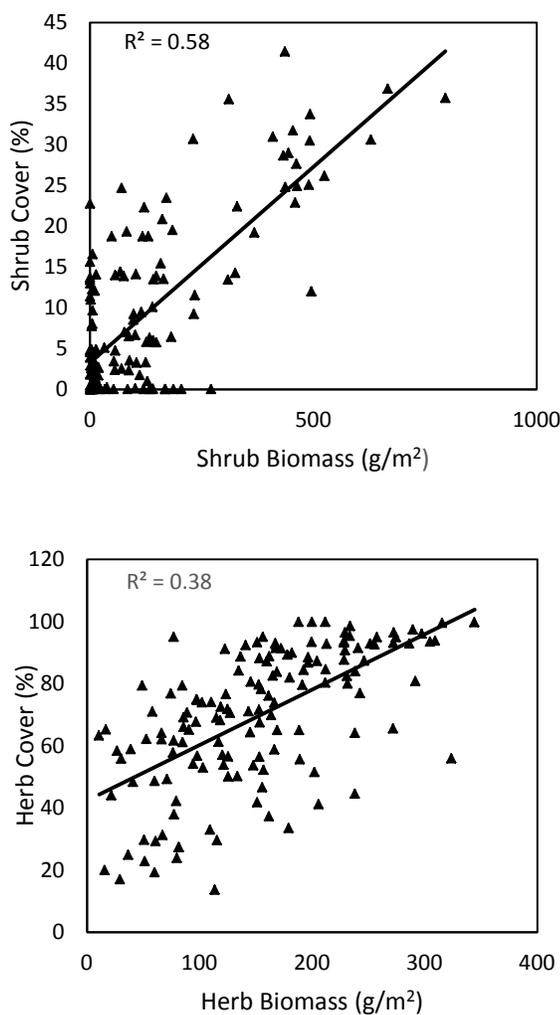


Figure 3.5 Relationship between cover and biomass in shrub and herb from (a) in-situ measurements and (b) imputation for all $n=144$ plots. There is a significant relationship between shrub biomass and cover in the field measurements. Imputed estimation showed that the correlation between shrub biomass and cover decreased while correlation between herb biomass and cover increased relative to field measurements

3.6 Conclusion

Large scale estimation of vegetation attributes is important to understand climate regulation through carbon storage and to better manage conservation efforts for ecological services. Using multi-temporal Landsat 8 OLI spectral data and a machine

learning approach, we have demonstrated the large scale mapping of biomass and percent cover in semi-arid rangeland in western United States. We validated our model with an independent data set and underscore the importance of soil adjusted vegetation indices as important predictors in a study site with sparse vegetation. We also validated lidar as an important remote sensing technique, especially to measure shrub attributes. We have demonstrated our approach to be successful in mapping herb/shrub characteristics but it is also too coarse to be able to sub classify into native-non-native or between herbs, forbs or grasses.

These findings are reflective of the *in-situ* data collected for model run. The biases which might have been introduced during field collection are extended to the model runs. This resulted in the imputed map with the majority of plots statistically skewed towards high herbaceous cover. This might explain the reason behind the comparatively low performance of the independent OLI validation model. Similarly, our models are limited in their capacity to address the heterogeneity in the field site. Larger field plots with a higher field samples are desirable to keep the biases low. Given the heterogeneity in drylands, no one technique may be sufficient enough to map the vegetation characteristics at a larger scale. Hence, we stress developing models with a fusion of spectral and lidar data to improve the prediction capabilities and keep uncertainties low. This study can be used to plan future remote sensing surveys in drylands. It can also provide baseline information on shrub biomass and cover, albeit with error, for informed management decisions for ecosystem management, wildfire mitigation and increasing resilience to climate change.

3.7 References

- Ahmed, Oumer S., Steven E. Franklin, Michael A. Wulder, and Joanne C. White. 2015. "Characterizing Stand-Level Forest Canopy Cover and Height Using Landsat Time Series, Samples of Airborne LiDAR, and the Random forest Algorithm." *ISPRS Journal of Photogrammetry and Remote Sensing* 101 (March): 89–101. doi:10.1016/j.isprsjprs.2014.11.007.
- Andersen, H.E., McGaughey, R.J. and Reutebuch, S.E., 2005. Estimating forest canopy fuel parameters using LIDAR data. *Remote sensing of Environment*, 94(4), pp.441-449.
- Anderson, Jay E., and Richard S. Inouye. 2001. "Landscape-Scale Changes in Plant Species Abundance and Biodiversity of a Sagebrush Steppe over 45 Years." *Ecological Monographs* 71 (4): 531–56. doi:10.2307/3100035.
- Anderson, Kyle. 2014. "Vegetation Measurement in Sagebrush Steppe Using Terrestrial Laser Scanning." Idaho State University.
- Angell, Raymond F, Tony Svejcar, Jon Bates, Nicanor Z Saliendra, and Douglas A Johnson. 2001. "Bowen Ratio and Closed Chamber Carbon Dioxide Flux Measurements over Sagebrush Steppe Vegetation." *Agricultural and Forest Meteorology* 108 (2): 153–61. doi:10.1016/S0168-1923(01)00227-1.
- Asner, Gregory P., George VN Powell, Joseph Mascaro, David E. Knapp, John K. Clark, James Jacobson, Ty Kennedy-Bowdoin, et al. 2010. "High-Resolution Forest Carbon Stocks and Emissions in the Amazon." *Proceedings of the National Academy of Sciences* 107 (38): 16738–42.
- Avitabile, Valerio, Alessandro Baccini, Mark A. Friedl, and Christiane Schmullius. 2012. "Capabilities and Limitations of Landsat and Land Cover Data for Aboveground Woody Biomass Estimation of Uganda." *Remote Sensing of Environment, Remote Sensing of Urban Environments*, 117 (February): 366–80. doi:10.1016/j.rse.2011.10.012.

- Baig, Muhammad Hasan Ali, Lifu Zhang, Tong Shuai, and Qingxi Tong. 2014. "Derivation of a Tasselled Cap Transformation Based on Landsat 8 at-Satellite Reflectance." *Remote Sensing Letters* 5 (5): 423–31. doi:10.1080/2150704X.2014.915434.
- Balch, Jennifer K., Bethany A. Bradley, Carla M. D'Antonio, and José Gómez-Dans. 2013. "Introduced Annual Grass Increases Regional Fire Activity across the Arid Western USA (1980–2009)." *Global Change Biology* 19 (1): 173–83. doi:10.1111/gcb.12046.
- Balch, Jennifer K., Daniel C. Nepstad, and Lisa M. Curran. 2009. "Pattern and Process: Fire-Initiated Grass Invasion at Amazon Transitional Forest Edges." In *Tropical Fire Ecology*, 481–502. Springer Praxis Books. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-77381-8_17.
- Barbour, Michael G., and William Dwight Billings. 2000. *North American Terrestrial Vegetation*. Cambridge University Press.
- Basso, Bruno, Davide Cammarano, and Pasquale De Vita. "Remotely sensed vegetation indices: Theory and applications for crop management." *Rivista Italiana di Agrometeorologia* 1 (2004): 36-53.
- Bonham, Charles D. 2013. *Measurements for Terrestrial Vegetation*. John Wiley & Sons.
- Booth, D. Terrance, Samuel E. Cox, and Robert D. Berryman. 2006. "Point Sampling Digital Imagery with 'SamplePoint.'" *Environmental Monitoring and Assessment* 123 (1-3): 97–108. doi:10.1007/s10661-005-9164-7.
- Bork, Edward W., and Jason G. Su. 2007. "Integrating LIDAR Data and Multispectral Imagery for Enhanced Classification of Rangeland Vegetation: A Meta Analysis." *Remote Sensing of Environment* 111 (1): 11–24. doi:10.1016/j.rse.2007.03.011.
- Branson, F. A., Reuben F. Miller, and I. S. McQueen. 1976. "Moisture Relationships in Twelve Northern Desert Shrub Communities Near Grand Junction, Colorado." *Ecology* 57 (6): 1104–24. doi:10.2307/1935039.
- Breiman, Leo. 1999. "Random forests--Random Features." Statistics Technical Reports. University of California at Berkeley, Berkeley, California: Statistics Department,

University of California, Berkeley. <http://stat-reports.lib.berkeley.edu/accessPages/567.html>.

- Breiman, Leo. 2001. "Random forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Brooks, Matthew L., Carla M. D'Antonio, David M. Richardson, James B. Grace, Jon E. Keeley, Joseph M. DiTomaso, Richard J. Hobbs, Mike Pellant, and David Pyke. 2004. "Effects of Invasive Alien Plants on Fire Regimes." *BioScience* 54 (7): 677–88. doi:10.1641/0006-3568(2004)054[0677:EOIAP0]2.0.CO;2.
- Cardon, Zoe G., John M. Stark, Patrick M. Herron, and Jed A. Rasmussen. 2013. "Sagebrush Carrying out Hydraulic Lift Enhances Surface Soil Nitrogen Cycling and Nitrogen Uptake into Inflorescences." *Proceedings of the National Academy of Sciences* 110 (47): 18988–93. doi:10.1073/pnas.1311314110.
- Chambers, Jeanne C., Bruce A. Roundy, Robert R. Blank, Susan E. Meyer, and A. Whittaker. 2007. "What Makes Great Basin Sagebrush Ecosystems Invasible by *Bromus Tectorum*?" *Ecological Monographs* 77 (1): 117–45. doi:10.1890/05-1991.
- Clark, Patrick E., Stuart P. Hardegree, Corey A. Moffet, and Fredrick B. Pierson. 2008. "Point Sampling to Stratify Biomass Variability in Sagebrush Steppe Vegetation." *Rangeland Ecology & Management* 61 (6): 614–22. doi:10.2111/07-147.1.
- Clinton, Nicholas E., Christopher Potter, Bob Crabtree, Vanessa Genovese, Peggy Gross, and Peng Gong. 2010. "Remote Sensing-Based Time-Series Analysis of Cheatgrass (*Bromus Tectorum* L.) Phenology." *Journal of Environmental Quality* 39 (3): 955–63. doi:10.2134/jeq2009.0158.
- Connelly, J., S. Knick, M. Schroeder, and S. Stiver. 2004. "Conservation Assessment of Greater Sage-Grouse and Sagebrush Habitats." *All U.S. Government Documents (Utah Regional Depository)*, June. <http://digitalcommons.usu.edu/govdocs/73>.
- Crist, E. P. 1983. "The Thematic Mapper Tasseled Cap - A Preliminary Formulation." In <http://ntrs.nasa.gov/search.jsp?R=19840030259>.

- Crookston, Nicholas L., and Andrew O. Finley. 2008. *yaImpute: An R Package for kNN Imputation*. <http://www.treesearch.fs.fed.us/pubs/29365>.
- Davies, Kirk W., and Aleta M. Nafus. 2013. "Exotic Annual Grass Invasion Alters Fuel Amounts, Continuity and Moisture Content." *International Journal of Wildland Fire* 22 (3): 353–58.
- Dhakal, Shital., A.Li., L.Spaete., D. Shinneman., R. Arkle., D. Pilliod., S. McIlroy., C. Baun., N. Glenn. 2016. "Airborne Lidar Based Estimation and Scaling of Semiarid Biomass Using Random Forest Variable Selection," Boise State University.
- Didham, Raphael K., Jason M. Tylianakis, Neil J. Gemmill, Tatyana A. Rand, and Robert M. Ewers. 2007. "Interactive Effects of Habitat Modification and Species Invasion on Native Species Decline." *Trends in Ecology & Evolution* 22 (9): 489–96. doi:10.1016/j.tree.2007.07.001.
- Dirnböck, T., and S. Dullinger. 2004. "Habitat Distribution Models, Spatial Autocorrelation, Functional Traits and Dispersal Capacity of Alpine Plant Species." *Journal of Vegetation Science* 15 (1): 77–84. doi:10.1111/j.1654-1103.2004.tb02239.x.
- Eitel, J. U. H., D. S. Long, P. E. Gessler, E. R. Hunt, and D. J. Brown. 2009. "Sensitivity of Ground-Based Remote Sensing Estimates of Wheat Chlorophyll Content to Variation in Soil Reflectance." *Soil Science Society of America Journal* 73 (5): 1715. doi:10.2136/sssaj2008.0288.
- Erdody, Todd L., and L. Monika Moskal. 2010. "Fusion of LiDAR and Imagery for Estimating Forest Canopy Fuels." *Remote Sensing of Environment* 114 (4): 725–37. doi:10.1016/j.rse.2009.11.002.
- Escuin, S., R. Navarro, and P. Fernández. 2008. "Fire Severity Assessment by Using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) Derived from LANDSAT TM/ETM Images." *International Journal of Remote Sensing* 29 (4): 1053–73. doi:10.1080/01431160701281072.

- Eskelson, Bianca N. I., Hailemariam Temesgen, Valerie Lemay, Tara M. Barrett, Nicholas L. Crookston, and Andrew T. ; Hudak. 2009. "The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases." *Scandinavian Journal of Forest Research* 24: 235–46.
- Estornell, J., L. A. Ruiz, B. Velázquez-Martí, and T. Hermosilla. 2012. "Estimation of Biomass and Volume of Shrub Vegetation Using LiDAR and Spectral Data in a Mediterranean Environment." *Biomass and Bioenergy*, International Conference on Lignocellulosic ethanol, 46 (November): 710–21.
doi:10.1016/j.biombioe.2012.06.023.
- Fernandes, Richard, and Sylvain G. Leblanc. 2005. "Parametric (modified Least Squares) and Non-Parametric (Theil–Sen) Linear Regressions for Predicting Biophysical Parameters in the Presence of Measurement Errors." *Remote Sensing of Environment* 95 (3): 303–16. doi:10.1016/j.rse.2005.01.005.
- Frank, T.D., S. A. Tweddale. 2006. "The Effect of Spatial Resolution on Measurement of Vegetation Cover in Three Mojave Desert Shrub Communities." *Journal of Arid Environments* 67: 88–99. doi:10.1016/j.jaridenv.2006.09.020.
- Fynn, R. W. S., and T. G. O'Connor. 2000. "Effect of Stocking Rate and Rainfall on Rangeland Dynamics and Cattle Performance in a Semi-Arid Savanna, South Africa." *Journal of Applied Ecology* 37 (3): 491–507.
- Gao, Bo-cai. 1996. "NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sensing of Environment* 58 (3): 257–66. doi:10.1016/S0034-4257(96)00067-3.
- Glenn, Edward P., Alfredo R. Huete, Pamela L. Nagler, and Stephen G. Nelson. 2008. "Relationship Between Remotely-Sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape." *Sensors* 8 (4): 2136–60. doi:10.3390/s8042136.
- Glenn, N. F., L. P. Spaete, T. T. Sankey, D. R. Derryberry, S. P. Hardegree, and J. J. Mitchell. 2011. "Errors in LiDAR-Derived Shrub Height and Crown Area on

- Sloped Terrain.” *Journal of Arid Environments* 75 (4): 377–82.
doi:10.1016/j.jaridenv.2010.11.005.
- Glenn, N. F., Amy Neuenschwander, Lee A Vierling, Lucas spete, Aihua Li, Douglas Shinneman, David S. Pilliod, Robert Arkle, Susan McIlro. In press. "Landsat 8 and ICESat-2: Synergies for vegetation structure in dryland ecosystem".
- Guan, H., Yu, J. and Luo, L., 2012. Random Forests-Based Feature Selection for Land-Use Classification Using LIDAR Data and Orthoimagery.*International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences*, 39, p.B7.
- Hall, S. A., I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker. 2005. “Estimating Stand Structure Using Discrete-Return Lidar: An Example from Low Density, Fire Prone Ponderosa Pine Forests.” *Forest Ecology and Management* 208 (1–3): 189–209. doi:10.1016/j.foreco.2004.12.001.
- Hartfield, Kyle A., Katheryn I. Landau, and Willem J. D. van Leeuwen. 2011. “Fusion of High Resolution Aerial Multispectral and LiDAR Data: Land Cover in the Context of Urban Mosquito Habitat.” *Remote Sensing* 3 (11): 2364–83.
doi:10.3390/rs3112364.
- Heitschmidt, Rodney K., and Jerry W. Stuth, eds. 1991. *Grazing Management: An Ecological Perspective*. Portland, Or: Timber Press, Incorporated.
- Hill, Sarah J., Peter J. Tung, and Michelle R. Leishman. 2005. “Relationships between Anthropogenic Disturbance, Soil Properties and Plant Invasion in Endangered Cumberland Plain Woodland, Australia.” *Austral Ecology* 30 (7): 775–88.
doi:10.1111/j.1442-9993.2005.01518.x.
- Hobbs, RJ, and L Atkins. 1990. “Fire-Related Dynamics of a Banksia Woodland in South-Western Western Australia.” *Australian Journal of Botany* 38 (1): 97–110.
- Hudak, Andrew T., Nicholas L. Crookston, Jeffrey S. Evans, David E. Hall, and Michael J. Falkowski. 2008. “Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data.” *Remote Sensing of Environment*,

- Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue, 112 (5): 2232–45. doi:10.1016/j.rse.2007.10.009.
- Huete, A, K Didan, T Miura, E. P Rodriguez, X Gao, and L. G Ferreira. 2002. “Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices.” *Remote Sensing of Environment*, The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, 83 (1–2): 195–213. doi:10.1016/S0034-4257(02)00096-2.
- Huete, A. R. 1988. “A Soil-Adjusted Vegetation Index (SAVI).” *Remote Sensing of Environment* 25 (3): 295–309. doi:10.1016/0034-4257(88)90106-X.
- Huete, A.R., HuiQing Liu, and W.J.D. van Leeuwen. 1997. “The Use of Vegetation Indices in Forested Regions: Issues of Linearity and Saturation.” In *Geoscience and Remote Sensing, 1997. IGARSS '97. Remote Sensing - A Scientific Vision for Sustainable Development., 1997 IEEE International*, 4:1966–68 vol.4. doi:10.1109/IGARSS.1997.609169.
- Hufkens, K., J. Bogaert, Q. H. Dong, L. Lu, C. L. Huang, M. G. Ma, T. Che, X. Li, F. Veroustraete, and R. Ceulemans. 2008. “Impacts and Uncertainties of Upscaling of Remote-Sensing Data Validation for a Semi-Arid Woodland.” *Journal of Arid Environments* 72 (8): 1490–1505. doi:10.1016/j.jaridenv.2008.02.012.
- Hunt Jr., E. Raymond, and Barrett N Rock. 1989. “Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances.” *Remote Sensing of Environment* 30 (1): 43–54. doi:10.1016/0034-4257(89)90046-1.
- Hyde, Peter, Ralph Dubayah, Wayne Walker, J. Bryan Blair, Michelle Hofton, and Carolyn Hunsaker. 2006. “Mapping Forest Structure for Wildlife Habitat Analysis Using Multi-Sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) Synergy.” *Remote Sensing of Environment* 102 (1–2): 63–73. doi:10.1016/j.rse.2006.01.021.
- Ismail, Riyad, Onesimo Mutanga, and Lalit Kumar. 2010. “Modeling the Potential Distribution of Pine Forests Susceptible to *Sirex Noctilio* Infestations in Mpumalanga, South Africa.” *Transactions in GIS* 14 (5): 709–26. doi:10.1111/j.1467-9671.2010.01229.x.

- Jackson, Ray D., and Alfredo R. Huete. 1991. "Interpreting Vegetation Indices." *Preventive Veterinary Medicine* 11 (3–4): 185–200. doi:10.1016/S0167-5877(05)80004-2.
- Kauth, R., and G. Thomas. 1976. "The Tasselled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT." *LARS Symposia*, January. http://docs.lib.purdue.edu/lars_symp/159.
- Knick, Steven T. 1999. "Requiem for a Sagebrush Ecosystem?" <https://research.wsulibs.wsu.edu:8443/xmlui/handle/2376/1169>.
- Lefsky, M. A., W. B. Cohen, Geoffrey G. Parker, and D. J. Harding. 2002. "Lidar Remote Sensing for Ecosystem Studies." <https://repository.si.edu/handle/10088/2924>.
- Lefsky, Michael A. 2010. "A Global Forest Canopy Height Map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System." *Geophysical Research Letters* 37 (15): L15401. doi:10.1029/2010GL043622.
- Li, P., Jiang, L., & Feng, Z. (2013). Cross-comparison of vegetation indices derived from Landsat-7 enhanced thematic mapper plus (ETM+) and Landsat-8 operational land imager (OLI) sensors. *Remote Sensing*, 6(1), 310-329.
- Li, Aihua, Nancy F. Glenn, Peter J. Olsoy, Jessica J. Mitchell, and Rupesh Shrestha. 2015. "Aboveground Biomass Estimates of Sagebrush Using Terrestrial and Airborne LiDAR Data in a Dryland Ecosystem." *Agricultural and Forest Meteorology* 213 (November): 138–47. doi:10.1016/j.agrformet.2015.06.005.
- Lin, Yi, Juha Hyypä, Antero Kukko, Anttoni Jaakkola, and Harri Kaartinen. 2012. "Tree Height Growth Measurement with Single-Scan Airborne, Static Terrestrial and Mobile Laser Scanning." *Sensors* 12 (9): 12798–813. doi:10.3390/s120912798.
- Liu, Jian, Ming Dong, Shi Li Miao, Zhen Yu Li, Ming Hua Song, and Ren Qing Wang. 2006. "Invasive Alien Plants in China: Role of Clonality and Geographical Origin." *Biological Invasions* 8 (7): 1461–70. doi:10.1007/s10530-005-5838-x.

- Liu, J. Y., D. F. Zhuang, D. Luo, and X. Xiao. 2003. "Land-Cover Classification of China: Integrated Analysis of AVHRR Imagery and Geophysical Data." *International Journal of Remote Sensing* 24 (12): 2485–2500. doi:10.1080/01431160110115582.
- Lu, D. 2005. "Aboveground Biomass Estimation Using Landsat TM Data in the Brazilian Amazon." *International Journal of Remote Sensing* 26 (12): 2509–25. doi:10.1080/01431160500142145.
- Lu, Dengsheng. 2006. "The Potential and Challenge of Remote Sensing-Based Biomass Estimation." *International Journal of Remote Sensing* 27 (April): 1297–1328. doi:10.1080/01431160500486732.
- Lu, Dengsheng., Qi Chen. 2012. "Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates." *International Journal of Forestry Research*. doi:10.1155/2012/436537.
- Marsett, R.C., Jianguo Qi. 2006. "Remote Sensing for Grassland Management in the Arid Southwest." *Rangeland Ecology & Management* 59 (5). doi:10.2111/05-201R.1.
- Miller, R. F., and J. A. Rose. 2006. "Fire History and Western Juniper Encroachment in Sagebrush Steppe." *Journal of Range Management Archives* 52 (6): 550–59.
- Miller, Richard F., Steven T. Knick, David A. Pyke, Cara W. Meinke, Steven E. Hanser, Michael J. Wisdom, and Ann L. Hild. 2011. "Characteristics of Sagebrush Habitats and Limitations to Long-Term Conservation." *Studies in Avian Biology* 38: 145–84.
- Mitchard, E. T. A., S. S. Saatchi, L. J. T. White, K. A. Abernethy, K. J. Jeffery, S. L. Lewis, M. Collins, et al. 2012. "Mapping Tropical Forest Biomass with Radar and Spaceborne LiDAR in Lopé National Park, Gabon: Overcoming Problems of High Biomass and Persistent Cloud." *Biogeosciences* 9 (1): 179–91. doi:10.5194/bg-9-179-2012.

- Mitchell, J.J., Shrestha, R., Moore-Ellison, C.A. and Glenn, N.F., 2013. Single and multi-date Landsat classifications of basalt to support soil survey efforts. *Remote Sensing*, 5(10), pp.4857-4876.
- Mitchell, Jessica J., Rupesh Shrestha, Lucas P. Spaete, and Nancy F. Glenn. 2015. "Combining Airborne Hyperspectral and LiDAR Data across Local Sites for Upscaling Shrubland Structural Information: Lessons for HypSIRI." *Remote Sensing of Environment*. Accessed July 5. doi:10.1016/j.rse.2015.04.015.
- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Mutanga, Onesimo, Elhadi Adam, and Moses Azong Cho. 2012. "High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random forest Regression Algorithm." *International Journal of Applied Earth Observation and Geoinformation* 18 (August): 399–406. doi:10.1016/j.jag.2012.03.012.
- Naidoo, L., M. A. Cho, R. Mathieu, and G. Asner. 2012. "Classification of Savanna Tree Species, in the Greater Kruger National Park Region, by Integrating Hyperspectral and LiDAR Data in a Random forest Data Mining Environment." *ISPRS Journal of Photogrammetry and Remote Sensing* 69 (April): 167–79. doi:10.1016/j.isprsjprs.2012.03.005.
- Nellis, M. Duane, and John M. Briggs. 1992. "Transformed Vegetation Index for Measuring Spatial Variation in Drought Impacted Biomass on Konza Prairie, Kansas." *Transactions of the Kansas Academy of Science (1903-)* 95 (1/2): 93–99. doi:10.2307/3628024.
- Olden, Julian D., Lawler, and N. LeRoy Poff. 2008. "Machine Learning Methods Without Tears: A Primer for Ecologists." *The Quarterly Review of Biology* 83 (2): 171–93. doi:10.1086/587826.
- Osumi, Katsuhiko, Shigeto Ikeda, and Toru Okamoto. 2003. "Vegetation Patterns and Their Dependency on Site Conditions in the Pre-Industrial Landscape of North-

- Eastern Japan.” *Ecological Research* 18 (6): 753–65. doi:10.1111/j.1440-1703.2003.00593.x.
- Pal, M. 2005. “Random forest Classifier for Remote Sensing Classification.” *International Journal of Remote Sensing* 26 (1): 217–22. doi:10.1080/01431160412331269698.
- Phua, Mui-How, and Hideki Saito. 2003. “Estimation of Biomass of a Mountainous Tropical Forest Using Landsat TM Data.” *Canadian Journal of Remote Sensing* 29 (4): 429–40. doi:10.5589/m03-005.
- Powell, Scott L., Warren B. Cohen, Sean P. Healey, Robert E. Kennedy, Gretchen G. Moisen, Kenneth B. Pierce, and Janet L. Ohmann. 2010. “Quantification of Live Aboveground Forest Biomass Dynamics with Landsat Time-Series and Field Inventory Data: A Comparison of Empirical Modeling Approaches.” *Remote Sensing of Environment* 114 (5): 1053–68. doi:10.1016/j.rse.2009.12.018.
- Prasad, Anantha M., Louis R. Iverson, and Andy Liaw. 2006. “Newer Classification and Regression Tree Techniques: Bagging and Random forests for Ecological Prediction.” *Ecosystems* 9 (2): 181–99. doi:10.1007/s10021-005-0054-1.
- Price, Kevin P., Xulin Guo, and James M. Stiles. 2002. “Optimal Landsat TM Band Combinations and Vegetation Indices for Discrimination of Six Grassland Types in Eastern Kansas.” *International Journal of Remote Sensing* 23 (23): 5031–42. doi:10.1080/01431160210121764.
- Purdie, Rosemary W., and R. O. Slatyer. 1976. “Vegetation Succession after Fire in Sclerophyll Woodland Communities in South-Eastern Australia.” *Australian Journal of Ecology* 1 (4): 223–36. doi:10.1111/j.1442-9993.1976.tb01111.x.
- Purevdorj, TS, R. Tateishi, T. Ishiyama, and Y. Honda. 1998. “Relationships between Percent Vegetation Cover and Vegetation Indices.” *International Journal of Remote Sensing* 19 (18): 3519–35. doi:10.1080/014311698213795.
- Qi, J., A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian. 1994. “A Modified Soil Adjusted Vegetation Index.” *Remote Sensing of Environment* 48 (2): 119–26. doi:10.1016/0034-4257(94)90134-1.

- Robinson, T.P., P. Novelly. 2009. "Towards an Approach for Remote Sensing-Based Rangeland Condition Assessment in North Western Australia," 899–913.
- Rosenqvist, Åke, et al. 2003 "A review of remote sensing technology in support of the Kyoto Protocol." *Environmental Science & Policy* 6.5 2003: 441-455.
- Ritchie, J. C., J. H. Everitt, D. E. Escobar, T. J. Jackson, and M. R. Davis. 2006. "Airborne Laser Measurements of Rangeland Canopy Cover and Distribution." *Journal of Range Management Archives* 45 (2): 189–93.
- Sala, O. E., and W. K. Lauenroth. 1982. "Small Rainfall Events: An Ecological Role in Semiarid Regions." *Oecologia* 53 (3): 301–4. doi:10.1007/BF00389004.
- Salford Predictive Modeler, 2013. User guide. Retrieved from: <http://www.salford-systems.com/products/spm/userguide>.
- Sant, Eric D., Gregg E. Simonds, R. Douglas Ramsey, and Randy T. Larsen. 2014. "Assessment of Sagebrush Cover Using Remote Sensing at Multiple Spatial and Temporal Scales." *Ecological Indicators* 43 (August): 297–305. doi:10.1016/j.ecolind.2014.03.014.
- Shinneman, D.J., R. Arkle, D. Pilliod, N.F. Glenn. 2011. Quantifying and Predicting Fuels and the Effects of Reduction Treatments Along Successional and Invasion Gradients in Sagebrush Habitats. Retrieved from: http://www.firescience.gov/JFSP_funded_project_detail.cfm?jdbid=%24%26ZO9V0%20%20%0A
- Shoshany, M. and Svoray, T., 2002. Multidate adaptive unmixing and its application to analysis of ecosystem transitions along a climatic gradient. *Remote Sensing of Environment*, 82(1), pp.5-20.
- Shiple, Lisa A., Tara B. Davila, Nicole J. Thines, and Becky A. Elias. 2006. "Nutritional Requirements and Diet Choices of the Pygmy Rabbit (*Brachylagus Idahoensis*): A Sagebrush Specialist." *Journal of Chemical Ecology* 32 (11): 2455–74. doi:10.1007/s10886-006-9156-2.

- Shrestha, Gyami, and Peter D. Stahl. 2008. "Carbon Accumulation and Storage in Semi-Arid Sagebrush Steppe: Effects of Long-Term Grazing Exclusion." *Agriculture, Ecosystems & Environment* 125 (1–4): 173–81. doi:10.1016/j.agee.2007.12.007.
- Sims, P.L. and Singh, J.S., 1978. The structure and function of ten western North American grasslands: II. intra-seasonal dynamics in primary producer compartments. *The Journal of Ecology*, pp.547-572.
- Sivanpillai, Ramesh, Steven D. Prager, and Thomas O. Storey. 2009. "Estimating Sagebrush Cover in Semi-Arid Environments Using Landsat Thematic Mapper Data." *International Journal of Applied Earth Observation and Geoinformation* 11 (2): 103–7. doi:10.1016/j.jag.2008.10.001.
- Smith, S. D., T. E. Huxman, S. F. Zitzer, T. N. Charlet, D. C. Housman, J. S. Coleman, L. K. Fenstermaker, J. R. Seemann, and R. S. Nowak. 2000. "Elevated CO₂ Increases Productivity and Invasive Species Success in an Arid Ecosystem." *Nature* 408 (6808): 79–82. doi:10.1038/35040544.
- Sripada, Ravi P., Ronnie W. Heiniger, Jeffrey G. White, and Alan D. Meijer. 2006. "Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn." *Agronomy Journal* 98 (4): 968. doi:10.2134/agronj2005.0200.
- Sternberg, Marcelo, and Maxim Shoshany. 2001. "Influence of Slope Aspect on Mediterranean Woody Formations: Comparison of a Semiarid and an Arid Site in Israel." *Ecological Research* 16 (2): 335–45. doi:10.1046/j.1440-1703.2001.00393.x.
- Stokes, C. J., A. Ash, and S. W. Howden. 2008. "Climate Change Impacts on Australian Rangelands." *Rangelands Archives* 30 (3): 40–45.
- Storch, Ilse. 2007. "Conservation Status of Grouse Worldwide: An Update." *Wildlife Biology* 13 (sp1): 5–12. doi:10.2981/0909-6396(2007)13[5:CSOGWA]2.0.CO;2.
- Streutker, David R., and Nancy F. Glenn. 2006. "LiDAR Measurement of Sagebrush Steppe Vegetation Heights." *Remote Sensing of Environment* 102 (1–2): 135–45. doi:10.1016/j.rse.2006.02.011.

- Su, J.G. and Bork, E.W., 2007. Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Applied Vegetation Science*, 10(3), pp.407-416.
- Swatantran, Anu, Ralph Dubayah, Dar Roberts, Michelle Hofton, and J. Bryan Blair. 2011. "Mapping Biomass and Stress in the Sierra Nevada Using Lidar and Hyperspectral Data Fusion." *Remote Sensing of Environment*, DESDynI VEG-3D Special Issue, 115 (11): 2917–30. doi:10.1016/j.rse.2010.08.027.
- Taylor, Kimberley, Tyler Brummer, Lisa J. Rew, Matt Lavin, and Bruce D. Maxwell. 2014. "Bromus Tectorum Response to Fire Varies with Climate Conditions." *Ecosystems* 17 (6): 960–73. doi:10.1007/s10021-014-9771-7.
- Thenkabail, P.S., A. D. Ward. 1994. "Thematic Mapper Vegetation Indices for Determining Soybean and Corn Growth Parameters." *Photogrammetric Engineering and Remote Sensing* 60: 437–42.
- Todd, S. W., R. M. Hoffer, and D. G. Milchunas. 1998. "Biomass Estimation on Grazed and Ungrazed Rangelands Using Spectral Indices." *International Journal of Remote Sensing* 19 (3): 427–38. doi:10.1080/014311698216071.
- Tucker, C.J., Vanpraet, C., Boerwinkel, E. and Gaston, E.A., 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel. *Remote Sensing of Environment*, 13(6), pp.461-474.
- USDI Bureau of Land Management. (2008). *Snake River Birds of Prey National Conservation Area Proposed Resource Management Plan and Final Environmental Impact Statement*, Boise, ID
- USGS (2013). Landsat 8, Fact Sheet 2013–3060. Retrieved from. <http://pubs.usgs.gov/fs/2013/3060/pdf/fs2013-3060.pdf>
- Vazirabad, Y.F. and Karslioglu, M.O., 2011. *Lidar for Biomass Estimation*. INTECH Open Access Publisher.
- van Wilgen, B. W., B. Reyers, D. C. Le Maitre, D. M. Richardson, and L. Schonegevel. 2008. "A Biome-Scale Assessment of the Impact of Invasive Alien Plants on Ecosystem Services in South Africa." *Journal of Environmental Management*,

- Economics of Invasive Species Management, 89 (4): 336–49.
doi:10.1016/j.jenvman.2007.06.015.
- Veraverbeke, S., I. Gitas, T. Katagis, A. Polychronaki, B. Somers, and R. Goossens. 2012. “Assessing Post-Fire Vegetation Recovery Using Red–near Infrared Vegetation Indices: Accounting for Background and Vegetation Variability.” *ISPRS Journal of Photogrammetry and Remote Sensing* 68 (March): 28–39. doi:10.1016/j.isprsjprs.2011.12.007.
- Vincenzi, Simone, Matteo Zucchetta, Piero Franzoi, Michele Pellizzato, Fabio Pranovi, Giulio A. De Leo, and Patrizia Torricelli. 2011. “Application of a Random forest Algorithm to Predict Spatial Distribution of the Potential Yield of *Ruditapes Philippinarum* in the Venice Lagoon, Italy.” *Ecological Modelling* 222 (8): 1471–78. doi:10.1016/j.ecolmodel.2011.02.007.
- Waite, R.B., 1994. The application of visual estimation procedures for monitoring pasture yield and composition in exclosures and small plots. *Tropical Grasslands*, 28, pp.38-38.
- Wallace, J.F. and Thomas, P., 1998. Rangeland monitoring in northern Western Australia using sequences of Landsat imagery, report for the WA Pastoral Lands Board.
- Watson, Robert T., Marufu C. Zinyowera, Richard H. Moss, and Intergovernmental Panel on Climate Change Working Group II. 1998. *The Regional Impacts of Climate Change: An Assessment of Vulnerability*. Cambridge University Press.
- Western Region Climate Center. (2012). *Swan Falls Power House, Idaho, Period of Record General Climate Summary*. Retrieved from: <http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?id8928>
- Whisenant, Steven G. 1990. “Changing Fire Frequencies on Idaho’s Snake River Plains: Ecological and Management Implications. In.” In *Station, Forest Service, U.S. Department of Agriculture*, 4–10.
- Wilcox, Bradford P. 2010. “Transformative Ecosystem Change and Ecohydrology: Ushering in a New Era for Watershed Management.” *Ecohydrology* 3 (1): 126–30. doi:10.1002/eco.104.

- Wilson, Adam M., John A. Silander, Alan Gelfand, and Jonathan H. Glenn. 2011. "Scaling up: Linking Field Data and Remote Sensing with a Hierarchical Model." *International Journal of Geographical Information Science* 25 (3): 509–21. doi:10.1080/13658816.2010.522779.
- Wisdom, M.J., M.M. Rowland, and R.J. Tausch. 2005. "Effective Management Strategies for Sage-Grouse and Sagebrush: A Question of Triage?" In *Transactions of the 70th North American Wildlife and Natural Resources Conference*, edited by R.D. Sparrowe and L.H. Carpenter, 70:206. Arlington, Virginia, USA: Wildlife Management Institute. http://www.fs.fed.us/rm/pubs_other/rmrs_2005_wisdom_m001.pdf.
- Wu, Hua, and Zhao-Liang Li. 2009. "Scale Issues in Remote Sensing: A Review on Analysis, Processing and Modeling." *Sensors* 9 (3): 1768–93. doi:10.3390/s90301768.
- Yang, Jian, Peter J. Weisberg, and Nathan A. Bristow. 2012. "Landsat Remote Sensing Approaches for Monitoring Long-Term Tree Cover Dynamics in Semi-Arid Woodlands: Comparison of Vegetation Indices and Spectral Mixture Analysis." *Remote Sensing of Environment* 119 (April): 62–71. doi:10.1016/j.rse.2011.12.004.
- Zandler, H., A. Brenning, and C. Samimi. 2015. "Quantifying Dwarf Shrub Biomass in an Arid Environment: Comparing Empirical Methods in a High Dimensional Setting." *Remote Sensing of Environment* 158 (March): 140–55. doi:10.1016/j.rse.2014.11.007.
- Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. and Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote sensing of Environment*, 87(2), pp.171-182.
- Zolkos, S. G., S. J. Goetz, and R. Dubayah. 2013. "A Meta-Analysis of Terrestrial Aboveground Biomass Estimation Using Lidar Remote Sensing." *Remote Sensing of Environment* 128 (January): 289–98. doi:10.1016/j.rse.2012.10.017.

4 CHAPTER FOUR: CONCLUSIONS

Both of the research papers presented in this thesis explore the potential of remote sensing data and statistical methods to better understand ecosystem information across a semi-arid rangeland ecosystem at large scales. In chapter 2, we incorporated airborne lidar with ground validation for scaling shrub and herb biomass from the plot level to a larger area covering the entire lidar footprint. We found a strong correlation of height based lidar metrics with field-measured biomass of shrub. We also compared raster processing techniques with point cloud processing and demonstrated that point cloud processing of lidar data significantly improves the estimation of biomass at coarser scales (e.g. 100 m); however at fine scales, raster processing is equivalent to point cloud processing. Similarly, in chapter 3, we leveraged a large number of Landsat 8 and lidar metrics for successful prediction of biomass and cover of shrubs at the regional scale. Our results demonstrate that Landsat performs better in estimating vegetation cover whereas lidar performs slightly better in estimating biomass of shrubs. We summarize the key findings of the research briefly as:

- 1) The vegetation cover and biomass of shrubs were successfully modeled using time-series multispectral imagery (Landsat 8) and airborne lidar.

- 2) We found that the best model to describe vegetation cover fractions included vegetation indices calculated from multiple acquisition dates of Landsat 8. The time-series data represented phenological transitions of the vegetation in the model.

3) Lidar was found to estimate shrub biomass slightly better than Landsat. The limitation of lidar is its high expense to map large areas.

4) Point cloud processing of lidar data significantly improves the estimation of biomass at the coarser scale (e.g. 100 m) compared to raster processing.

5) Lidar could not satisfactorily model herbaceous biomass ($R^2 < 2$) because of the low ground cover and low stature of the herbs.

6) As per our imputed map, the NCA contained ~ 344925600 g of herbaceous biomass and ~ 313420200 g of shrub biomass in 2013. Similarly, more than 70% of NCA has shrub biomass less than 100 g/m² and 50% of land with less than 100 g/m² of herb biomass. More than 68% of NCA has shrub cover of 10% or less.

These findings are reflective of the *in-situ* data collected for the model runs. The biases which might have been introduced during field collection are extended to the model runs. This resulted in the imputed map with the majority of plots statistically skewed towards high herbaceous cover. This might explain the reason behind the comparatively low performance of the independent OLI validation model. Similarly, our models are limited in their capacity to address the heterogeneity in the field site. A larger field plot with a higher number of samples are desirable to keep the biases low.

Our findings also illustrate the necessity of a critical assessment of available remote sensing platforms before implementing any project. This might mean, for example, the consideration of the tradeoff between the accuracy required and available resources before choosing between lidar and Landsat. For projects with limited funds or a large study area, freely available multi-temporal Landsat 8 might be the best resource. For other projects demanding higher structural information, the use of lidar dataset can be

much more conducive. Multisensor fusion can also be an alternative for applications that require additional predictive power for vegetation characteristics in low stature ecosystems.

The results presented herein are important steps toward scaling semi-arid rangeland ecological characteristics at the regional level. This is particularly significant in landscapes such as our study site which is rapidly transforming to non-native grassland from native shrub-steppe communities. The process, methods and results outlined here are important for modeling climate and hydrological dynamics, managing changes to the ecosystem, quantifying vegetation characteristics, estimating pre-fire and post-fire fuel loads, measuring carbon storage and assessing habitat quality. The remote sensing techniques described here will supplement the extensive and time-consuming field data collections over large and inaccessible areas. The remote sensing techniques may also help decision makers to take apt management decisions with lower uncertainties over large areas. More importantly, the methods and results of this research provide scientific references, resources and a framework to future studies trying to understand the ecosystem dynamics in semi-arid rangelands of the western United States using remote sensing.

5 REFERENCES

- Ahmed, Oumer S., Steven E. Franklin, Michael A. Wulder, and Joanne C. White. 2015. "Characterizing Stand-Level Forest Canopy Cover and Height Using Landsat Time Series, Samples of Airborne LiDAR, and the Random Forest Algorithm." *ISPRS Journal of Photogrammetry and Remote Sensing* 101 (March): 89–101. doi:10.1016/j.isprsjprs.2014.11.007.
- Andersen, H.E., McGaughey, R.J. and Reutebuch, S.E., 2005. Estimating forest canopy fuel parameters using LIDAR data. *Remote sensing of Environment*, 94(4), pp.441-449.
- Anderson, Jay E., and Richard S. Inouye. 2001. "Landscape-Scale Changes in Plant Species Abundance and Biodiversity of a Sagebrush Steppe over 45 Years." *Ecological Monographs* 71 (4): 531–56. doi:10.2307/3100035.
- Angell, Raymond F, Tony Svejcar, Jon Bates, Nicanor Z Saliendra, and Douglas A Johnson. 2001. "Bowen Ratio and Closed Chamber Carbon Dioxide Flux Measurements over Sagebrush Steppe Vegetation." *Agricultural and Forest Meteorology* 108 (2): 153–61. doi:10.1016/S0168-1923(01)00227-1.
- Barbour, Michael G., and William Dwight Billings. 2000. *North American Terrestrial Vegetation*. Cambridge University Press.
- Basso, Bruno, Davide Cammarano, and Pasquale De Vita. "Remotely sensed vegetation indices: Theory and applications for crop management." *Rivista Italiana di Agrometeorologia* 1 (2004): 36-53.
- Bonham, Charles D. 2013. *Measurements for Terrestrial Vegetation*. John Wiley & Sons.
- Bork, Edward W., and Jason G. Su. 2007. "Integrating LIDAR Data and Multispectral Imagery for Enhanced Classification of Rangeland Vegetation: A Meta Analysis." *Remote Sensing of Environment* 111 (1): 11–24. doi:10.1016/j.rse.2007.03.011.

- Branson, F. A., Reuben F. Miller, and I. S. McQueen. 1976. "Moisture Relationships in Twelve Northern Desert Shrub Communities Near Grand Junction, Colorado." *Ecology* 57 (6): 1104–24. doi:10.2307/1935039.
- Breiman, Leo. 1999. "Random forests--Random Features." Statistics Technical Reports. University of California at Berkeley, Berkeley, California: Statistics Department, University of California, Berkeley. <http://stat-reports.lib.berkeley.edu/accessPages/567.html>.
- Breiman, Leo. 2001. "Random forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.
- Campbell, James B., and Randolph H. Wynne. 2011. *Introduction to Remote Sensing, Fifth Edition*. 5th edition. New York: The Guilford Press.
- Cardon, Zoe G., John M. Stark, Patrick M. Herron, and Jed A. Rasmussen. 2013. "Sagebrush Carrying out Hydraulic Lift Enhances Surface Soil Nitrogen Cycling and Nitrogen Uptake into Inflorescences." *Proceedings of the National Academy of Sciences* 110 (47): 18988–93. doi:10.1073/pnas.1311314110.
- Chen, Qi, Gaia Vaglio Laurin, John J. Battles, and David Saah. 2012. "Integration of Airborne Lidar and Vegetation Types Derived from Aerial Photography for Mapping Aboveground Live Biomass." *Remote Sensing of Environment* 121 (June): 108–17. doi:10.1016/j.rse.2012.01.021.
- Clark, Patrick E., Stuart P. Hardegree, Corey A. Moffet, and Fredrick B. Pierson. 2008. "Point Sampling to Stratify Biomass Variability in Sagebrush Steppe Vegetation." *Rangeland Ecology & Management* 61 (6): 614–22. doi:10.2111/07-147.1.
- Connelly, John W., Michael A. Schroeder, Alan R. Sands, and Clait E. Braun. 2000. "Guidelines to Manage Sage Grouse Populations and Their Habitats." *Wildlife Society Bulletin* 28 (4): 967–85.
- Debouk, Haifa, Ramon Riera-Tatché, and Cristina Vega-García. 2013. "Assessing Post-Fire Regeneration in a Mediterranean Mixed Forest Using Lidar Data and Artificial Neural Networks." *Photogrammetric Engineering & Remote Sensing* 79 (12): 1121–30. doi:10.14358/PERS.79.12.1121.

- Ding, Yanling, Kai Zhao, Xingming Zheng, and Tao Jiang. 2014. "Temporal Dynamics of Spatial Heterogeneity over Cropland Quantified by Time-Series NDVI, near Infrared and Red Reflectance of Landsat 8 OLI Imagery." *International Journal of Applied Earth Observation and Geoinformation* 30 (August): 139–45. doi:10.1016/j.jag.2014.01.009.
- Dubayah, Ralph, Robert Knox, Michelle Hofton, J. Bryan Blair, and Jason Drake. 2000. "Land surface characterization using lidar remote sensing." *Spatial information for land use management* : 25-38.
- Eberhardt, Lester E., Eric E. Hanson, and Larry L. Cadwell. 1984. "Movement and Activity Patterns of Mule Deer in the Sagebrush-Steppe Region." *Journal of Mammalogy* 65 (3): 404–9. doi:10.2307/1381086.
- Estornell, J., L. A. Ruiz, B. Velázquez-Martí, and T. Hermosilla. 2012. "Estimation of Biomass and Volume of Shrub Vegetation Using LiDAR and Spectral Data in a Mediterranean Environment." *Biomass and Bioenergy*, International Conference on Lignocellulosic ethanol, 46 (November): 710–21. doi:10.1016/j.biombioe.2012.06.023.
- Farid, A., Goodrich, D. C., Bryant, R., & Sorooshian, S. (2008). Using airborne lidar to predict Leaf Area Index in cottonwood trees and refine riparian water-use estimates. *Journal of Arid Environments*, 72(1), 1–15. <http://doi.org/10.1016/j.jaridenv.2007.04.010>
- Fernandes, Richard, and Sylvain G. Leblanc. 2005. "Parametric (modified Least Squares) and Non-Parametric (Theil–Sen) Linear Regressions for Predicting Biophysical Parameters in the Presence of Measurement Errors." *Remote Sensing of Environment* 95 (3): 303–16. doi:10.1016/j.rse.2005.01.005.
- García-Gutiérrez, Jorge, Eduardo González-Ferreiro, Daniel Mateos-García, Jose C. Riquelme-Santos, and David Miranda. 2011. "A Comparative Study between Two Regression Methods on LiDAR Data: A Case Study." In *Hybrid Artificial Intelligent Systems*, edited by Emilio Corchado, Marek Kurzyński, and Michał

- Woźniak, 311–18. Lecture Notes in Computer Science 6679. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-642-21222-2_38.
- Glenn, N. F., L. P. Spaete, T. T. Sankey, D. R. Derryberry, S. P. Hardegree, and J. J. Mitchell. 2011. “Errors in LiDAR-Derived Shrub Height and Crown Area on Sloped Terrain.” *Journal of Arid Environments* 75 (4): 377–82. doi:10.1016/j.jaridenv.2010.11.005.
- Golzarian, M. R., Frick, R. A., Rajendran, K., Berger, B., Roy, S., Tester, M., & Lun, D. S. 2011. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods*, 7, 2. <http://doi.org/10.1186/1746-4811-7-2>
- Guan, H., Yu, J. and Luo, L., 2012. Random Forests-Based Feature Selection for Land-Use Classification Using LIDAR Data and Orthoimagery. *International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences*, 39, p.B7.
- Hall, S. A., I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker. 2005. “Estimating Stand Structure Using Discrete-Return Lidar: An Example from Low Density, Fire Prone Ponderosa Pine Forests.” *Forest Ecology and Management* 208 (1–3): 189–209. doi:10.1016/j.foreco.2004.12.001.
- He, Q., Chen, E., An, R., & Li, Y. (2013). Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest. *Forests*, 4(4), 984–1002. <http://doi.org/10.3390/f4040984>
- Hudak, Andrew T., Nicholas L. Crookston, Jeffrey S. Evans, David E. Hall, and Michael J. Falkowski. 2008. “Nearest Neighbor Imputation of Species-Level, Plot-Scale Forest Structure Attributes from LiDAR Data.” *Remote Sensing of Environment, Earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue*, 112 (5): 2232–45. doi:10.1016/j.rse.2007.10.009.
- Hudak, A., Evans, J., & Smith, A. 2009. LiDAR Utility for Natural Resource Managers. *USDA Forest Service / UNL Faculty Publications*. Retrieved from <http://digitalcommons.unl.edu/usdafsfacpub/202>.

- Heitschmidt, R. K., & Stuth, J. W. (Eds.). 1991. *Grazing Management: An Ecological Perspective*. Portland, Or: Timber Press, Incorporated.
- Jones, Hamlyn G., and Robin A. Vaughan. 2010. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. 1 edition. Oxford ; New York: Oxford University Press.
- Knick, Steven T. 1999. "Requiem for a Sagebrush Ecosystem?"
<https://research.wsulibs.wsu.edu:8443/xmlui/handle/2376/1169>.
- Laurin Vaglio, Gaia, Qi Chen, Jeremy A. Lindsell, David A. Coomes, Fabio Del Frate, Leila Guerriero, Francesco Pirotti, and Riccardo Valentini. 2014. "Above Ground Biomass Estimation in an African Tropical Forest with Lidar and Hyperspectral Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 89 (March): 49–58. doi:10.1016/j.isprsjprs.2014.01.001.
- Lefsky, Michael A., Warren B. Cohen, David J. Harding, Geoffrey G. Parker, Steven A. Acker, and S. Thomas Gower. 2002. "Lidar Remote Sensing of above-Ground Biomass in Three Biomes." *Global Ecology and Biogeography* 11 (5): 393–99. doi:10.1046/j.1466-822x.2002.00303.x.
- Li, Aihua, Nancy F. Glenn, Peter J. Olsoy, Jessica J. Mitchell, and Rupesh Shrestha. 2015. "Aboveground Biomass Estimates of Sagebrush Using Terrestrial and Airborne LiDAR Data in a Dryland Ecosystem." *Agricultural and Forest Meteorology* 213 (November): 138–47. doi:10.1016/j.agrformet.2015.06.005.
- Li, Peng, Luguang Jiang, and Zhiming Feng. 2013. "Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors." *Remote Sensing* 6 (1): 310–29. doi:10.3390/rs6010310.
- Luscier, Jason D., William L. Thompson, John M. Wilson, Bruce E. Gorham, and Lucian D. Dragut. 2006. "Using Digital Photographs and Object-Based Image Analysis to Estimate Percent Ground Cover in Vegetation Plots." *Frontiers in Ecology and the Environment* 4 (8): 408–13. doi:10.1890/1540-9295(2006)4[408:UDPAOI]2.0.CO;2.

- Marsett, Jianguo Qi. 2006. "Remote Sensing for Grassland Management in the Arid Southwest." *Rangeland Ecology & Management* 59 (5). doi:10.2111/05-201R.1.
- Miller, Richard F., Steven T. Knick, David A. Pyke, Cara W. Meinke, Steven E. Hanser, Michael J. Wisdom, and Ann L. Hild. 2011. "Characteristics of Sagebrush Habitats and Limitations to Long-Term Conservation." *Studies in Avian Biology* 38: 145–84.
- Mitchell, J J, and N F Glenn. 2009. "Subpixel Abundance Estimates in Mixture-Tuned Matched Filtering Classifications of Leafy Spurge (*Euphorbia Esula* L.)." *International Journal of Remote Sensing* 30 (23): 6099–6119.
- Mitchell, J J, N Glenn, T Sankey, D Derryberry, M Anderson, and R Hruska. 2011. "Small-Footprint Lidar Estimations of Sagebrush Canopy Characteristics." *Photogrammetric Engineering & Remote Sensing* 77 (5).
- Mitchell, Jessica J., Rupesh Shrestha, Lucas P. Spaete, and Nancy F. Glenn. 2015. "Combining Airborne Hyperspectral and LiDAR Data across Local Sites for Upscaling Shrubland Structural Information: Lessons for HypSIRI." *Remote Sensing of Environment*, Special Issue on the Hyperspectral Infrared Imager (HypSIRI), 167 (September): 98–110. doi:10.1016/j.rse.2015.04.015.
- Mitchell, J.J., Shrestha, R., Moore-Ellison, C.A. and Glenn, N.F., 2013. Single and multi-date Landsat classifications of basalt to support soil survey efforts. *Remote Sensing*, 5(10), pp.4857-4876.
- Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Mutanga, Onesimo, Elhadi Adam, and Moses Azong Cho. 2012. "High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random forest Regression Algorithm." *International Journal of Applied Earth Observation and Geoinformation* 18 (August): 399–406. doi:10.1016/j.jag.2012.03.012.

- Pal, M. 2005. "Random forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26 (1): 217–22.
doi:10.1080/01431160412331269698.
- Pierson, Frederick B., Peter R. Robichaud, Kenneth E. Spaeth, and Corey A. Moffet. 2003. *Impacts of Fire on Hydrology and Erosion in Steep Mountain Big Sagebrush Communities*. <http://www.treesearch.fs.fed.us/pubs/23541>.
- Polley, H. Wayne, Brian J. Wilsey, and Justin D. Derner. 2007. "Dominant Species Constrain Effects of Species Diversity on Temporal Variability in Biomass Production of Tallgrass Prairie." *Oikos* 116 (12): 2044–52.
doi:10.1111/j.2007.0030-1299.16080.x.
- Rengsirikul, Kannika, Assadaporn Kanjanakuha, Yasuyuki Ishii, Kunn Kangvansaichol, Prapa Sripichitt, Vittaya Punsuvon, Pilanee Vaithanomsat, Ganda Nakamane, and Sayan Tudsri. 2011. "Potential Forage and Biomass Production of Newly Introduced Varieties of *Leucaena* (*Leucaena Leucocephala* (Lam.) de Wit.) in Thailand." *Grassland Science* 57 (2): 94–100. doi:10.1111/j.1744-697X.2011.00213.x.
- Riaño, D., E. Chuvieco, S. L. Ustin, J. Salas, J. R. Rodríguez-Pérez, L. M. Ribeiro, D. X. Viegas, J. M. Moreno, and H. Fernández. 2007. "Estimation of Shrub Height for Fuel-Type Mapping Combining Airborne LiDAR and Simultaneous Color Infrared Ortho Imaging." *International Journal of Wildland Fire* 16 (3): 341–48.
- Richardson, Arthur J., and James H. Everitt. 1992. "Using Spectral Vegetation Indices to Estimate Rangeland Productivity." *Geocarto International* 7 (1): 63–69.
doi:10.1080/10106049209354353.
- Ritchie, J. C., J. H. Everitt, D. E. Escobar, T. J. Jackson, and M. R. Davis. 2006. "Airborne Laser Measurements of Rangeland Canopy Cover and Distribution." *Journal of Range Management Archives* 45 (2): 189–93.
- Roy, D.P., M.A. Wulder, T.R. Loveland, Woodcock C.E., R.G. Allen, M.C. Anderson, D. Helder, et al. 2014. "Landsat-8: Science and Product Vision for Terrestrial Global

- Change Research.” *Remote Sensing of Environment* 145 (April): 154–72.
doi:10.1016/j.rse.2014.02.001.
- Sala, O. E., and W. K. Lauenroth. 1982. “Small Rainfall Events: An Ecological Role in Semiarid Regions.” *Oecologia* 53 (3): 301–4. doi:10.1007/BF00389004.
- Sims, P.L. and Singh, J.S., 1978. The structure and function of ten western North American grasslands: II. intra-seasonal dynamics in primary producer compartments. *The Journal of Ecology*, pp.547-572.sa
- Schlesinger, William H., James F. Reynolds, Gary L. Cunningham, Laura F. Hueneke, Wesley M. Jarrell, Ross A. Virginia, and Walter G. Whitford. 1990. “Biological Feedbacks in Global Desertification.” *Science* 247 (4946): 1043–48.
doi:10.1126/science.247.4946.1043.
- Shinneman, D. J., W. L. Baker, and P. Lyon. 2008. “Ecological Restoration Needs Derived from Reference Conditions for a Semi-Arid Landscape in Western Colorado, USA.” *Journal of Arid Environments* 72 (3): 207–27.
doi:10.1016/j.jaridenv.2007.06.002.
- Shiple, Lisa A., Tara B. Davila, Nicole J. Thines, and Becky A. Elias. 2006. “Nutritional Requirements and Diet Choices of the Pygmy Rabbit (*Brachylagus Idahoensis*): A Sagebrush Specialist.” *Journal of Chemical Ecology* 32 (11): 2455–74.
doi:10.1007/s10886-006-9156-2.
- Shrestha, Gyami, and Peter D. Stahl. 2008. “Carbon Accumulation and Storage in Semi-Arid Sagebrush Steppe: Effects of Long-Term Grazing Exclusion.” *Agriculture, Ecosystems & Environment* 125 (1–4): 173–81. doi:10.1016/j.agee.2007.12.007.
- Storch, Ilse. 2007. “Conservation Status of Grouse Worldwide: An Update.” *Wildlife Biology* 13 (sp1): 5–12. doi:10.2981/0909-6396(2007)13[5:CSOGWA]2.0.CO;2.
- Streutker, David R., and Nancy F. Glenn. 2006. “LiDAR Measurement of Sagebrush Steppe Vegetation Heights.” *Remote Sensing of Environment* 102 (1–2): 135–45.
doi:10.1016/j.rse.2006.02.011.

- Su, J.G. and Bork, E.W., 2007. Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Applied Vegetation Science*, 10(3), pp.407-416.
- Ursino, N. 2007. "Modeling Banded Vegetation Patterns in Semiarid Regions: Interdependence between Biomass Growth Rate and Relevant Hydrological Processes." *Water Resources Research* 43 (4): W04412.
doi:10.1029/2006WR005292.
- Vierling, Lee A., Yanyin Xu, Jan U.H. Eitel, and John S. Oldow. 2012. "Shrub Characterization Using Terrestrial Laser Scanning and Implications for Airborne LiDAR Assessment." *Canadian Journal of Remote Sensing* 38 (06): 709–22.
doi:10.5589/m12-057.
- Waite, R.B., 1994. The application of visual estimation procedures for monitoring pasture yield and composition in exclosures and small plots. *Tropical Grasslands*, 28, pp.38-38.
- West, N. E. 1999. Managing for biodiversity of rangelands. *Biodiversity in agroecosystems*, 101-126.
- Wilson, Adam M., John A. Silander, Alan Gelfand, and Jonathan H. Glenn. 2011. "Scaling up: Linking Field Data and Remote Sensing with a Hierarchical Model." *International Journal of Geographical Information Science* 25 (3): 509–21. doi:10.1080/13658816.2010.522779.
- Wilcox, Bradford P. 2010. "Transformative Ecosystem Change and Ecohydrology: Ushering in a New Era for Watershed Management." *Ecohydrology* 3 (1): 126–30. doi:10.1002/eco.104.
- Wilson, Adam M., John A. Silander, Alan Gelfand, and Jonathan H. Glenn. 2011. "Scaling up: Linking Field Data and Remote Sensing with a Hierarchical Model." *International Journal of Geographical Information Science* 25 (3): 509–21.
doi:10.1080/13658816.2010.522779.
- Wisdom, M.J., M.M. Rowland, and R.J. Tausch. 2005. "Effective Management Strategies for Sage-Grouse and Sagebrush: A Question of Triage?" In *Transactions of the 70th North American Wildlife and Natural Resources Conference*, edited by R.D.

Sparrowe and L.H. Carpenter, 70:206. Arlington, Virginia, USA: Wildlife Management Institute.

http://www.fs.fed.us/rm/pubs_other/rmrs_2005_wisdom_m001.pdf.

Zandler, H., A. Brenning, and C. Samimi. 2015. "Quantifying Dwarf Shrub Biomass in an Arid Environment: Comparing Empirical Methods in a High Dimensional Setting." *Remote Sensing of Environment* 158 (March): 140–55.
doi:10.1016/j.rse.2014.11.007.

Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C. and Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote sensing of Environment*, 87(2), pp.171-182.

6 APPENDIX

6.1 An Application: Effect of Wildfire on Biomass And Cover

What follows below is a simplified approach to study the effect of wildfire on biomass and cover by using historical fire data for the NCA from 1957 to 2014, demonstrating one of many applications of this research. For this analysis, the biomass and cover maps obtained from the Landsat imputations were compared against fire frequency for the entire NCA. We studied the resulting relationships, individually for herbaceous and shrub.

6.1.1 Fire Data

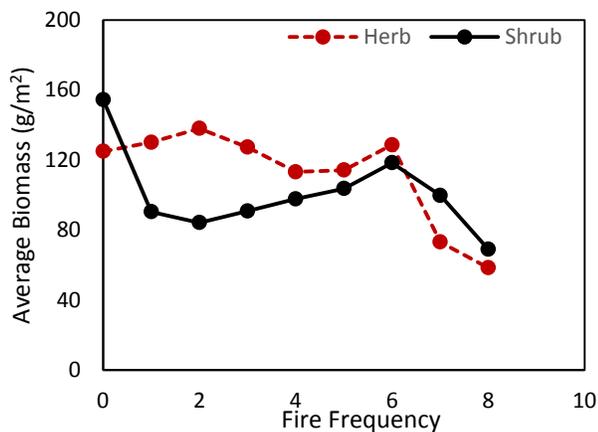
We used the US Geological Survey (USGS) historic fire perimeter data from 1957 to 2014 for the fire analyses. The data was compiled from GeoMAC (Geospatial Multi-Agency Coordination Group), MTBS (Monitoring Trends in Burn Severity) and USGS databases (see Balch et al. 2013). The fire analyses were performed using fire frequency, earliest recorded fire year and most recent fire year. For fire perimeters, we assumed that a given location could only burn once per year and hence removed any duplicate perimeters.

6.1.2 Relation with Fire Frequency

An analysis of imputed herb and shrub percent cover and biomass with the fire data revealed differences in the response of the vegetation with respect to fire. As shown in **figure A. 1** after each successive fire event, herbaceous communities were generally more efficient in recovering than shrub communities. Also, in the second and sixth fire

counts, the post-fire herb cover and biomass was found to surpass the pre-fire cover and biomass (see concave shape of graph, fig. A. 1). However, shrub was found to regenerate slower after successive fire events (convex shape in fig. A.1).

(a)



(b)

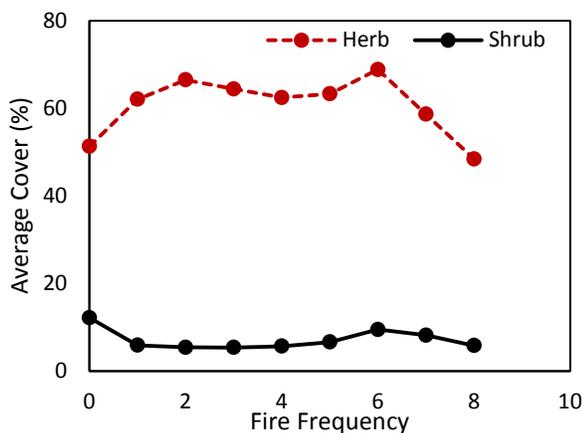


Figure A.1 Graph showing change in a) average biomass and b) average percent cover with respect to fire frequency in the study site. The biomass and cover value are computed from the respective imputed map.

We also analyzed historical fire events with *in-situ* field plots. The frequency distribution of fire events showed that among all 141 plots, 35% of plots were never

burned and about 1% of plots were burned 7 times (**Fig. A. 3**). None of the plots were burned 6 times.

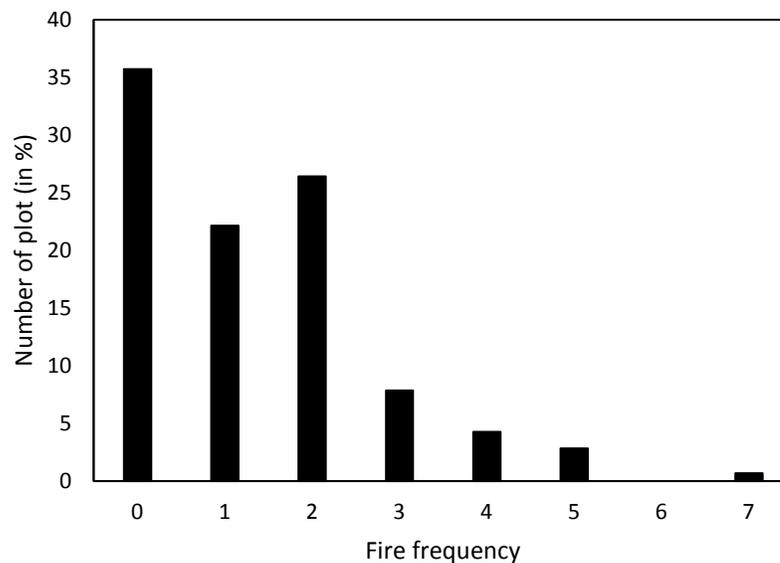
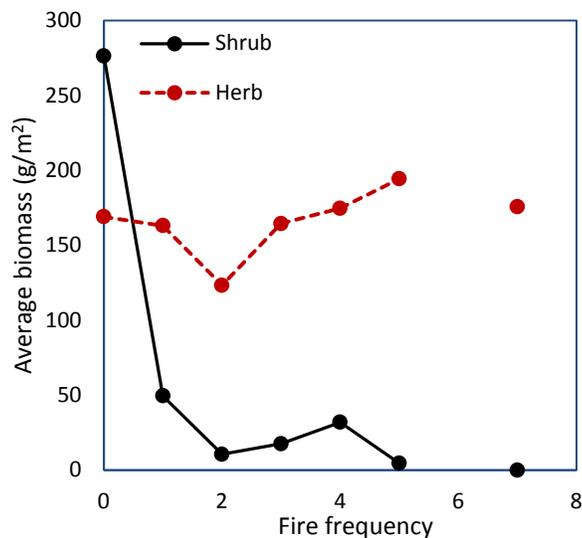


Figure A.2 Frequency distribution of plots with respect to fire events.

The *in-situ* field plots were first grouped with respect to their burn frequencies, and then the mean values of the vegetation characteristics were calculated for each plot. The *in-situ* values versus fire events revealed a similar pattern as the imputed results. As shown in **figure A. 3**, shrubs were found to be less resilient to regeneration compared to herbs after successive fire events.

(a)



(b)

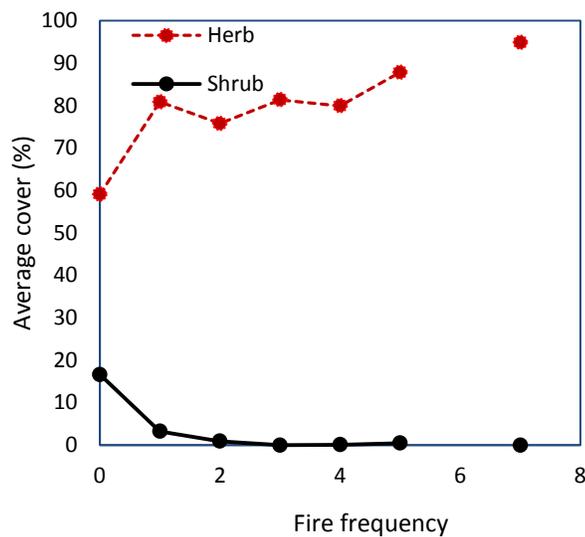


Figure A.3 Graph showing change in a) field average biomass and b) field average percent cover with respect to fire frequency. None of the field plots burned six times.

6.1.3 Implication of Fire Events

Herbaceous plants were more efficient at recovering from wildfire than shrubs in our field site. Studies have found non-native grasses to be resilient to fire and even

benefit from it, especially in areas of low rainfall (D'Antonio and Vitousek 1992; Balch et al. 2009; Taylor et al. 2014). Nonnative annuals like cheatgrass are a widespread herb community in North America and they are known to displace native shrub after subsequent fire (Chambers et al. 2007; Knapp 1996; Clinton et al. 2010). This is largely due to the "head start" they get for rapid growth in spring after the seed germination in fall or early winter. After the fire in summer, these annual exotic can again grow in fall whereas the perennial native shrubs like sagebrush need decades to recover (Klemmedson and Smith 1964). Moreover, increased fire occurrence, intensity and severity have been associated with these nonnative annual grasses because of their fine fuel biomass, increased flammability and high percent cover increasing fuel continuity and density (Whisenant 1990; Brooks et al. 2004; Balch et al. 2013; Davies and Nafus 2013). Whisenant (1990) found a fire return interval of 3-5 years in cheatgrass-dominated rangeland whereas the estimation for native sagebrush-dominated rangeland was 60-100 years. Balch et al. (2013) found cheatgrass-dominated rangelands to be nearly four times more prone to fire than native land cover. This positive feedback loop created by the vicious grass-fire circle replaces the native shrub in North American shrubland with non-native herbs.

Beside fire, climate change can be a major factor determining the growth and production of vegetation in drylands, especially for annuals (Watson et al. 1998; Stokes et al. 2008). Cheatgrass cover and biomass are promoted by wet and warm conditions during the fall and spring (Knapp 1998). Sala and Lauenroth (1982) demonstrated the ecological significance of precipitation in semiarid land and Fynn and O'Connor (2000) observed marked effect on variability, especially in herbaceous production. Smith et al.

(2000) showed incremental trend in above ground production of invasive grasses in arid climate by the atmospheric increase of CO₂.

The produced maps here were also studied in the context of historical events to determine a cursory effect of fire in biomass and cover. However, we cannot ignore the fact that the regeneration of vegetation is a highly complex phenomenon and doesn't necessarily depend on fire events only. Hence, further study specifically targeted towards the temporal change of plant communities is important to determine the major causes of fluctuations in percent cover and biomass in our study site.

6.2 Stepwise Regression Between Biomass and Lidar Metrics

Regression is the predictive modeling technique most broadly used for relating field and remotely sensed data (Hudak et al. 2008). In addition to RF variable selection and Ordinary Regression, Stepwise Regression (with backward elimination) were also performed separately for both total biomass (AGB) and shrub biomass as independent variables and 35 lidar derived vegetation metrics as the independent variables. The results from these two regression models provided an independent check on the RF result.

A stepwise regression also showed a high correlation between lidar metrics and total and shrub biomass. The regression was performed with backward elimination. We created a model with total biomass (Y variable) and only twelve lidar metrics (X variable) that had the highest correlation (Pearson's coefficient) with total biomass. Based on P-values we eliminated variables that performed poorly in the model until we had two of the most important variables and a satisfactory R². This gave us the following regression equation for total biomass with an R² of 0.76 and both P values and significance of $F \leq 0.05$:

AGB=109987.16 x Standard Deviation of Height -126780.89 x Mean Absolute Deviation from Height -74.63.

Similarly, the stepwise regression was performed for shrub biomass (SB) only and resulted in the following equation with an R^2 of 0.77 and both P values and significance of $F \leq 0.05$:

SB=137950.9 x Standard Deviation of Height 8 - 164885 x Mean Absolute Deviation from Height - 143.53. The negative shrub biomass shown in **figure A. 5** are likely artifacts from height filtering contributed by herbaceous plants, picked up by the lidar but not taken into account by shrub biomass model.

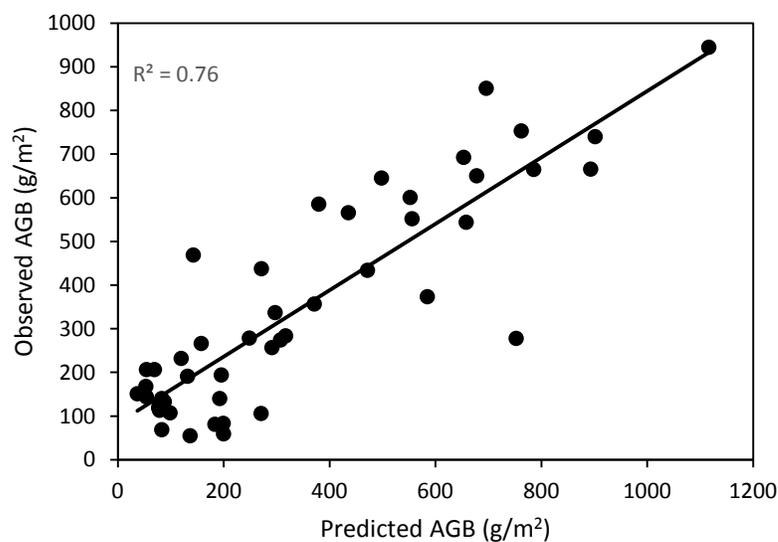


Figure A.4 Comparison between observed AGB and AGB predicted from stepwise regression.

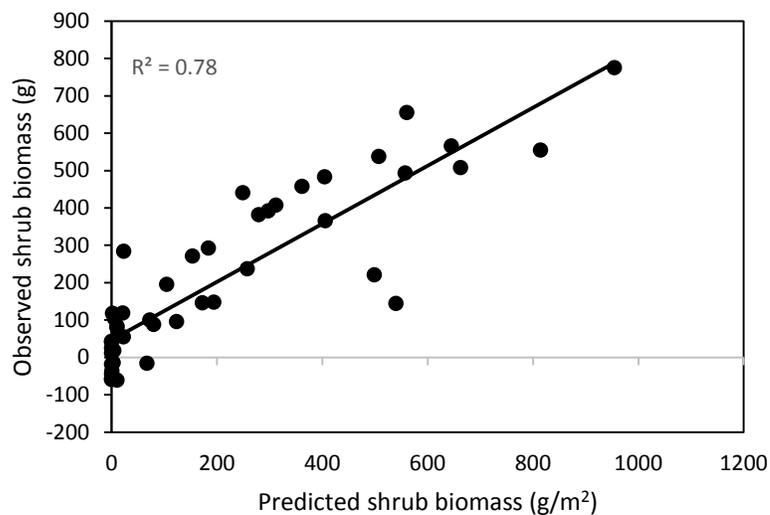


Figure A.5 Comparison between observed shrub biomass and shrub biomass predicted from stepwise regression.

The analysis of the residual showed a random pattern signifying homoscedasticity and normality of the data (**Fig. A. 6 and A. 7**).

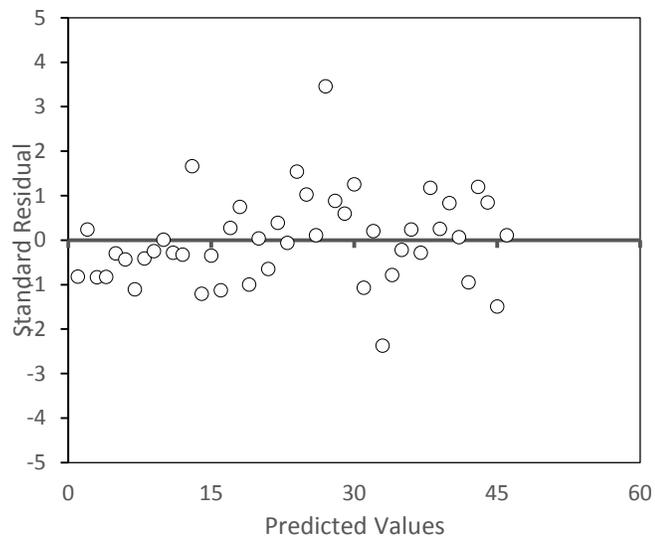


Figure A.6 Residual analysis of stepwise regression for AGB

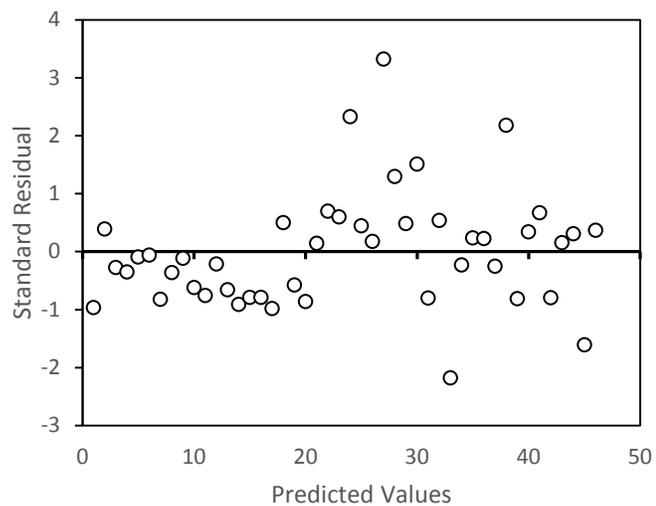


Figure A.7 Residual analysis of stepwise regression for shrub biomass

These results further support our conclusion that variables of height are important metrics to estimate biomass in rangeland.

6.3 R Code for Imputation

```
library(raster)

library(rgdal)

require(parallel)

library(randomForest)

library(snow)

library(yaImpute)

#This is the training data csv used in Salford Systems
  training <- read.csv("C:/test/summer_new.csv")

#select the column of the target variable
y <- subset(training, select = c(Sagebrush))

#select the columns of predictors
x <- subset(training, select = c(bB12, bB33, bB36))

  #the randomforest model. Include
rfMode="regression" for regression or leave out for
classification
sagebrush.rf <- yai(x=x, y=y, method="randomForest",
rfMode="regression", ntree= 2000)

  #this is where the output file is saved. Make sure name
matches the column name used in y
outfile <- list(Sagebrush = "C:/Sagebrush.asc")

  #Create a list of the ascii grids to be used for
the imputation. Make sure the names match the #column
names used in x
```

```
xfile <-list(bB12 = "C:/re_12.asc", bB33  
="C:/re_33.asc", bB36= "C:/re_36.asc")  
  
# The imputation. Make sure you the same name of the  
randomforest model is being called  
  
AsciiGridImpute(sagebrush.rf, xfile, outfile)
```

6.4 MATLAB Code for Calculating Vegetation Metrics

% This code reads LAS files in a folder and calculates the vegetation metrics (#35) with point cloud without using BCAL

LIDAR tool

```
clear all; close all;

%A=LASreadAll('43116C4405_ 3.las');

file=dir('H:\project\paperI\subsetLAS\2012_2013_100m\*.las'
);

for q=1:length(file)

    A=LASreadAll(file(q).name);

    classA=A.classification;

    vegA=A.pointSourceID;

    n=length(vegA);

    Zfactor=0.01;

    Z=vegA*Zfactor;

    vegZ=Z(classA==3);

    CT=0.15; %Crown threshold value

    GT=0.05; %Ground threshold value

    i=0;

    j=0;
```

```

nV=length(vegZ(vegZ>CT));%number of vegetation return
FLAG
nG=length(Z(Z<GT));%number of Ground return FLAG
VegDent=(nV/nG)*100; %FLAG
VegCover=(nV/n)*100;

%% For FHD
FHD=0;
for nb=1:20 %number of bins
    bht=0.1*nb; %bin height
    np(nb)=length(vegZ(vegZ<bht & vegZ>(nb-1)*0.1 &
vegZ> CT));
    pnp(nb)=(np(nb)/nV);
    if pnp(nb)~=0
        FHD=FHD+(pnp(nb)*log(pnp(nb)));
    end
end
FHD=-(FHD);

%% For FHD-Points above ground threshold
FHDGT=0;
for nb=1:(max(vegZ)/0.1) %alternate way to find the
number of bins
    bht=0.1*nb; %bin height
    np(nb)=length(vegZ(vegZ<bht & vegZ>(nb-1)*0.1 &
vegZ> GT));

```

```

    pnp(nb)=(np(nb)/nG);
    %if pnp(nb)~=0
        FHDGT=FHDGT+(pnp(nb)*log(pnp(nb)));%FHDGT-Points
above ground threshold
    end
    FHDGT=-(FHDGT);
    %For calculation of Percent of vegetaion in height
range
    p1=length(vegZ(vegZ<1&vegZ>0));
    p2=length(vegZ(vegZ<2.5&vegZ>1));
    p3=length(vegZ(vegZ>2.5 & vegZ<10));
    p4=length(vegZ(vegZ>10 & vegZ<20));
    p5=length(vegZ(vegZ>20 & vegZ<30));
    p6=length(vegZ(vegZ>30));

    MinH=min(vegZ);
    MaxH=max(vegZ);
    RangeH=MaxH-MinH;
    MeanH=mean(vegZ);
    MedianH=median(vegZ);
    MAD=1.4826*median(abs(vegZ-MedianH));
    AAD=mean(abs(vegZ-MeanH));
    VarH=var(vegZ);
    SDH=std(vegZ);
    SkewH=skewness(vegZ);
    KurtH=kurtosis(vegZ);

```

```
IQRH=iqr(vegZ);
Coeffvar=SDH/MeanH;
a=prctile(vegZ,5);
b=prctile(vegZ,10);
c=prctile(vegZ,25);
d=prctile(vegZ,50);
e=prctile(vegZ,75);
f=prctile(vegZ,90);
g=prctile(vegZ,95);
LiDAR_Return=n;%flag
CanRR=((MeanH-MinH)/(MaxH-MinH));
TH=std(vegZ(vegZ>GT & vegZ<CT));
GRet=(nG/n)*100; %flag

P1=(p1/n)*100;
P2=(p2/n)*100;
P3=(p3/n)*100;
P4=(p4/n)*100;
P5=(p5/n)*100;
P6=(p6/n)*100;

%saving all the matrices in one variable
%VegMat=zeros(q,35);%preallocating for speed
VegMat(q,1)=MinH;
```

VegMat (q, 2) =MaxH;
VegMat (q, 3) =RangeH;
VegMat (q, 4) =MeanH;
VegMat (q, 5) =MAD;
VegMat (q, 6) =AAD;
VegMat (q, 7) =VarH;
VegMat (q, 8) =SDH;
VegMat (q, 9) =SkewH;
VegMat (q, 10) =KurtH;
VegMat (q, 11) =IQRH;
VegMat (q, 12) =Coeffvar;
VegMat (q, 13) =a;
VegMat (q, 14) =b;
VegMat (q, 15) =c;
VegMat (q, 16) =d;
VegMat (q, 17) =e;
VegMat (q, 18) =f;
VegMat (q, 19) =g;
VegMat (q, 20) =LiDAR_Return;
VegMat (q, 21) =nV;
VegMat (q, 22) =nG;
VegMat (q, 23) =VegDent;
VegMat (q, 24) =VegCover;
VegMat (q, 25) =GRet;
VegMat (q, 26) =P1;
VegMat (q, 27) =P2;

```
VegMat(q,28)=P3;  
VegMat(q,29)=P4;  
VegMat(q,30)=P5;  
VegMat(q,31)=P6;  
VegMat(q,32)=CanRR;  
VegMat(q,33)=TH;  
VegMat(q,34)=FHD;  
VegMat(q,35)=FHDGT;
```

```
%exporting as an excel file
```

```
end
```

```
%xlswrite('vegmatrix.xlsx',VegMat);
```

```
%end
```

6.5 Abbreviation of the Vegetative Indices Used

Table A.1 Abbreviation of the VI used

Indices and Ancillary data	Formula
Greenness Condition Index(GI) / Simple Vegetation Index (SVI)	NIR/Red
Vegetation Condition Index(VCI)	MIR/NIR MIR=SWIR2
Normalized Difference Vegetation Index (NDVI)	NIR-Red/NIR+Red
Transformed Vegetation Index (TVI)	$\sqrt{(\text{NDVI}+0.5)}$
Differenced Vegetation Index (DVI)	NIR2-MIR MIR=SWIR2 NIR2=SWIR1
Mid-IR/Red Reflectance Index(MIRI)	MIR-Red/MIR+Red MIR=SWIR2 B7-B4/B7+B4
Tasseled cap brightness(BI)	0.3029Blue+0.2786Green+0.4733Red+0. 5599NIR+ 0.508SWIR1+0.1872SWIR2
Tasseled cap greenness(GVI)	-0.2941Blue-0.243Green- 0.5424Red+0.7276NIR-0.0713SWIR1- 0.1608SWIR2 -0.2941B2-0.243B3- 0.5424B4+0.7276B5-0.0713B6-0.1608B7
Tasseled cap wetness(WI)	0.1511Blue+0.1973Green+0.3283Red+0. 3407NIR-0.7117SWIR1-0.4559SWIR2
Soil Adjusted Vegetation Index(SAVI)	$[(\text{NIR1}-\text{Red}) / (\text{NIR1}+\text{Red}+\text{L})] \times (1+\text{L})$ L=Soil Correction factor

	$[(B5-B4)/(B5+B4+0.25)] \times (1+0.25)$
Green Soil Adjusted Vegetation Index (GSAVI)	$[(NIR-green)/(NIR+green+L)] \times (1+L)$ $[(B5-B3)/(B5+B3+0.25)] \times (1+0.25)$
Modified Soil Adjusted Vegetation Spectral Index (MSAVI)	$[2 \times NIR+1-\sqrt{((2 \times NIR+1)^2 - 8 \times (NIR - Red))}]/2$ $[2 \times B5+1-\sqrt{((2 \times B5+1)^2 - 8 \times (B5 - B4))}]/2$
Soil Adjusted Total Vegetation Index (SATVI)	$[(NIR1-Red)/(NIR1+Red+L)] \times (1+L) - NIR2/2$ $[(B5-B4)/(B5+B4+0.25)] \times (1+0.25) - B6/2$ L=Soil Correction factor NIR2=SWIR1
Stress-related Vegetation Index 1 (STVI-1)	MIR x Red/NIR MIR=SWIR2
Moisture Stress Index(MSI)	SWIR/NIR B6/B5
Moisture Stress Index2(MSI2)	SWIR2/NIR B7/B5
Normalized Difference Water Index(NDWI)	$(NIR-SWIR)/(NIR+SWIR)$ $(B5-B6)/(B5+B6)$
Normalized Difference Water Index2(NDWI2)	$(NIR-SWIR2)/(NIR+SWIR2)$
Normalized Burn Ratio2 (NBR)	$(SWIR1-SWIR2)/(SWIR1+SWIR2)$ $(B6-B7)/(B6+B7)$