COMPLEXITY AND DYNAMICS OF SEMI-ARID VEGETATION STRUCTURE, FUNCTION AND DIVERSITY ACROSS SPATIAL SCALES FROM FULL WAVEFORM LIDAR

by

Ginikanda Yapa Mudiyanselage Nayani Thanuja Ilangakoon

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Geosciences
Boise State University

May 2020

© 2020

Ginikanda Yapa Mudiyanselage Nayani Thanuja Ilangakoon

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Ginikanda Yapa Mudiyanselage Nayani Thanuja Ilangakoon

Dissertation Title: Complexity and Dynamics of Semi-arid Vegetation Structure, Function and Diversity Across Spatial Scales from Full Waveform Lidar

Date of Final Oral Examination: 20 February 2020

The following individuals read and discussed the dissertation submitted by student Ginikanda Yapa Mudiyanselage Nayani Thanuja Ilangakoon, and they evaluated her 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
Shawn Benner, Ph.D.	Member, Supervisory Committee
Jodi Brandt, Ph.D.	Member, Supervisory Committee
Dylan Mikesell, Ph.D.	Member, Supervisory Committee

The final reading approval of the dissertation was granted by Nancy F. Glenn, Ph.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

DEDICATION

This dissertation is dedicated to my wonderful daughter Amelia and to my Appachchi (Dad), the strength of my life.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and gratitude to my advisor Dr. Nancy F. Glenn for her guidance, support, and patience throughout this journey.

I would also like to convey sincere thanks to the dissertation committee: Dr. Shawn Benner, Dr. Jodi Brandt, and Dr. Dylan Mikesell for their enormous support and constructive feedback during my PhD.

I would like to thank the NASA Terrestrial Ecology (NASA TE -NNX14AD81G) and the NASA Earth and Space Science Fellowship (NESSF - 17-EARTH17F-0209) programs for supporting me with funding.

I sincerely appreciate Dr. Thomas Painter, Dr. Cathie Olschanowsky, Dr. Fabian Schneider, Dr. Steven Hancock, and Dr. Tristan Goulden, for their invaluable support during my PhD including journal publications.

Thank you to USDA ARS Northwest Watershed Research Center for access and support to Reynolds Creek Experimental Watershed (RCEW) where this study took place.

I would like to thank all BCAL group members for their generous support during my PhD.

I would also like to thank Boise State Research Computing for their enormous support and co-operation with me when I use many cores of the cluster at the same time.

I convey my gratitude to the faculty, staff, and students of the Department of Geosciences for their help during my stay at Boise State University.

Finally, I extend my heartfelt thanks to my family and friends who always stood by me during this journey.

ABSTRACT

Semi-arid ecosystems cover approximately 40% of the earth's terrestrial landscape and show high dynamicity in ecosystem structure and function. These ecosystems play a critical role in global carbon dynamics, productivity, and habitat quality. Semi-arid ecosystems experience a high degree of disturbance that can severely alter ecosystem services and processes. Understanding the structure-function relationships across spatial extents are critical in order to assess their demography, response to disturbance, and for conservation management. In this research, using state-of-the-art full waveform lidar (airborne and spaceborne) and field observations, I developed a framework to assess the complexity and dynamics of vegetation structure, function and diversity across spatial scales in a semi-arid ecosystem.

Difficulty in differentiating low stature vegetation from bare ground is the key remote sensing challenge in semi-arid ecosystems. In this study, I developed a workflow to differentiate key plant functional types (PFTs) using both structural and biophysical variables derived from the full waveform lidar and an ensemble random forest technique. The results revealed that waveform lidar pulse width can clearly distinguish shrubs from bare ground. The models showed PFT classification accuracy of 0.81–0.86% and 0.60–0.70% at 10 m and 1 m spatial resolutions, respectively. I found that structural variables were more important than the biophysical variables to differentiate the PFTs in this study area. The study further revealed an overlap between the structural features of different PFTs (e.g. shrubs from trees).

Using structural features, I derived three main functional traits (canopy height, plant area index and foliage height diversity) of shrubs and trees that describe canopy architecture and light use efficiency of the ecosystem. I evaluated the trends and patterns of functional diversity and their relationship with non-climatic abiotic factors and fire disturbance. In addition to the fine resolution airborne lidar, I used simulated large footprint spaceborne lidar representing the newly launched Global Ecosystem Dynamics Investigation system (GEDI, a lidar sensor on the International Space Station) to evaluate the potential of capturing functional diversity trends of semi-arid ecosystems at global scales. The consistency of diversity trends between the airborne lidar and GEDI confirmed GEDI’s potential to capture functional diversity. I found that the functional diversity in this ecosystem is mainly governed by the local elevation gradient, soil type, and slope. All three functional diversity indices (functional richness, functional evenness and functional divergence) showed a diversity breakpoint near elevations of 1500 m – 1700 m. Functional diversity of fire-disturbed areas revealed that the fires in our study area resulted in a more even and less divergent ecosystem state. Finally, I quantified aboveground biomass using the structural features derived from both the airborne lidar and GEDI data. Regional estimates of biomass can indicate whether an ecosystem is a net carbon sink or source as well as the ecosystem’s health (e.g. biodiversity). Further, the potential of large footprint lidar data to estimate biomass in semi-arid ecosystems are not yet fully explored due to the inherent overlapping vegetation responses in the ground signals that can be affected by the ground slope. With a correction to the slope effect, I found that large footprint lidar can explain 42% of variance of biomass with a RMSE of 351 kg/ha (16% RMSE). The model estimated 82% of the study area with less than 50%

uncertainty in biomass estimates. The cultivated areas and the areas with high functional richness showed the highest uncertainties. Overall, this dissertation establishes a novel framework to assess the complexity and dynamics of vegetation structure and function of a semi-arid ecosystem from space. This work enhances our understanding of the present state of an ecosystem and provides a foundation for using full waveform lidar to understand the impact of these changes to ecosystem productivity, biodiversity and habitat quality in the coming decades. The methods and algorithms in this dissertation can be directly applied to similar ecosystems with relevant corrections for the appropriate sensor. In addition, this study provides insights to related NASA missions such as ICESat-2 and future NASA missions such as NISAR for deriving vegetation structure and dynamics related to disturbance.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGEMENTS v

ABSTRACT vii

TABLE OF CONTENTS x

LIST OF TABLES xiv

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xx

CHAPTER ONE: INTRODUCTION 1

CHAPTER TWO: CONSTRAINING PLANT FUNCTIONAL TYPES IN A SEMI-ARID ECOSYSTEM WITH WAVEFORM LIDAR 7

- Abstract 7
- Introduction 8
- Materials 14
  - Study area 14
  - Field data 15
  - Small-footprint waveform lidar data 16
- Methods 17
  - Decomposition of waveform lidar signals 17
  - Additional waveform features derived from Gaussian decomposition 19
  - Frequency-domain deconvolution of lidar waveforms 21

Plant functional types classification....26

Implementation of random forest classification....26

Results....30

Influence of the viewing geometry on waveform features....30

Important waveform features for PFT classification....30

RF model performance....34

Discussion....35

Differentiating PFTs with waveform features....35

Waveform derived heights and differentiating bare ground from vegetation....41

Scale dependency of PFT classification....42

Upscaling to large-footprint waveforms....43

Conclusions....45

CHAPTER THREE: SPACEBORNE LIDAR REVEALS TRENDS AND PATTERNS OF FUNCTIONAL DIVERSITY IN A SEMI-ARID ECOSYSTEM....46

Abstract....46

Introduction....47

Materials and Methods....50

Study area....50

Field data....51

Environmental data....52

Airborne lidar data....52

Derivation of bare ground....53

ASO Functional traits....54

GEDI waveform simulation and functional traits 56
Functional diversity 58
Statistical analysis 59
Results 60
Discussion 76
Mapping functional traits 76
Functional diversity in semi-arid ecosystems 78
Conclusion 82
CHAPTER FOUR: ESTIMATING ABOVEGROUND BIOMASS IN A SEMI-ARID ECOSYSTEM FROM LARGE FOOTPRINT LIDAR DATA: INSIGHTS FOR GEDI .83
Abstract 83
Introduction 84
Methods 88
Study area 88
Field data 89
Airborne lidar data (ASO) 90
GEDI simulation 90
Reference biomass 92
GEDI biomass model development 93
Results 96
Discussion 103
Conclusion 109
CHAPTER FIVE: CONCLUSION 110
REFERENCES 117

APPENDIX A 137

- Supplementary figures for chapter 3 137
- Supplementary figures for chapter 4 142

LIST OF TABLES

Table 2.1 Summary of waveform features derived from individual waveforms and their biophysical relationships to the target. 11

Table 2.2 Features extracted from waveform backscatter lidar. 25

Table 2.3 Ensemble random forest (E-RF) models used to evaluate the waveform attribute selection and PFT classification. 29

Table 2.4 Producer and user accuracy of each PFT (ASP, DF, JP, BT, SG, and GD) in each RF model described in Table 3.. 35

LIST OF FIGURES

Figure 2.1 Reynolds Creek Experimental Watershed study area with field sample locations (n=103 plots) of plant functional types (ASP, DF, JP, BT, SG, and GD) and waveform lidar trajectories. Field photos depict the sparse to dense shrub and tree communities (from top to bottom photo). 15

Figure 2.2 Workflow illustrating the steps used to derive waveform features from the raw backscattered lidar waveform. Processing includes georeferencing, data alignment, Gaussian decomposition, deconvolution, and calibration. 18

Figure 2.3 Illustration of information contained in a lidar waveform. Peak 1, 2, and 3 are the echoes from three scatterers detected by the waveform. Three Gaussian functions (Gaussian pulse 1, 2, and 3) were fitted to the raw waveform. The peak amplitude is the maximum amplitude of the first echo after Gaussian fitting. Pulse width is the full width at half maximum. The stars are the trigger locations of each echo. Leading edge is the time distance from trigger to the max amplitude. The backscattered area represents the scattered cross-section from the first echo........................ 24

Figure 2.4 Ensemble random forest PFT classification workflow. Feature selection and base classifiers were trained using training samples. The selected base classifier models were applied to the validation samples and ensembled to make the final decision. 28

Figure 2.5 Box plots showing the variability of values of the most important waveform features at 1 m scale among PFTs. Definitions of PFTs are: ASP-aspen, DF-Douglas fir, JP-juniper, BT-bitterbrush, SG-sagebrush, GD-bare ground. From left to right, top row: variability of standard deviation of 90th percentile energy, 10th percentile height, and 75th percentile energy, respectively; middle row: variability of pulse width of first returns (first width), 50th percentile energy, and pulse widths of all returns (total width), respectively; bottom row: variability of standard deviation of backscatter coefficient of first returns (first γ), rise time, and 90th percentile energy, respectively. Note the differentiation of shrubs and bare ground with the pulse width and rise time. 32

Figure 2.6 Box plots showing the variability of values of the most important waveform features at 10 m scale among PFTs. From left to right, top row: variability of pulse width of first return, rise time, and standard deviation of total backscattering coefficients, respectively; middle row: variability of energy at 50th percentile height, and standard deviation of first return pulse widths, and standard deviation of target cross-section, respectively; bottom row: variability of kurtosis of first return amplitudes and skewness of last return amplitudes, respectively. 34

Figure 2.7 Distribution of total backscattered coefficient (total γ) and target cross-section (σ) of PFTs at 1 m (a) and 10 m (b) spatial resolution. Total backscatter coefficient shows less variability than the target cross section within PFTs as well as among PFTs. 38

Figure 2.8 Distribution of amplitude of the selected PFTs at 1 m (a) and 10 m (b) spatial scale. Distribution of heights above ground of the PFT samples at 1 m (c) and 10 m (d) spatial resolution. 40

Figure 3.1 Reynolds Creek Experimental Watershed, SW Idaho with the topographic gradient and stream network. The black stars represent 10 m x 10 m field plots across the watershed. 51

Figure 3.2 Summary workflow for deriving plant area index from small footprint waveform lidar. 54

Figure 3.3 Correlation between a) field measured max vegetation heights and ASO derived vegetation heights, b) field measured plant area index (PAI) and ASO derived PAI, c) field measured max vegetation heights and GEDI derived vegetation heights, and d). field measured PAI and GEDI derived PAI 61

Figure 3.4 Correlation between a) field measured max vegetation heights and ASO derived PAI, b) ASO derived FHD and ASO derived PAI, c) field measured max vegetation heights and GEDI derived PAI, and d). GEDI derived FHD and GEDI derived PAI 62

Figure 3.5 ASO and GEDI derived functional traits distribution across the RCEW. Top row: a). ASO derived PAI, b). ASO derived FHD, c). ASO derived CH. Bottom row: d). GEDI derived PAI, e). GEDI derived FHD, f). GEDI derived CH. The ASO based maps were derived at 10 m spatial resolution. GEDI data are displayed at footprint scale. The ASO and GEDI canopy heights are displayed as log canopy heights to enhance the visualization of canopy height distribution across the study area. 65

Figure. 3.6 Functional diversity derived using 500 m spatial neighborhood from ASO (top row- richness, evenness, divergence), and GEDI (bottom row- richness, evenness, divergence). Functional richness, functional evenness and functional divergence of RCEW derived from functional traits distribution across 500 m neighborhood to each 10 m x 10 m pixel in ASO and 25 m x 25 m footprints in GEDI. 66

Figure. 3.7 Correlation between ASO and GEDI based diversity indices at 500 m spatial neighborhood. 67

Figure. 3.8 Functional diversity – environmental gradient trends of RCEW. The red and blue curves represent the mean variation of diversity indices from ASO and GEDI respectively and the surrounding gray area represents the standard deviation. From top to bottom variation of functional diversity with altitude (a,b,c), aspect (d,e,f), slope (g,h,i), topographic wetness index (TWI) (j,k,l), and distnace to water (m,n,o) are displayed. 70

Figure. 3.9Variance of functional diversity explained by each abiotic factor; a). ASO functional richness, b). ASO functional evenness, c). ASO functional divergence, d). GEDI functional richness, e). GEDI functional evenness, f). GEDI functional divergence. TWI – topographic wetness index, DTW – distance to the nearest stream in meters. 70

Figure. 3.10 Frequency distribution of functional traits within the fire disturbed and surrounding undisturbed areas of RCEW. The pink represents the functional traits from burned areas while the green represents the functional traits of undisturbed areas. 72

Figure. 3.10 Frequency distribution of functional diversity indices of burned and unburned areas of RCEW. The pink represents the functional diversity from burned areas while the green represents the functional diversity of undisturbed areas. 73

Figure. 3.11 Functional diversity variability with search neighborhood radius at four different burned regions. Pink represents the burned areas and the blue represents the surrounding unburned areas. 75

Figure. 4.1 Assimilation of vegetation signal with ground signal in large footprint lidar data over sloped terrain. 87

Figure. 4.2 a) Reynolds Creek Experimental Watershed (RCEW) study area vegetation. b) ASO derived reference biomass distribution. The red dots in a) represents the spatial distribution of field plots. 89

Figure. 4.3 Processing workflow of GEDI biomass and biomass uncertainty estimates. 95

Figure. 4.4 ASO point cloud derived a) maximum vegetation heights and, b) percent vegetation cover maps at 5 m resolution. 96

Figure. 4.5 Correlation coefficients between slope corrected GEDI RH metrics and field observed max vegetation heights. 97

Figure. 4.6 Reference ASO biomass versus predicted GEDI footprint biomass (kg/ha) with the 1:1 line. 99

Figure. 4.7 a) ASO reference biomass with vegetation above 3m masked and b) predicted GEDI footprint biomass. 100

Figure. 4.8 a) GEDI footprint level biomass; and b) prediction uncertainty from the 95% credible interval; and c) percent uncertainty. 101

Figure. 4.9 ASO and GEDI biomass comparison including a) reference ASO biomass resampled to 1 km; b) predicted GEDI biomass resampled to 1 km; and c) biomass differences between reference and GEDI at 1 km. 102

Figure. 4.10 Uncertainty of biomass with upscaling to 1 km including a) predicted GEDI footprint level biomass; b) range of biomass within each 1 km pixel and c) standard deviation of biomass at 1 km pixels derived from the 25 m footprint biomass. 103

Figure. 4.11 A sample simulated GEDI waveform with the reference ASO waveform and the cumulative energy profile. The reference ASO waveform is contained within the GEDI RH16 and RH68. 104

Figure. 4.12 Correlation of percent uncertainty with mean footprint vegetation height and percent vegetation cover. 107

Figure. 4.13 GEDI footprint density and standard deviation of biomass at 1 km pixels. 108

Figure S.1 Boxplots of CH, PAI, and FHD derived from GEDI of burned and surrounding unburned areas of Koke fire (a, b, c), Whiskey fire (d, e, f), Break fire (g, h, i), and Rabbit Creek fire (j, k, l). 138

Figure S.2 Boxplots of functional richness, functional evenness, and functional divergence calculated using GEDI functional traits of burned and surrounding unburned areas of Koke fire (a, b, c), Whiskey fire (d, e, f), Break fire (g, h, i), and Rabbit Creek fire (j, k, l). The first column represents functional richness while the second and third columns represent functional evenness and functional divergence respectively. 139

Figure S.3 GEDI footprints colored by functional diversity of the four different fires studied. The first column represents functional richness while the second and third columns represent functional evenness and functional divergence respectively. 140

Figure S.4 RCEW precipitation and temperature variation with elevation gradient. Precipitation: average for the period of 1963-2010, temperature: mean normal air temperature for the period of 1984 – 2014 (PRISM Climate Group, 2016, CZO Dataset: Reynolds Creek, 2016) 141

Figure S.5 Frequency distribution of a) predicted GEDI biomass and b) percent uncertainty at GEDI footprint scale. 142

Figure S.6 Variance of predicted footprint biomass and percent uncertainty explained by Slope, Vegetation heights and percent vegetation cover. 142

LIST OF ABBREVIATIONS

ASO	Airborne Snow Observatory
AGB	Aboveground biomass
ALS	Airborne laser scanning
ASP	Aspen
BT	Bitterbrush
CH	Canopy height
CHM	Canopy height models
DEM	Digital elevation models
DTW	Distance to water
FHD	Foliar height diversity
GD	Bare ground
GEDI	Global Ecosystem Dynamic Investigation
DF	Douglas fir
JP	Juniper
PAI	Plant Area Index
PFT	Plant functional types
NASA	National Aeronautics and Space Administration
SG	Sagebrush
TWI	Topographic wetness index

CHAPTER ONE: INTRODUCTION

Over the last few decades, land degradation has become a critical challenge for terrestrial ecosystems. Climate and human driven disturbances modify the structure and function of natural ecosystems. Altered ecosystem structure and function provide adverse effects on ecosystem services and processes including productivity, biodiversity and habitat quality. Understanding the structure and function of global terrestrial ecosystems improves the understanding of their interactions with the biosphere, atmosphere, and hydrosphere including the cycling of the major biogeochemical elements and water (Diaz et al., 2007; Dietze et al., 2017). Further, assessing the effects of climate and human driven changes at different levels provide necessary information for national and international policy discussion around mitigation targets (Arnell, Lowe, Challinor, & Osborn, 2019; Rödig et al., 2019). Quantitative assessments of ecosystem structure, function and their spatial diversity at regional to global scales are fundamental to monitor the ecosystem state, and the impact to the atmosphere, biosphere, and hydrosphere under changing conditions. Among others, semi-arid ecosystems experience a high degree of land degradation (Fusco, Rau, Falkowski, Filippelli, & Bradley, 2019) due to both climatic (drought, fire, invasion and encroachment, erosion etc.) and anthropogenic (grazing, land use, agriculture) disturbances. Semi-arid ecosystems cover approximately 40% of the global terrestrial surface and are home to about 20% of the world's population (Li et al., 2018; (Nautiyal, Bhaskar, & Khan, 2015). These ecosystems are typically heterogeneous, low-stature, and sparsely vegetated. Semi-arid ecosystems comprise a range of intra and inter species structural and functional characteristics. These heterogeneous vegetation characteristics provide habitat and biodiversity to unique fauna and flora as well as two billion people worldwide (Nautiyal, Bhaskar, & Khan, 2015). Further, semi-arid ecosystems play a critical role in global carbon dynamics (Ahlstrom et al., 2015; Poulter et al., 2015) and show that afforestation could offset the climate warming effects and cool the planet (Yosef et al., 2018).

Availability of remote sensing data at fine to coarse spatial and temporal scales facilitates monitoring the retrospective and prospective states of ecosystems across spatial scales needed for ecosystem service management (Abelleira Martínez et al., 2016). Importantly, waveform lidar, which digitizes the total amount of lidar return energy at high vertical resolution (~1 ns = 15 cm) provides unprecedented opportunities to accurately quantify ecosystem structure and function at local to regional scales (Hovi, Korhonen, Vauhkonen, & Korpela, 2016; Qi, Lee, et al., 2019; Yao, Krzystek, & Heurich, 2012). With the launch of the Global Ecosystem Dynamics Investigation (GEDI) mission, we have new opportunities to map functional types, traits and diversity at global scales (Duncanson et al., 2019; Qi, Lee, et al., 2019; Qi et al., 2019; Rödig et al., 2019).

The abundance and distribution of plant functional types (PFTs) are important indicators for monitoring ecosystem state, as well as its resistance and resilience to climate and human driven disturbances (Lavorel, McIntyre, Landsberg, & Forbes, 1997; Poulter et al., 2015; Schimel, Asner, & Moorcroft, 2013). Thus, PFTs are frequently used as inputs for vegetation dynamics and earth system models (Krinner et al., 2005; Sitch et al., 2003; Wullschleger et al., 2014). However, uncertainty in PFTs, especially in semi-arid ecosystems between shrub, grass and forest classes reduces the accuracy of these models (Hartley, MacBean, Georgievski, & Bontemps, 2017). In semi-arid ecosystems, the influence of soil background on the remote-sensing signals is a major challenge. Improved methods to capture plant functional types (PFTs) in semi-arid ecosystems are needed to accurately assess the ecosystem state.

A wealth of research has shown that functional traits are the best representatives of ecosystem processes (Bardgett & van der Putten, 2014; Hooper et al., 2006). Research evidence further indicates that, though net primary productivity (NPP), nutrient retention, and disturbance regimes can describe facets of ecosystem functioning, none of these variables can directly quantify the observed diversity in ecosystem functioning (Gough et al., 2016). Moreover, disturbance-driven alterations and their ecological impacts are highly dynamic in space and time. Morphological, physiological and phenological traits within and between species of an ecosystem can represent the ecosystem demography and response strategies to the disturbances (Serbin et al., 2019). Thus, remotely sensed functional traits and their diversity are widely utilized in forested ecosystems to predict variations in ecosystem structure – function relationships (Funk et al., 2017, Wieczynski et al., 2019). Yet, several gaps remain in our understanding of how the complexity and dynamics of functional diversity in semi-arid ecosystems vary with respect to the environmental gradient and in response to disturbance, especially post fire.

Another important vegetation functional trait is the canopy aboveground biomass (ABG). AGB serves to characterize, quantify, understand, and predict whether ecosystems are a net carbon sink or source (Duncanson et al., 2019; Li et al., 2015; Qi et al., 2019). Hence, accurate estimates of ABG at regional to global scales improves the understanding of carbon fluxes associated with the ecosystem and provides significant implications to constrain global vegetation/carbon dynamics. AGB can further help assess ecosystem health including biodiversity. Fusco, Rau, Falkowski, Filippelli, & Bradley, (2019) demonstrated that both shrublands and woodlands account for significant carbon storage, especially in semi-arid ecosystems. Ahlström et al. (2015) showed that semi-arid ecosystems control inter-annual variability of global carbon. Nevertheless, estimating ABG from remote sensing data, especially over semi-arid ecosystems at regional scale has been a long-standing challenge due to the short canopies and their sparse distribution in space.

The overarching goal of this dissertation is to develop novel remote sensing-based methods for and to understand the complexity and dynamics of the vegetation structure, function and diversity across spatial scales in a semi-arid ecosystem. To address this, three main research questions are considered, including:

1. **How can key plant functional types including shrubs, trees, and bare ground be differentiated using state-of-the-art full waveform lidar data?**
2. **What are the trends and patterns of functional diversity in the study area and their abiotic controls?**
    a. **What is the potential of the newly launched GEDI, the spaceborne lidar system to capture functional diversity trends in a semi-arid ecosystem?**
3. **What is the uncertainty of regional AGB estimates in semi-arid ecosystems using the GEDI system?**

For a test study area, I used the Reynolds Creek Experimental Watershed (RCEW), a semi-arid ecosystem of approximately 270 km^2 within the Great Basin ecoregion in the Western US. The RCEW has a range of topography (1100 m – 2200 m) and a diverse vegetation community. While the unique and important sagebrush-steppe with many grass and forbs dominates the low elevations, tree communities mark the high elevations. In addition, riparian vegetation with cottonwood and willow are found within valleys, and along streams across the watershed. The study area is further characterized by a mean annual temperature and precipitation that varies between 4.6–9.2 ° C and 230-959 mm, along the elevation gradient, respectively. The study area has experienced prescribed and natural fires and supports grazing. Consequently, invasion of cheatgrass in native shrub areas and juniper encroachment have occurred in this study area during the last few decades. The diverse topography, vegetation, and disturbance followed by invasion history of the study area provided a unique setting to elucidate the main research questions of this dissertation.

To answer the research questions, I first developed a novel methodological workflow for state-of-the-art full waveform lidar to differentiate key plant functional types. In this, I decoded structural and biophysical characteristics of vegetation and bare ground embedded in the lidar signal (Chapter 2). Then, I investigated the relationships between both functional diversity and environmental gradients (altitude, slope, aspect, topographic wetness index, and distance to water) and functional diversity and disturbance relationships (e.g. fire). This work focuses on understanding the ecosystem demography and response strategies to disturbance (Chapter 3). Finally, I estimated the uncertainty in assessing the AGB of this heterogeneous, low-stature, semi-arid ecosystem (Chapter 4).

In this, I used spatially explicit vegetation structure derived from simulated GEDI lidar, especially in support of new measurement capabilities for satellite missions and global vegetation/carbon dynamics.

CHAPTER TWO: CONSTRAINING PLANT FUNCTIONAL TYPES IN A SEMI-ARID ECOSYSTEM WITH WAVEFORM LIDAR

This chapter has been published as:

Ilangakoon, N. T., Glenn, N. F., Dashti, H., Painter, T. H., Mikesell, T. D., Spaete, L. P., Jessica J. Mitchell, & Shannon, K. (2018). Constraining plant functional types in a semi-arid ecosystem with waveform lidar. *Remote Sensing of Environment, 209*, 497-509.

Abstract

Accurate classification of plant functional types (PFTs) reduces the uncertainty in global biomass and carbon estimates. Airborne small-footprint waveform lidar data are increasingly used for vegetation classification and above-ground carbon estimates at a range of spatial scales in woody or homogeneous grass and savanna ecosystems. However, a gap remains in understanding how waveform features represent and ultimately can be used to constrain the PFTs in heterogeneous semi-arid ecosystems. This study evaluates lidar waveform features and classification performance of six major PFTs, including shrubs and trees, along with bare ground in the Reynolds Creek Experimental Watershed, Idaho, USA. Waveform lidar data were obtained with the NASA Airborne Snow Observatory (ASO). From these data we derived waveform features at two spatial scales (1 m and 10 m rasters) by applying a Gaussian decomposition and a frequency-domain deconvolution. An ensemble random forest algorithm was used to assess classification performance and to select the most important waveform features. Classification models developed with the 10 m waveform features outperformed those at 1 m (Kappa (κ) = 0.81–0.86 vs. 0.60–0.70, respectively). At 1 m resolution, lidar height features improved the PFT classification accuracy by 10% compared to the analysis without these features. However, at 10 m resolution, the inclusion of lidar derived heights with other waveform features decreased the PFT classification performance by 4%. Pulse width, rise time, percent energy, differential target cross section, and radiometrically calibrated backscatter coefficient were the most important waveform features at both spatial scales. A significant finding is that bare ground was clearly differentiated from shrubs using pulse width. Though the overall accuracy ranges between 0.72 – 0.89 across spatial scales, the two shrub PFTs showed 0.45 - 0.87 individual classification success at 1 m, while bare ground and tree PFTs showed high (0.72 – 1.0) classification accuracy at 10 m. We conclude that small-footprint waveform features can be used to characterize the heterogeneous vegetation in this and similar semi-arid ecosystems at high spatial resolution. Furthermore, waveform features such as pulse width can be used to constrain the uncertainty of terrain modeling in environments where vegetation and bare ground lidar returns are close in time and space. The dependency on spatial resolution plays a critical role in classification performance in tree-shrub co-dominant ecosystems.

Introduction

Climate and human driven disturbances in dryland ecosystems have adverse effects on biodiversity, ecosystem services, carbon storage, and desertification (Ahlstrom et al., 2015; Poulter et al., 2011). Furthermore, aridity in drylands is expected to increase in the future, causing expansion of land degradation and desertification (Huang et al., 2017). Ultimately, changes in the abundance and distribution of plant functional types (PFTs) in drylands can alter productivity and the capacity of these lands for carbon storage (Chen et al., 2017). Thus, PFTs are important indicators for monitoring the state of an ecosystem, as well as its resistance and resilience to climate and human driven disturbances (Lavorel et al., 1997; Poulter et al., 2015; Schimel, Asner, & Moorcroft, 2013). PFTs are frequently used as inputs for vegetation dynamics and earth system models (Krinner et al., 2005; Sitch et al., 2003; Wullschleger et al., 2014). However, uncertainty in PFTs, especially in dryland ecosystems between shrub, grass and forest classes reduces the accuracy of these models (Hartley et al., 2017). Hence, improved methods to capture the structure and function of PFTs in drylands are needed to accurately model carbon storage flux in these systems.

Due to its ability to capture three dimensional structure and some radiometric properties, light detection and ranging (lidar) is used to derive vegetation heights and digital terrain models, as well as to classify vegetation species, function and structure (Dalponte & Coomes, 2016). These products are further used for automated forest inventory estimates such as biomass and carbon stocks (Coomes et al., 2017; Dalponte & Coomes, 2016; Ene et al., 2017), as well as for ecosystem demography models (Thomas et al., 2008) to estimate carbon flux. Waveform lidar, which digitizes the total amount of lidar return energy at high vertical resolution (~1 ns = 15 cm), provides potential species-specific information about the illuminated target (Hancock et al., 2015; Hancock, Disney, Muller, Lewis, & Foster 2011; Roncat, Bergauer, & Pfeifer, 2011; Wagner, Ullrich, Ducic, Melzer, & Studnicka, 2006). The shape of the returning waveform results from a convolution of the temporal shape of the emitted pulse and system impulse (together called “system response/waveform”) with the target cross-section. Thus the backscattered waveform contains target characteristics such as size, orientation, and spatial arrangement, as well as radiometric characteristics of individual vegetation species (Hovi & Korpela, 2014; Korpela, Hovi, & Korhonen, 2013; Wagner et al., 2006).

Each echo in a waveform signal corresponds to an individual reflection target or set of targets. Thus, an echo can be used to detect individual target properties, the position and the orientation in 3D space. Through optimal waveform processing techniques, such as the commonly used Gaussian decomposition (Wagner et al., 2006), linear fitting or other asymmetric fitting techniques (Jutzi & Stilla, 2006; Mallet et al., 2010; Roncat et al., 2011; Wu, van Aardt, & Asner, 2011), numerous features can be derived from backscattered waveforms. Some of these additional waveform features and their biophysical relationships to the target are summarized in Table 2.1.

However, many of these waveform features (e.g. amplitude, pulse width, and backscatter cross section) are sensitive to system parameters such as incident angle, range and flying height (Abed, Mills, & Miller, 2012; Hovi & Korpela, 2014; Lin, 2015; Wagner, 2010). Thus, it is necessary to correct the influence of these system parameters on waveform features prior to application (Bruggisser, Roncat, Schaepman, & Morsdorf, 2017; Fieber et al., 2013; Wagner, 2010).

Table 2.1 Summary of waveform features derived from individual waveforms and their biophysical relationships to the target.

ATTRIBUTE	BIOPHYSICAL RELATIONSHIP	REFERENCE
Pulse width	Surface roughness and slope	Fieber et al., 2013
Amplitude	Optical response of the target to the emitted lidar wavelength	Fieber et al., 2013
Backscatter cross-section	Horizontal scattered cross-section of the target with respect to the deployed system wavelength, range, and incident angle	Wagner et al., 2006
Backscatter coefficient	The area-normalized backscatter cross-section corrected for incidence angle. A function of the target reflectance.	Wagner, Hollaus, Briese, & Ducic, 2008; Wagner, 2010
Differential target cross section	Laser system independent true target profile	Roncat et al., 2011
Rise time	Vertical structural distribution of target (e.g. in trees the vertical distribution of leaves and branches)	Ranson & Sun, 2000
Number of echoes	Vertical distribution and height of target	Heinzel & Koch, 2011
Height/height variability	Vertical distribution of target and its separation from ground	Fieber et al. 2013
Secondary explanatory features derived from any of the above parameters	N/A	Heinzel & koch, 2011

Waveform features and height information have been used to estimate vegetation structure as well as plant functional type and structural traits at both fine (< 2 m) and regional spatial scales (Alexander, Deák, Kania, Mücke, & Heilmeier, 2015; Wagner, Hollaus, Briese, & Ducic, 2008). Classification of plant functional types and individual species in tree dominant ecosystems show great improvement of classification accuracy with inclusion of one or several of these waveform features (Hovi et al., 2016). The pulse width and location characterize the vegetation components along the waveform path and have been used to classify deciduous and coniferous species (Reitberger, Krzystek, & Stilla, 2008; Yao, Krzystek, & Heurich, 2012). Wagner et al. (2008) shows that the scattering shape of backscattered signals can be used to separate vegetation from no vegetation with an accuracy up to 89%. Pulse widths can be used to classify vegetation in different patch conditions such as within varying soil roughness, understory and density (Hollaus, Aubrecht, Höfle, Steinnocher, & Wagner, 2011). Vaughn, Moskal, & Turnblom (2012) show that inclusion of frequency-domain full-waveform lidar features improve a five-species classification accuracy by 6% over discrete-return lidar alone, from 79 to 85%.

Numerous studies using combined features from discrete and waveform datasets have improved classification performance of tree and grass species (Heinzel and Koch, 2011; Neuenschwander, 2009; Vaughn et al., 2012). Backscatter cross-section alone can be used to distinguish ground, grass, and trees from each other (Fieber et al., 2013; Wagner et al., 2008). Further, lidar-derived height and energy related features have been used to delineate individual trees in object-based image analysis (OBIA) studies as the OBIA eliminates the discontinuity that is common in pixel-based classification (Zahidi, Yusuf, Hamedianfar, Shafri, & Mohamed, 2015).

In most of these studies, lidar-derived heights or height-based products such as canopy height models (CHM) and digital elevation models (DEM) play a critical role in delineation of individual tree crowns as well as in differentiating vegetation from bare ground (Hovi et al., 2016). Some vegetation studies use lidar returns above a certain height threshold (e.g. ~ 2 m above ground) for classification (Ene et al., 2017; Zahidi et al., 2015). However, in low-height vegetation, lidar does not return a separate energy peak unless the vegetation height is above the range resolution of the lidar system. Thus, bare ground lidar responses are typically mixed with low-height vegetation such as shrubs and grasses. This causes difficulties to measure the fractions of bare ground and vegetation, an important criterion for plant functional distribution mapping in dryland ecosystems (Hartley et al., 2017). Numerous studies in low-height ecosystems have documented that lidar heights underestimate vegetation heights (e.g. Streutker & Glenn, 2006). Similar underestimations and uncertainties appear in almost all studies which use lidar-based height features to model low-stature vegetated ecosystems across the world, which significantly affects regional ecosystem modeling and upscaling attempts (Hopkinson et al., 2005; Rango et al., 2000). Fortunately, waveform lidar is sensitive to the occurrence of low vegetation, where echoes often have a wider pulse than echoes from the bare ground. Although this limits the use of traditional lidar heights to separate ground from vegetation, the derivation of additional waveform features provides the opportunity to uncover hidden vegetation characteristics in the datasets.

In addition, vegetation distributions in many semi-arid ecosystems are topographically controlled and low-height vegetation often coexist with taller tree communities. The topographic and species complexity in these ecosystems makes classification using optical data challenging. In many instances, classification studies, even at high spatial resolution, consider all shrub species in one category (e.g. Zahidi et al., 2015). The complexity of heterogeneous semi-arid ecosystems further emphasizes the importance of understanding the effects of resolution in retrieving species type and diversity to guide future trade-offs in spaceborne sensors (e.g. GEDI and ICESat-2) (Abdalati et al., 2010; Endres, 2016; Qi & Dubayah, 2016) and ultimately, global ecosystem modeling. Semi-arid ecosystems cover a significant portion of the global land surface, and thus, the ability to map the density of shrubs and trees in these ecosystems will advance dynamic global vegetation models that account for vegetation demography (Fisher et al. 2018). For example, the clumping of foliage affects the exposure of bare ground and ultimately the land surface water vapor, carbon, and energy exchange.

The objectives of this study are three-fold. First, we aim to identify small-footprint waveform features to distinguish characteristics of two major shrub types from each other, from bare ground and from three dominant tree species in a pixel-based classification scheme in the Reynolds Creek Experimental Watershed (RCEW), Idaho. Second, we explore the influence of waveform-derived height features to differentiate these vegetation types and bare ground. Third, we test the effect of scale on waveform features used to classify the study site. For this we use two different pixel sizes (1 m and 10 m) to represent the waveforms and vegetation.

Materials

Study area

RCEW is characterized by a range of topography (1100 m – 2200 m) and plant functional types (PFTs) (Figure. 2.1). The study area consists of many varieties of grass, forbs, shrubs, trees, and riparian species. This study focuses on major PFTs of low stature shrubs (sagebrush (*Artemisia tridentata*), bitterbrush (*Purshia tridentata*)), and trees (Aspen (*Populus tremuloides*), juniper (*Juniperus occidentalis*), and Douglas fir (*Pseudotsuga menziesii*)). Dense tree canopies are observed at higher elevations, within valleys, and along streams. Shrubs and grass dominate throughout RCEW with species, density and structure varying by elevation. Further, the study area experiences topography-dependent mean annual temperature and precipitation regimes that vary between 4.6 – 9.2 ° C and 230-959 mm, respectively.

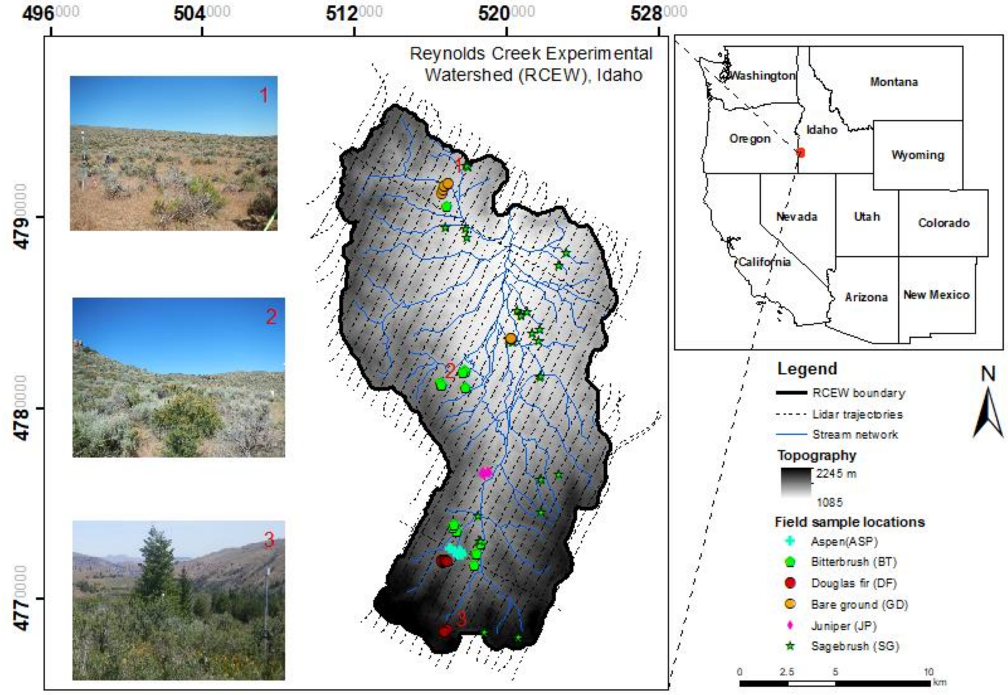

496000
504000
512000
520000
528000
4790000
4780000
4770000
Reynolds Creek Experimental Watershed (RCEW), Idaho
1
2
3
Washington
Montana
Oregon
Idaho
Wyoming
Nevada
Utah
Colorado
California
Arizona
New Mexico
Legend
RCEW boundary
Lidar trajectories
Stream network
Topography
2245 m
1085
Field sample locations
Aspen(ASP)
Bitterbrush (BT)
Douglas fir (DF)
Bare ground (GD)
Juniper (JP)
Sagebrush (SG)
N
0
2.5
5
10
km

Figure 2.1 Reynolds Creek Experimental Watershed study area with field sample locations (n=103 plots) of plant functional types (ASP, DF, JP, BT, SG, and GD) and waveform lidar trajectories. Field photos depict the sparse to dense shrub and tree communities (from top to bottom photo).

Field data

Reference field data of plant functional types (trees – aspen (ASP), juniper (JP), Douglas fir (DF)), shrubs – sagebrush (SG), and bitterbrush (BT), grass (native and invasive collectively), and bare ground (GD) were collected at 10 m x 10 m plots randomly selected over the study area (Figure 1). The plots were divided into PFTs based on the majority cover type within each plot. A line intercept method (Canfield, 1941) was employed to measure the percent vegetation cover in each shrub-dominated plot. The plot boundaries were collected using a RTK GPS and 5 transects established at 1 m, 3 m, 5 m, 7 m, and 9 m. Shrub type and the beginning and end points for each occurrence of a shrub intercepting a transect were recorded. The total lengths of intercepts for all five transects were calculated and summarized into percent cover by type per plot. In each shrub-dominated plot, we randomly selected six shrubs and collected their geographic position within the plot, species, height, and major/minor widths. For trees, we collected species information for several trees from each plot, avoiding mixed crowns. In many cases our tree plots were within 1 km of each other due to limited accessibility (steep valleys and ridges) and low dominance of trees overall in the watershed. In summary, we collected 103 plot-level (10 m x 10 m) samples containing 178 shrubs, 56 trees, and 23 bare ground samples.

Small-footprint waveform lidar data

Small-footprint waveform lidar data were acquired in August 2014 using the NASA Airborne Snow Observatory’s Riegl LMS-Q1560 (RIEGL Laser Measurement Systems GmbH, Horn, Austria), which is a dual laser scanner (1064 nm wavelength). The mean above ground level of ASO was 1000 m (700 – 1300 m due to terrain conditions) for a footprint of 20 – 60 cm. The scanning angle was ± 30^o. The study area was scanned at a pulse repetition rate of 400 kHz per laser and the backscattered signal was sampled at 1 ns per sample. The data were recorded using the low power channel. The resulting average point density was 10-14 pts/m^2. Numerous flight lines (38 parallel and 2 cross flights) were collected across the study area (Figure 1), resulting in multiple acquisition characteristics (scan angle, range, point density, and amplitude).

Methods

Decomposition of waveform lidar signals

In this study, we implemented a Gaussian decomposition technique for echo detection and analysis of both emitted and backscattered waveforms (measured in units of digital numbers (DN)) because the Riegl LMS-Q series emitted pulse is Gaussian (Wagner et al, 2006). We observed nearly symmetric pulses in the backscattered waveforms. Thus, we fit Gaussians to the raw waveforms recorded by the instrument. For the decomposition, waveforms that had raw amplitudes above a noise level of 6 DNs were considered. This noise level was defined based on other studies which have used Riegl's LMS-Q series (Mallet et al., 2010; Reitberger et al., 2008).

For echo detection using Gaussian decomposition, the maximum number of Gaussian echoes was limited to 7 per waveform. The number of observed echoes was always below 7, even at sites with tall trees (> 5 m) due to dense canopies and the laser footprint size (20 – 60 cm). The initial amplitudes and their position in space to initialize the Gaussian fit were derived using the maxima of Savitzky-Golay smoothed second derivatives of the original waveform. The second derivative was used because it helps to detect overlapping echoes with complex waveforms which are not detectable only using the local maxima of the first derivative (which is commonly used) (Bruggisser et al., 2017; Lin, Mills, & Smith-Voysey, 2010). The trigger for echo detection with the second derivative was defined as when the amplitude exceeded 4 and the spacing between echoes was larger than half of the initial pulse width. The initial pulse width was defined to equal that of the corresponding emitted waveform. We used a non-negative least square fitting algorithm developed in MATLAB (2016b) (The MathWorks Inc., Natick, MA, 2016) with the above initial Gaussian parameters to fit the backscatter signals. From the fitted Gaussians, we extracted the number of individual Gaussians in each waveform along with their maximum amplitudes, their position in the waveform (which was later used to calculate the range in meters), and the pulse widths at full width at half maximum (FWHM). We implemented a lower boundary condition of an amplitude of 17 DN, and a pulse width equivalent to that of the corresponding emitted waveform. The 17 DN is the marginal maximum amplitude that can produce a trigger amplitude (~ 36% of the 17 DN) above the noise level of 6 DN. The algorithm to extract waveform features from the raw lidar waveforms is illustrated in Figure. 2.2.

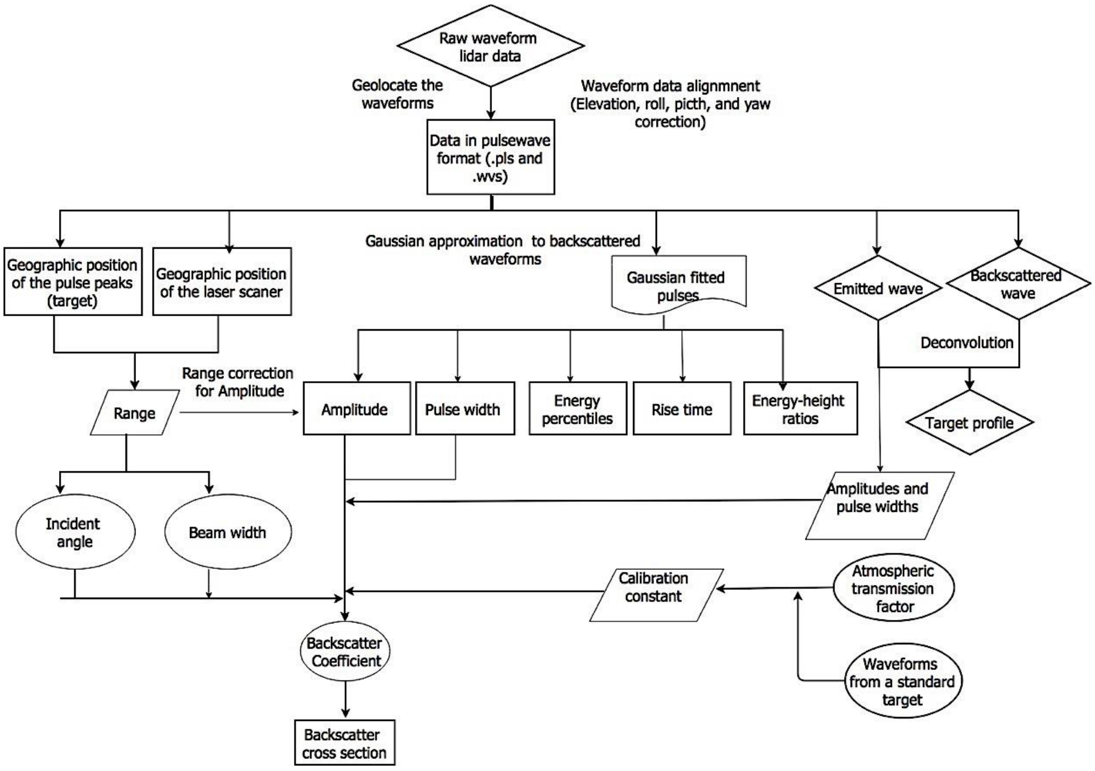

Raw waveform lidar data
Geolocate the waveforms
Waveform data alignmnent (Elevation, roll, picth, and yaw correction)
Data in pulsewave format (.pls and .wvs)
Geographic position of the pulse peaks (target)
Geographic position of the laser scaner
Gaussian approximation to backscattered waveforms
Gaussian fitted pulses
Emitted wave
Backscattered wave
Deconvolution
Range
Range correction for Amplitude
Amplitude
Pulse width
Energy percentiles
Rise time
Energy-height ratios
Target profile
Incident angle
Beam width
Amplitudes and pulse widths
Calibration constant
Atmospheric transmission factor
Waveforms from a standard target
Backscatter Coefficient
Backscatter cross section

Figure 2.2 Workflow illustrating the steps used to derive waveform features from the raw backscattered lidar waveform. Processing includes georeferencing, data alignment, Gaussian decomposition, deconvolution, and calibration.

Additional waveform features derived from Gaussian decomposition

We used the Gaussian fitted waveforms to derive a number of features (Table 2.2), including those shown in Figure. 2.3. The number of echoes and their maximum amplitudes and locations in each backscattered waveform (detailed above) were used to recognize the trigger amplitudes (~ 36% of max amplitude in the leading edge) and their georeferenced location in space. These locations were considered the target locations. The time duration from the trigger amplitude location to the maximum amplitude location is the rise time of each echo. Furthermore, using the spatial (x, y, and z) locations of the trigger and the scanner, we calculated the distance from the laser scanner to each echo (referred to as the range (R) hereafter) and the echo incident angle (θ). To facilitate the subsequent comparison of echo amplitude and energy values from overlapping flight lines at various ranges, the waveforms were first corrected using the model driven approach explained in Höfle & Pfeifer (2007). From the amplitude corrected waveforms, we integrated amplitudes from the trigger location of the first echo to the end location of the last echo in each waveform and used these as the cumulative energy of each footprint.

The end location of the last echo was defined as the last amplitude above the noise amplitude in the tailing edge of each echo. Using the cumulative energy curve (top to bottom), we calculated height at five energy percentiles from the total energy (10th, 25th, 50th, 75th, and 90th). This explains the waveform shape and energy distribution along the range. The total height was extracted by subtracting the ground elevations from the waveform location in 3D space. To obtain the absolute heights, we used a 1 m digital elevation model (DEM) derived from the point clouds of the same data set. Further, we calculated the cumulative energy at certain height percentiles (bottom to top) from total height. These calculations were made because differences in vegetation structure typically result in variations in the energy distribution in the returned waveform. For example, a dense canopy may have concentrated energy at the beginning of the waveform, whereas less dense canopy with ground exposure will cause larger energy concentrations near the end of the waveform. Further, different canopy structure or partial hits of the waveforms along the canopy edge will result in different waveform shapes.

The backscatter coefficient of each echo (γ_i) was calculated from equation (1) (Wagner, 2010).

$$\gamma_i = C_{cal} \frac{R_i^2 \hat{P} s_{p,i}}{\hat{S} \eta_{atm}} \tag{1}$$

In our study, we calculated the backscatter coefficient independent from the flying altitude (Wagner, 2010). The backscatter coefficient (γ_i) can be directly calculated using the calibration constant C_{cal}, the range R (in meters), the amplitude of the backscattered echo $\hat{P}$, the standard deviation of echo width $s_{p,i}$, the amplitude of the system's pulse $\hat{S}$, and the atmospheric transmission factor η_{atm}. The calibration constant C_{cal} was calculated using the backscattered waveforms of a 10 m x 10 m white standard reflectance (58% reflectance) tarp at the study site during airborne data collection. We also collected reflectance data of the tarp using a FieldSpec® Pro spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) and used the reflectance at 1064 nm (equal to lidar wavelength) for calibration. The emitted and backscattered waveforms of the tarp were extracted. The waveforms were Gaussian decomposed to extract amplitude, pulse widths, range and incident angle following the workflow in Figure. 2.2. Using the reflectance (ρ_d) and the incident angle (θ), the backscattering coefficient (γ_{CT}) per waveform was calculated from the equation (2) below (Wagner 2010).

$$\gamma_{CT} = 4\rho_d \cos\theta \tag{2}$$

With these calculated backscattering coefficients, pulse widths, and amplitudes, the average calibration constant was calculated using equation (3) and used as the calibration constant for the study (Wagner, 2010).

$$C_{cal} = \frac{1}{N_{CT}} \sum_{j=1}^{N_{CT}} \frac{\hat{S}_j \eta_{atm}}{R_j^2 \hat{P}_j s_{p,j}} \gamma_{CT} \, , \tag{3}$$

where γ_{CT}, N_{CT} are the backscatter coefficients of the calibration tarp and the number of echoes from the tarp used for the calibration, respectively. The η_{atm} is calculated from equation (4), where *a* is the atmospheric attenuation coefficient in dB/km (Höfle & Pfeifer, 2007).

$$\eta_{atm} = 10^{-2Ra/10000} \tag{4}$$

Frequency-domain deconvolution of lidar waveforms

Target cross section (σ) is another waveform lidar derived parameter and is a function of the target reflectivity (ε) with respect to the given laser wavelength and the illuminated target area (*dA*) (equation (5))

$$\sigma = \frac{4\pi}{\Omega} \varepsilon dA, \tag{5}$$

where Ω is the scattering solid angle of the target (Roncat et al., 2011). Although the raw backscattered waveform is a function of the emitted waveform and the laser system configuration, the target cross-section does not depend on instrument specifications.

Thus, the target cross-section values can be directly used to classify target properties, rather than the raw backscattered signal. As prior information about the target reflectivity and scattering solid angle are limited, Wagner (2010) shows that the target cross section can be directly calculated from the backscattering coefficient (γ) and the laser footprint area (A_{lf}) (equation (6)).

$$\sigma = \gamma A_{lf} \tag{6}$$

However, all the lidar parameters described in section 3.2, including the backscattering coefficient, depend on the assumed Gaussian behavior of the emitted and backscattered waveforms. Thus, the target cross-section calculated using equation (6) also becomes a Gaussian function in time. A backscattered waveform can be considered as a convolution of the emitted waveform and the derivative of the interacting target cross-section (Roncat et al., 2011). To estimate the target cross-section without the Gaussian assumption, we deconvolved the emitted waveform from the backscattered waveform. We converted each received backscattered waveform (*bw*) and the emitted waveform (*ew*) into Fourier frequency domain. In frequency domain (*f*), deconvolution is a spectral division of the backscattered waveform by the emitted waveform (Equation (7)).

$$\sigma(f) = \frac{bw(f)}{ew(f)} \tag{7}$$

To suppress division by small numbers (e.g. 0) we used a water-level regularization algorithm, which added a small value to the denominator and prevented noise enhancement in the deconvolution. In this way, we extracted the target cross-section from each laser backscattered waveform in the frequency domain. The frequency-domain target cross-section was transformed back into the time-dependent target cross-section (referred to as the differential target cross-section, DEC in Table 1). From the differential target cross-section, we calculated the target profile max amplitudes and integrated cross-section. The target cross-section (σ) is a function of the target reflectivity (ε) with respect to the given laser wavelength and the illuminated target area (dA) (equation (8))

$$\sigma = \frac{4\pi}{\Omega} \varepsilon dA, \tag{8}$$

where $\boldsymbol{\Omega}$ is the scattering solid angle of the target (Roncat et al., 2011). The number of echoes, echo amplitudes and the total energy (integration of the target cross-section) were extracted from the deconvolved target cross-section as predictor features.

Once we completed the waveform feature extraction (sections 3.2 and 3.3), a correlation analysis was performed between all features derived from individual backscattered waveforms and the incident angle to ensure that the features were not biased by viewing geometry.

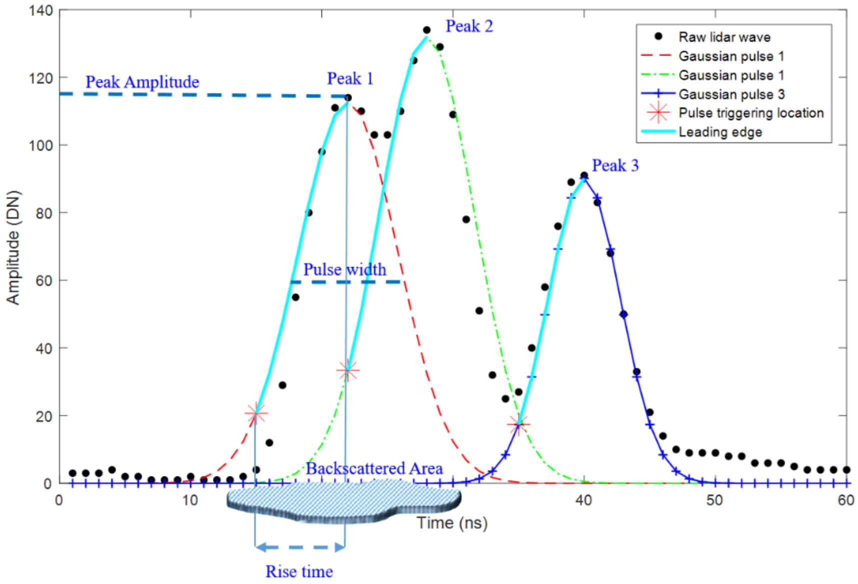


Figure 2.3 Illustration of information contained in a lidar waveform. Peak 1, 2, and 3 are the echoes from three scatterers detected by the waveform. Three Gaussian functions (Gaussian pulse 1, 2, and 3) were fitted to the raw waveform. The peak amplitude is the maximum amplitude of the first echo after Gaussian fitting. Pulse width is the full width at half maximum. The stars are the trigger locations of each echo. Leading edge is the time distance from trigger to the max amplitude. The backscattered area represents the scattered cross-section from the first echo.

Table 2.2 Features extracted from waveform backscatter lidar.

Code	Variable	Description
Amplitude (First, Last, and Total)†	Amplitude (echo maximum) in DNs	Mean of digital numbers (DN) of all peaks corrected for the range, atmospheric influence and incident angle within a given pixel.
Width (First, Last, and Total)†	Pulse width (full width at half max)	Mean of pulse widths measured from Gaussian decomposition (ns) within a given pixel.
X, Y, and Z	Echo coordinates (X, Y, Z)	Georeferenced easting, northing and elevation coordinate of each echo triggering location in meters.
Rise time*	Rise time of all pulses	Number of time bins between 10% - 90% energy at rising edge of each pulse.
Fall time*	Fall time of all pulses	Number of time bins between 10% - 90% energy at trailing edge of each pulse.
θ	Incident angle	Wave incident angle in degrees.
Height	Heights at percent energies in each waveform Absolute height from the ground to first and last echo positions of each waveform	Absolute heights at cumulative energy percentiles (10th, 25th, 50th, 75th, 90th). The absolute height was derived by subtracting the elevation of the last location of the last echo in each waveform from elevation at each percentile. Units are in m.
Energy	Waveform energy at heights from first echo triggering location	Cumulative energy at height percentiles (10th, 25th, 50th, 75th, 90th) as sum of DNs divided by 100.
γ (First, Last, and Total)†	Backscatter coefficient (per pulse and per waveforms)	Backscatter coefficient calculated as in (Wagner, 2010).
Differential σ	Differential target cross-section	Target waveform profile by deconvolution (the system waveform influence was removed from the backscatter signal). The target profile is

		an amplitude profile with respect to the range in meters.
Deconvolved Amplitude	Deconvolved wave amplitudes	Digital numbers of echo maximums in the differential target cross-section.
σ	Target cross-section	Integral of the Differential target cross-section.

† First and Last – variable measured from the first & last pulse in multi-pulse backscattering waveforms respectively. † Total – Sum of the variable measured from all the pulses from multi-pulse backscattering waveforms. *The rise time and the fall time are equal in value because we use Gaussian decomposition.

Plant functional types classification

We classified the PFTs at two different spatial scales (1 and 10 m) to account for the impact of canopy size variation between shrubs and trees, and to assess the potential for upscaling to large-footprint waveform acquisition. Based on the average canopy area of shrub (<= 1 m^2) and tree (>3 m^2) PFTs, and assuming an individual tree is more likely to be contained within a 10 m pixel, we expected that waveforms derived from 1 m and 10 m would better characterize shrubs and trees, respectively. We used all waveforms in 1 m and 10 m pixels and derived the mean, standard deviation, skewness and kurtosis of each waveform feature listed in Table 1. The response feature was the PFT categories (sagebrush, bitterbrush, ground, aspen, juniper, and Douglas fir).

Implementation of random forest classification

We used an ensemble random forest (E-RF) (Ko, Sohn, Remmel, & Miller, 2016) to classify the PFTs at the plot level (using 10 m pixel size) and at individual locations (using 1 m pixel size). We used an ensemble approach to reduce classification bias (Ko, Sohn, Remmel, & Miller, 2016). The random forest algorithm itself is an ensemble classifier where the final classification labels are obtained by combining multiple classification trees for categorical predictors using approximately 63% of the data for training (in-bag data) and 37% of the data (out-of-bag (OOB) data) for validation (Breiman, 2001). We trained a set of base classifiers using this traditional random forest classification algorithm and ensembled the base classifiers to provide the final class using the majority vote approach (Ko, Sohn, Remmel, & Miller, 2016). We used binary based classifiers because this approach allows “unknown”, or unclassified data in the final classification product. In comparison, a traditional supervised random forest classification classifies the whole field study area during imputation. The traditional random forest model was computed to compare to the ensemble model performance.

We used 257 individual samples (1 m) and 103 plot scale samples (10 m) for the random forest model development. In each spatial scale, we selected 50 % of the response PFT categories for training and used the remainder for validation. The selection of 50 % was chosen to provide enough samples from each category to train the base classifiers. We trained six binary base classifiers (sagebrush (SG)/non-sagebrush, bitterbrush (BT)/ non-bitterbrush, ground (GD)/non-ground, aspen (ASP)/non-aspen, juniper (JP)/non-juniper, Douglas fir (DF) /non-Douglas fir) with and without height-based features and at both 1 m and 10 m spatial scales (Figure. 2.4). This produced four ensemble RF models (Table 3).

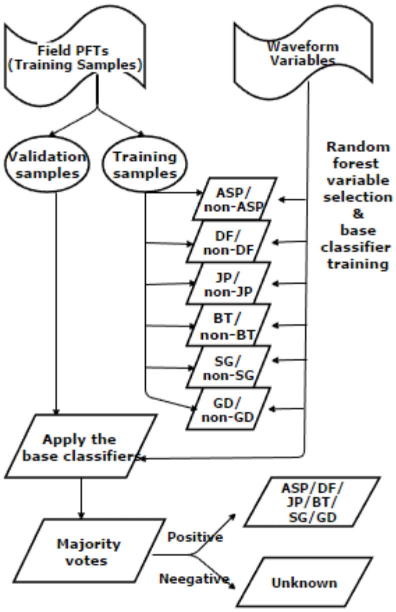

Field PFTs
(Training Samples)
Waveform
Variables
Validation
samples
Training
samples
Random
forest
variable
selection
&
base
classifier
training
ASP/
non-ASP
DF/
non-DF
JP/
non-JP
BT/
non-BT
SG/
non-SG
GD/
non-GD
Apply the
base classifiers
ASP/DF/
JP/BT/
SG/GD
Majority
votes
Positive
Neegative
Unknown

Figure 2.4 Ensemble random forest PFT classification workflow. Feature selection and base classifiers were trained using training samples. The selected base classifier models were applied to the validation samples and ensembled to make the final decision.

To train the RF models, important features were selected using the “varselRF” package in R software (Diaz-Uriarte & Alvarez de Andres, S., 2005). This package was chosen as it selects the important features using iterative backward feature elimination until the OOB error stabilizes and has been used successfully in previous lidar studies (e.g. Chen, Li, Wang, Chen, & Liu, 2014). For each base classifier we set 5000 trees for the first forest and 2000 trees for each additional forest for variable selection. We set 0.2 as the variable drop factor to exclude the features at the next iteration. From the selected important features in each case, a RF model was generated and applied to the validation data set. In E-RF, the base classifiers simultaneously classify sagebrush (non- sagebrush), bitterbrush (non- bitterbrush), ground (non-ground), aspen (non-aspen), juniper (non-juniper), and Douglas (non-Douglas) in the validation data set. In cases where there are no conflicts in decisions made among the base classifiers, the final decision is made by the classifier voted for by a positive case. If there is a conflict in decision, the final decision is made by the class that has the majority positive vote from all base classifiers. Where all six classifiers vote negatively, the class is labeled as “unknown” (Ko, Sohn, Remmel, & Miller, 2016). E-RF model performance at both spatial scales was assessed using the overall accuracy and Kappa coefficient (κ). The overall accuracy is the ratio between the number of correctly classified PFT samples and total reference PFT observations tested. Kappa coefficient (κ) is a measure of agreement between overall (observed) accuracy with an expected accuracy from random chance (Jensen, 2005). We also tested the classification success of each PFT using producer and user accuracies to evaluate the best practice. Producer accuracy is the probability of the reference data being correctly classified by the method employed. The user accuracy measures how well the classified results represent what is observed in the ground (Jensen, 2005).

Table 2.3 Ensemble random forest (E-RF) models used to evaluate the waveform attribute selection and PFT classification.

RF Model	Description
NH_ E-RF_1 m	Ensemble random forest model without lidar derived height features in predictor space at 1 m
H_ E-RF_1 m	Ensemble random forest model with lidar derived height features in predictor space at 1 m
NH_E-RF_10 m	Ensemble random forest model without lidar derived height features in predictor space at 10 m
H_E-RF_10 m	Ensemble random forest model with lidar derived height features in predictor space at 10 m

Results

Influence of the viewing geometry on waveform features

Waveform features derived from all individual backscattered waveforms used in this study indicated a low correlation (-0.07 – 0. 22) with incident angle (θ). The maximum correlation was with the first return pulse width (0.22). Although the maximum possible scan angle of the instrument was 28°, the local incident angle of the tested samples varied between 0.7 and 32° due to the rough terrain of the study area. The amplitude and pulse width of the system waveforms had negligible variability. However, wherever necessary (e.g. for initial pulse width during Gaussian decomposition, backscattered coefficient estimation) we used amplitude and pulse width values from each individual emitted waveform with each respective backscattered waveform to derive our features instead of applying a constant emitted pulse width or amplitude.

Important waveform features for PFT classification

In almost all of the ensemble RF models we produced (Table 3), percentile energy (e.g. 10th, 50th, and 75th percentiles), statistical moments of target cross-section (standard deviation of σ), rise time, statistical moments of backscatter coefficient (standard deviation of first and total γ), and pulse widths were selected as the most important waveform features. These results were observed even when lidar-derived height features were included (except for the ASP/non-ASP and SG/non-SG in 1 m). Overall, more height features were selected as the most important features in the 1 m than in the 10 m classifications. Further, inclusion of heights resulted in a more complex model than those without heights at both spatial scales. In comparison between scales, the target cross-section and the corresponding standard deviation (σ and standard deviation of σ) frequently appeared among 1 m base classifiers, while varieties of backscatter coefficient such as standard deviation of first and total γ more often appeared in 10 m base classifiers. The number of peaks was not among the most significant features at any spatial scale in this study.

Figure. 2.5 illustrates the most important features at the 1 m scale analysis for each PFT. All tree PFTs (ASP, DF, and JP) stand out by having higher standard deviation of 90th percentile energy, 10th percentile height, total width, rise time, and 90th percentile height. The vertical structure distribution of tall vegetation tends to generate long smeared waveforms with slow rise and multiple peaks. From the selected tree PFTs, ASP shows the highest variability in several waveform features. DF stands out by having the highest standard deviation of 90th percentile energy, total width, and rise time and may represent the tall, dense internal vegetation structure compared to other tree PFTs used in this study. The JP PFT had the highest first width and standard deviation of first γ responses. Shrub PFTs (BT and SG) and the ground class (GD) show relatively lower means of standard deviation of 90th percentile energy, 10th percentile height, total width, rise time, and 90th percentile height. However, BT and SG show higher means than bare ground for 75th percentile energy, first width, and 50th percentile energy. GD shows significantly low values of first width (< 3.2 ns threshold) and rise time (< 6.2 ns threshold) reflecting the narrow single pulses from bare ground. Thus, these features can be used to distinguish bare ground lidar signals from vegetation signals.

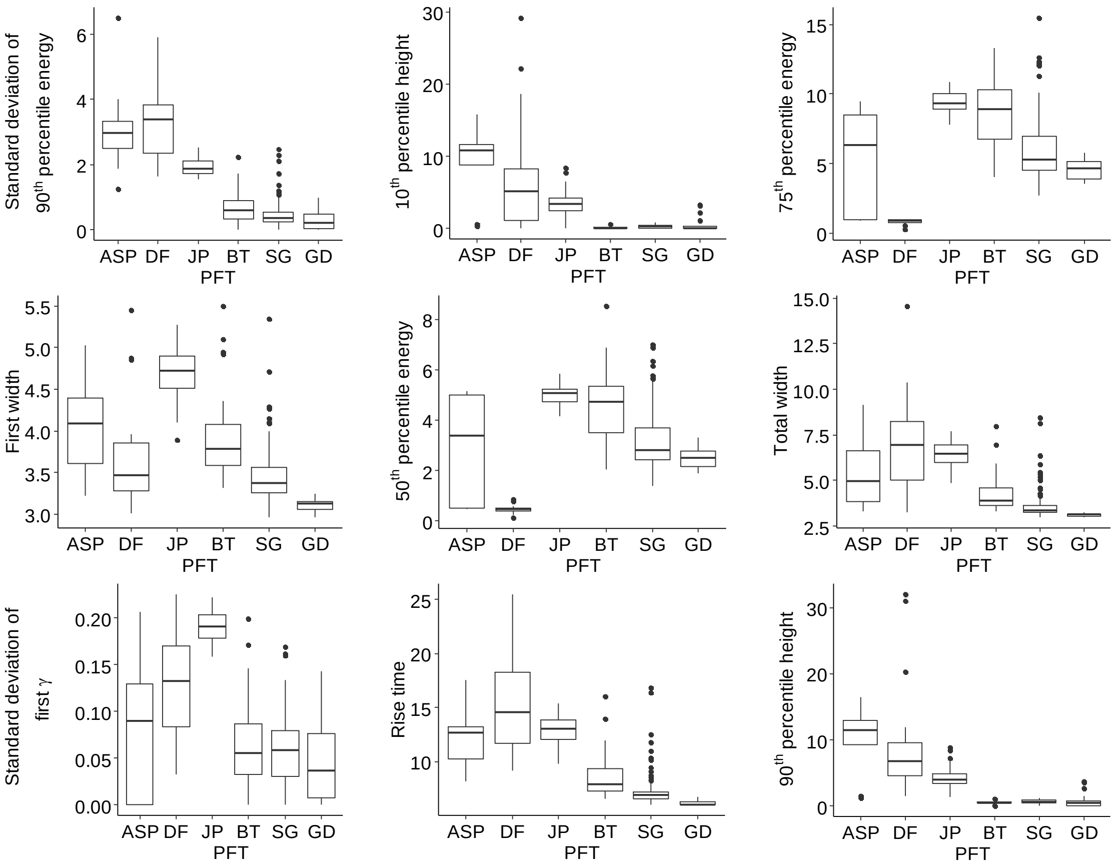


Figure 2.5 Box plots showing the variability of values of the most important waveform features at 1 m scale among PFTs. Definitions of PFTs are: ASP-aspen, DF-Douglas fir, JP-juniper, BT-bitterbrush, SG-sagebrush, GD-bare ground. From left to right, top row: variability of standard deviation of 90th percentile energy, 10th percentile height, and 75th percentile energy, respectively; middle row: variability of pulse width of first returns (first width), 50th percentile energy, and pulse widths of all returns (total width), respectively; bottom row: variability of standard deviation of backscatter coefficient of first returns (first γ), rise time, and 90th percentile energy, respectively. Note the differentiation of shrubs and bare ground with the pulse width and rise time.

First width, rise time, and 50th percentile energy were always among the most important features regardless of the spatial scale. However, different trends and higher ranges were observed at 10 m compared to the 1 m spatial scale (Figure. 2.6). Tree PFTs show relatively high values for first width, rise time, and standard deviation of σ at 10 m spatial scale. For cumulative energy at 50th percentile energy, DF shows the smallest mean in 1 m and the highest in 10 m. Standard deviation of σ of PFTs at both spatial scales were similar. At 10 m, DF has the highest value of 50th percentile energy as well as the highest variability of rise time, standard deviation of σ, and skewness of last amplitude. Although the mean values of many waveform features of JP were closely aligned with DF, the 50th percentile energy and standard deviation of first width metrics show an opposite trend allowing JP to be distinct from DF. ASP is distinct at 10 m having the highest values for first width, standard deviation of total γ, and standard deviation of first width with low variability. Bare ground is well discriminated from all vegetation classes at 10 m with first width, rise time, kurtosis of first amplitude and skewness of last amplitude. The two shrub PFTs fall between the value ranges for ground and tree waveform features. However, with the exception of skewness of last amplitude, bitterbrush has higher means than sagebrush for all waveform features.

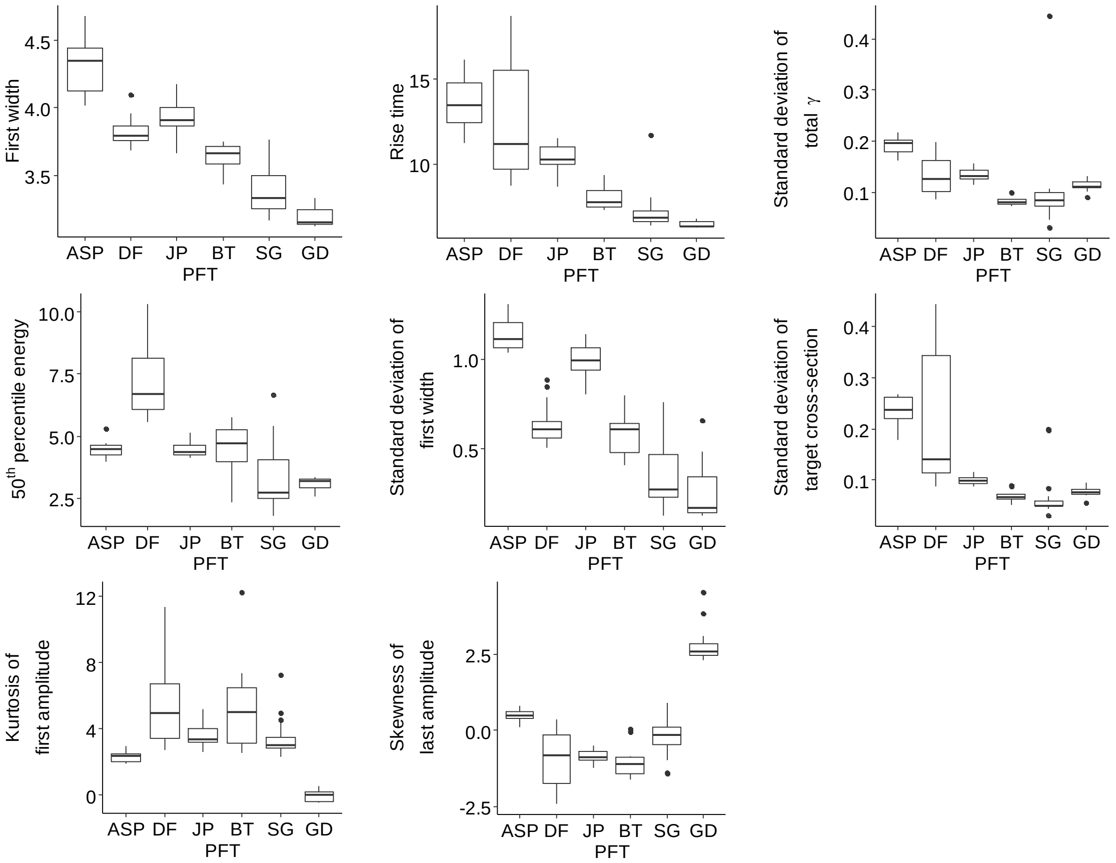


Figure 2.6 Box plots showing the variability of values of the most important waveform features at 10 m scale among PFTs. From left to right, top row: variability of pulse width of first return, rise time, and standard deviation of total backscattering coefficients, respectively; middle row: variability of energy at 50th percentile height, and standard deviation of first return pulse widths, and standard deviation of target cross-section, respectively; bottom row: variability of kurtosis of first return amplitudes and skewness of last return amplitudes, respectively.

RF model performance

The classification success of each model was assessed based on Kappa and overall accuracy values using the validation data sets. Each model was iterated 100 times with randomly selected training and validation data sets without replacement. All 10 m models outperformed the 1 m models (Table 4). The random forest model at 10 m spatial scale without height (NH_E-RF_10m) showed the highest classification success (κ = 0.86, overall acc. = 0.89). The inclusion of waveform derived heights decreased the accuracy by 4%. However, the inclusion of heights increased the accuracy of PFT discrimination by approximately 10 % at the 1m spatial scale (κ /overall acc. = 0.70/0.80 and 0.60/0.72, with and without height features, respectively).

Aspen (ASP) and shrub PFTs had the highest producer accuracies with waveform heights at 1 m and without waveform heights at 10 m whereas the opposite effect of waveform heights was found at 1 m and 10 m for DF and JP. Shrubs had similarly high producer's accuracy at both 1 m and 10 m, whereas trees and bare ground generally had stronger classifications at 10 m (producer acc. 0.93-1).

Table 2.4 Producer and user accuracy of each PFT (ASP, DF, JP, BT, SG, and GD) in each RF model described in Table 3.

RF Model	Model accuracy		Producer Accuracy						User Accuracy					
1m														
	Kappa	Overall	ASP	DF	JP	BT	SG	GD	ASP	DF	JP	BT	SG	GD
NH_ E-RF_1 m	0.6	0.72	0.14	**0.87**	**0.93**	0.45	0.8	0.42	**0.93**	0.9	0.92	0.66	0.76	0.7
H_ E-RF_1 m	**0.7**	**0.8**	**0.47**	0.86	0.92	**0.56**	**0.84**	**0.55**	0.9	**0.95**	**0.99**	**0.75**	**0.81**	**0.85**
10 m														
NH_E-RF_10 m	**0.86**	**0.89**	**0.95**	0.96	0.91	**0.57**	**0.85**	**1**	1	**0.95**	0.94	**0.83**	**0.91**	**1**
H_E-RF_10 m	0.81	0.86	0.72	**0.98**	**0.93**	0.33	0.81	**1**	**1**	0.92	**0.97**	0.65	0.9	**1**

Discussion

Differentiating PFTs with waveform features

Pulse widths (width), rise time, and percent energy (10th, 50th, 75th, and 90th percentiles) were among the most important waveform features at both spatial scales to distinguish all vegetation types. These results closely align with findings of Hovi, Korhonen, Vauhkonen, & Korpela, (2016) to differentiate tree species. Pulse width and the rise time of an echo is a measure of the canopy roughness along the laser path. Thus, the lowest pulse width and rise time result from bare ground while the highest can result from tree PFTs (Figure. 2.6 and Figure. 2.7). Within tree PFTs, a range of pulse widths are observed from the heterogeneity of the distribution of leaves and branches. For example, high-headed deciduous tree canopies like in aspen showed mostly single pulse backscattered waveforms with large pulse widths. Equal canopy density in aspen resulted in a lower variability of rise time in the first echo. The pyramidal shaped canopy, with closely arranged branches from canopy top to bottom in Douglas fir and juniper, are likely to produce smeared backscattered waveforms with one or more echoes. As a result, pulse width and rise time had a higher variability than in aspen. The percent energy is a measure of backscattered radiation at different heights of the target along the laser path. The significant differences in percentile energy (at both 50th and 75th) of Douglas fir and juniper may be due to differences in the fraction of woody material and needles interacting with the laser beam. Thus, the variability of pulse width, rise time, and percentile energy can be used to differentiate broadleaf deciduous and evergreen conifers with distinct canopy distributions without adopting an initial individual tree delineation process.

In this study, we included both backscatter coefficient (γ) and target cross-section (σ) as waveform features. Some studies have used backscatter coefficient alone or with other waveform features (Fieber et al., 2013; Mallet, Bretar, Roux, Soergel, & Heipke, 2011); however, we found the target cross-sections (σ) to be useful for classification. The backscatter coefficient (γ) is a measure of the electromagnetic energy intercepted and reflected by objects back towards the sensor. On the other hand, target cross section describes the target's structure distribution as a function of range and is independent of the laser scanning system and decomposition parameters deployed (Roncat et al., 2011;

Roncat, Pfeifer, & Briese, 2012). Neither γ nor σ were among the most important features. Our results in Figure. 2.7 show γ has low variability among and within PFTs compared to the σ at both spatial scales. These results confirm that the structure (target cross-section) is more important than lidar derived radiometric parameters to separate PFTs.

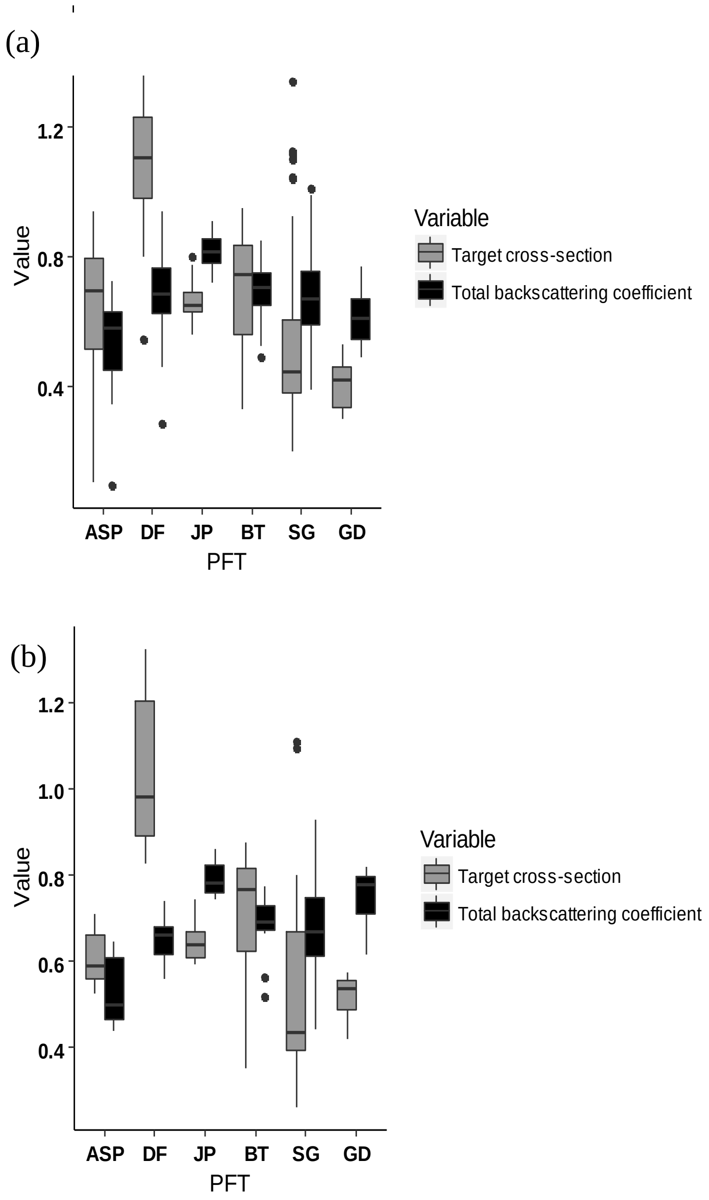


Figure 2.7 Distribution of total backscattered coefficient (total γ) and target cross-section (σ) of PFTs at 1 m (a) and 10 m (b) spatial resolution. Total backscatter coefficient shows less variability than the target cross section within PFTs as well as among PFTs.

Surprisingly, the number of peaks and amplitude were not among the most important features (except for the SG and GD classifiers at 10 m spatial scale) as per other similar studies (Bruggisser et al., 2017; Cao et al., 2016; Heinzel & Koch, 2011; Mallet et al., 2011). This may be due to the range of incident angles (0.7-32 degrees) in our data, resulting in the possibility of some laser beams hitting trees at lower heights and producing single pulse backscattered signals. Furthermore, dense tree canopies may produce smeared single pulses or high energy first peaks with following echoes being equivalent to the noise level in the system. Thus, with the potential to produce single pulses from almost all vegetation PFTs, the number of peaks may not be a distinguishing characteristic in our study site.

Peak amplitude was not a strong waveform feature in differentiating PFTs. The amplitude is a function of target reflectance at 1064 nm and is further affected by other factors such as geometric properties. The similar amplitudes of shrubs and ground (Figure. 2.8) is likely due to the influence of high soil and litter exposure (and thus reflectance) in this shrubland ecosystem at both spatial scales. Similarly, the large variation in amplitude in the aspen and Douglas fir is likely due to varying cover between these two PFTs.

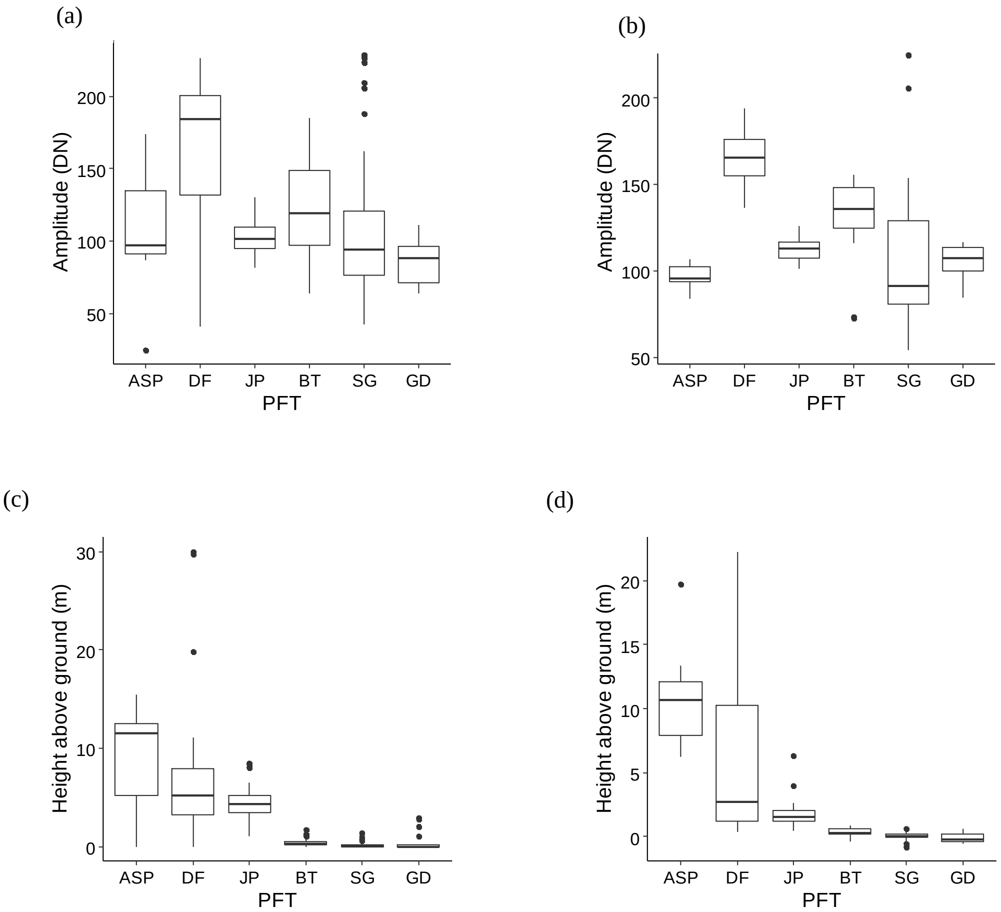


Figure 2.8 Distribution of amplitude of the selected PFTs at 1 m (a) and 10 m (b) spatial scale. Distribution of heights above ground of the PFT samples at 1 m (c) and 10 m (d) spatial resolution.

Although the classification results were highly correlated with field observed shrub PFTs at both spatial scales, regional scale attribution of these models misclassified some of the relatively dense canopy sagebrush (e.g. Wyoming big sagebrush (*Artemisia tridentate* spp. *wyomingensis*)) as bitterbrush. The variability of waveform features in response to size, age, and genotype within the same PFT needs to be further evaluated. In the PFT classification, the edges of the tree patches were classified as “unknown”. The confusion and edge effects may be reduced by coupling our waveform data with optical data from multispectral or hyperspectral data. In addition, the development of physical radiative transfer (RT) models incorporating simulated lidar waveforms using known reference target structural and optical properties as well as detailed sensor specifications (spatial and spectral resolutions, sensor viewing geometry, platform altitude, and range) should be explored. These models will help constrain waveform features as well as quantify the influence of vegetation structure on canopy reflectance and image texture (Bruniquel-Pinel & Gastellu-Etchegorry, 1998; Gastellu-Etchegorry et al., 2015, 2016; Malenovsky et al., 2008). Further, by inversion of waveform lidar RT models with known sensor specifications, we have the potential to extract vegetation biophysical parameters such as fractional vegetation cover, leaf-area index, and vegetation heights and extensions (Koetz et al., 2006).

Waveform derived heights and differentiating bare ground from vegetation

In this study, height was one of the least significant features (except for the ASP/non-ASP and BT/non-BT classifiers at 1 m spatial scale). However, similar studies have noted that airborne waveform lidar derived heights at percent energy (Neuenschwander, 2009) or median energy height (Cao et al., 2016) are among the most significant features for classification. Many of these studies used tree species in which the vegetation height is sufficiently separable from the ground. A similar study to ours in an open grassland did not include height as a variable and instead incorporated other leaf-off season waveform features (Alexander et al., 2015). The inclusion of height to differentiate species requires accurate bare ground elevations as well as low intra-species, but high inter-species height variability. In our study the application of a height threshold is limited because the variability in heights between shrubs and ground, and between tree species were too low. For example, previous studies have shown uncertainty in lidar heights to be approximately 30% of shrub height (e.g. Glenn et al., 2011).

Regardless, lidar-derived height plays a critical role for digital elevation modeling (DEM), digital terrain modeling (DTM), digital surface modeling (DSM), and many object-based classification studies (Reitberger et al., 2008; Yao et al., 2012), as well as hydrological studies (Painter et al., 2016). The results of this study show that waveform features such as rise time, pulse width, statistical distribution of amplitude (skewness and kurtosis), backscattering coefficient, and target cross-section can be used as prior information to filter ground returns (e.g. pulse width < 3.2 ns) from vegetation returns in shrub and savanna ecosystems to constrain the uncertainty in elevation model development. Further, the ability to separate bare ground from low-height vegetation such as shrub and grass helps accurately map the vegetation and soil fractions, a key controller in land surface model uncertainty, especially in drylands.

Scale dependency of PFT classification

Our overall results show that small-footprint lidar waveform features can differentiate PFTs with high accuracy at both 1 m and 10 m spatial scales (80–89%) in semi-arid ecosystems. Shrubs showed relatively high accuracy at 1 m, whereas trees and bare ground were best classified at 10 m. Our study demonstrates that the individual PFT classification accuracy depends on both the average individual stand size (shrub or tree) and the selection of the pixel size. Most shrubs were contained within a 1 m radius, and thus, the waveform features can capture the variation at this scale. However, trees have a larger canopy area and although we used the canopy center location to extract waveform features of tree PFTs at 1 m, the intra-structure (stem and branches) are beyond this radius. This was clearly observed in most aspen stands. Further, waveforms with a higher viewing geometry may belong to more than one pixel. Thus, at 1m, the extracted tree waveform features have less potential to represent actual tree parameters.

Our study area comprises heterogeneous vegetation species and structure (shrubs, forbs, grasses, evergreen and deciduous trees), topography, and sub climatic regions. Hence, the suite of waveform features (e.g. pulse width of 3.2 (ns) for ground and shrub separation, variability of pulse width, rise time, and percentile energy for deciduous and evergreen separation) is likely capable of adapting to ecosystems with similar shrub and tree communities. Moreover, additional field attributes such as individual tree height age, soil type, and moisture condition of the site, may help to explain the intra-species variability of waveform features of tree communities (Hovi, Korhonen, Vauhkonen, & Korpela, 2016). The waveform lidar derived high accuracy PFTs and their uncertainty with respect to individual stand and pixel sizes can also be used in dynamic global vegetation models in which the PFTs and their structure play a critical role, to evaluate the model performance.

Upscaling to large-footprint waveforms

Our study results inform the application of future large-footprint waveform data such as GEDI (Abdalati et al., 2010) over semi-arid ecosystems. For example, the pulse width of the emitting waveform of GEDI (<= 15.6 ns) (ISS: GEDI, retrieved 9_24_2017) is much larger than the ground and shrub pulse widths in our study area (3.7 ns). While this pulse width is not ideal for differentiating shrubs or shrubs and bare ground in sparsely vegetated ecosystems, simulation techniques and a binned (1 ns / 15 cm in both the ASO system and GEDI) energy distribution pattern analysis may help identify vegetation hits along the waveforms and hence, to approximate related parameters such as vegetation fraction and leaf-area index relevant to dryland ecosystems. However, the accuracy of these simulation and pattern analysis methods heavily depends on the superiority of noise tracking and reduction. Our 10 m results can be used as intermediate scale observations to bridge high (ASO 20-40 cm footprint) and low (GEDI's 25 m footprint) spatial resolution studies over vegetation to understand scale dependency of target retrieval (similar to Holm, Nelson, & Ståhl, 2017). However, direct retrieval of structural parameters such as canopy height from arid vegetation at the satellite level is still limited due to the compromised coarse pulse widths, digitization interval, and laser footprint sizes to maintain high signal to noise ratio (SNR) at far range measurements (Hassebo, 2012).

This study was developed using small-footprint slant-range waveforms with a high degree of incident angles. Although the vertical resolution is high, slant-range waveforms in complex vegetation and terrain conditions does not necessarily represent the true target structure. While waveform features were extracted after individual object delineation, slant-range backscattered waveforms within individual objects show a range of patterns based on where it hits the target and at what angle. This can be partially avoided by aggregating the waveforms at the raw level rather than aggregation of features either at the object or pixel scale. Our future studies intend to aggregate waveforms preserving incident angle information of component waveforms. This will allow us to evaluate not only the limitations caused by small-footprint waveforms, but also the capability of large-footprint waveforms (using pseudo footprint produced by aggregation of raw waveforms) in complex terrain for vegetation studies.

Conclusions

While computationally more challenging than discrete return lidar data, deriving waveform features from the ASO data provides new information to map semi-arid ecosystems. The Gaussian decomposition and frequency-domain deconvolution techniques yield robust predictors in our study area. Pulse width, rise time, percent energy, target cross-section, and radiometrically calibrated backscatter coefficient were the most important waveform features at both spatial scales to differentiate the shrubs, trees, and bare ground from each other. Among those, pulse width and backscatter coefficient had the lowest variability within each PFT in comparison to the other waveform features.

The ability to differentiate vegetation from bare ground with ASO waveform data is a significant improvement for ecosystems where vegetation height is lower than the height uncertainty in discrete return lidar (often +/- 15 cm). The lack of importance of the number of peaks for PFT classification demonstrates the capability of waveform features to differentiate PFTs even within a lower pulse density scenario. Overall, this study shows exciting promise to constrain PFTs in heterogeneous semi-arid ecosystems, providing new opportunities for automated inventorying and monitoring and estimating biomass and carbon from waveform lidar data.

CHAPTER THREE: SPACEBORNE LIDAR REVEALS TRENDS AND PATTERNS OF FUNCTIONAL DIVERSITY IN A SEMI-ARID ECOSYSTEM

Abstract

Assessing functional diversity and its abiotic controls at continuous spatial scales are crucial to understand changes in ecosystem processes and services. Semi-arid ecosystems cover large portions of the global terrestrial surface and provide carbon cycling, habitat, and biodiversity, among other important ecosystem processes and services. Yet, the trends and patterns of functional diversity in semi-arid ecosystems and their abiotic controls are unclear. Here, we mapped the functional diversity in a semi-arid ecosystem using airborne small footprint lidar data (ALS) and evaluated the potential of a newly launched spaceborne lidar system (GEDI) to capture functional diversity trends of the same ecosystem. Our results revealed that functional diversity in this ecosystem is mainly governed by elevation gradient, soil, and slope. All three functional diversity indices (functional richness, functional evenness and functional divergence) showed a diversity breakpoint at elevations around 1500 m – 1700 m. We found a more even and less divergent ecosystem in fire disturbed regions of our study area. The consistency of diversity trends between airborne lidar and GEDI confirmed GEDI’s potential of capturing functional diversity of semi-arid ecosystems. The number of GEDI footprints relative to the size of the fire-disturbed areas restricted the ability to capture the full effects of fire disturbance. Future opportunities to fuse GEDI with ICESat-2 and

TanDEM-X data to fill gaps, will improve our ability to capture disturbance-driven diversity in semi-arid ecosystems.

Introduction

Understanding the drivers of ecosystem processes and services at regional and global scales provide pivotal knowledge to assess ecosystem responses under changing conditions (Diaz et al., 2007; Isbell et al., 2015). A wealth of research use functional trait-based approaches, as functional diversity show a greater effect on ecosystem processes rather than species diversity (Bardgett & van der Putten, 2014; Hooper et al., 2006; Violle, Reich, Pacala, Enquist, & Kattge, 2014). The distribution of functional traits within and between species of an ecosystem represent the demography and response strategies to disturbance (Poorter & Markesteijn, 2008; Díaz et al., 2004; Wright, 2004, Serbin et al., 2019). Both single and multi-trait indices have been shown to predict variations in ecosystem processes under changing conditions (Zhu, Jiang, & Zhang, 2016). Hence, direct observations of functional traits are widely utilized at various spatial and temporal scales, mostly in forested ecosystems to elucidate overall structure, function, and diversity (Funk et al., 2017, Wieczynski et al., 2019). In addition, previous research has utilized functional traits in assessing community assembly processes across a variety of traits (e.g., Pakeman & Stockan, 2014; Medeiros et al, 2019). However, it is important to use multiple dimensions of functional diversity as it is not possible to completely represent the functional diversity of a community as a single number (Ludwig and Reynolds 1988).

Moreover, selection of representative functional traits and scaling those traits are critically important as ecosystem to global scale processes are a function of combined traits of co-occurring species and their abundance (Funk et al., 2017). However, large-scale functional diversity measurements are strongly limited due to the lack of spatially continuous data sets (Jetz et al., 2016). Availability of remote sensing data at fine to coarse spatial and temporal scales facilitates upscaling traits relevant for ecosystem service management (Martínez et al., 2018; Braun, Damm, Hein, Petchey, & Schaepman, 2020). Among many, the traits that represent canopy architecture (morphological functional traits) show direct relationships between carbon storage (Rödig et al., 2019), habitat distribution and quality (Bae et al., 2019), and biodiversity (Bagaram, Giuliarelli, Chirici, Giannetti, & Barbati, 2018), and are widely used to characterize regional to global scale ecosystem processes. Importantly, lidar remote sensing provides opportunities to accurately calculate the morphological functional traits in order to map the morphological diversity at local and regional scales (Schneider et al. 2017). With the launch of the Global Ecosystem Dynamics Investigation (GEDI) mission, we have new opportunities to map functional traits and biodiversity at global scales (Marselis et al., 2019; Rödig et al., 2019; Schimel & Schneider, 2019). Nevertheless, the performance of GEDI in estimating functional traits and diversity in different ecosystems, and especially in semi-arid ecosystems with short and sparse vegetation is yet to be investigated (Qi et al., 2019).

Semi-arid ecosystems cover approximately 40% of the terrestrial landscape and show high dynamicity in ecosystem structure and function (Conti & Díaz, 2013). Hence, semi-arid ecosystems show an unprecedented influence in global carbon dynamics, productivity, and habitat quality (Poulter et al., 2015). Structure-functioning relationships in frequently disturbed semi-arid ecosystems are unclear, largely due to gaps in spatially continuous data and the weak response of sparse and short height vegetation in optical remote sensing (Kulawardhana, Popescu, & Feagin, 2017, Stavros et al., 2017). In particular, understanding semi-arid ecosystem responses to global change is challenging due to the complex and dynamic interactions among multiple ecosystem functions. To unravel this complexity, assessing the spatial patterns of functional diversity and the abiotic controls are critically important (Schlesinger et al. 1990, D'Odorico, Bhattachan, Davis, Ravi, & Runyan. 2013). Especially, understanding how community assembly is controlled by the balance of abiotic drivers is important in predicting the response of ecosystems to environmental change (Pakeman & Stockan, 2014). Further, mapping functional diversity at continuous spatial scales helps to constrain model accuracies of ecosystem processes at landscape scales in different regions across the globe (Braghiere et al., 2019; Dashti et al., in review; Stavros et al., 2017).

In this study, we demonstrate trends and patterns of functional diversity derived from three functional traits using both airborne, the gold standard of measuring morphological traits, and spaceborne lidar (GEDI) in a semi-arid ecosystem. We selected three functional traits: canopy height (CH), plant area index (PAI) and foliage height diversity (FHD) that can be accurately measured from airborne lidar and are widely used to evaluate ecosystem structure - function relationships in a range of ecosystems from forest to savanna. We evaluated the spatial pattern of functional diversity derived from NASA Airborne Snow Observatory (ASO) small footprint airborne lidar (ASO hereafter) and GEDI with respect to abiotic controls to demonstrate GEDI's potential of quantifying the semi-arid ecosystem processes at regional scales. We selected a range of abiotic controls including topography, distance to water, topographic wetness index, soil, and disturbance.

Materials and Methods

Study area

The study was carried out in the Reynolds Creek Experimental Watershed (RCEW) in southwest Idaho (Figure. 3.1). The study area is characterized by a range of topography (1100 m – 2200 m) and vegetation communities. While many varieties of grass, forbs, and shrubs dominate the low elevations, trees of aspen (*Populus tremuloides*), juniper (*Juniperus occidentalis*), and Douglas fir (*Pseudotsuga menziesii*) mark the high elevations. However, low stature sagebrush (*Artemisia tridentata*), bitterbrush (*Purshia tridentata*), and grass of varying densities and cover are found throughout RCEW. In addition, riparian vegetation with cottonwood and willow are found within valleys, and along streams across the watershed. Further, the study area experiences topography-dependent mean annual temperature and precipitation regimes that vary between 4.6–9.2 ^{o} C and 230-959 mm, respectively. The study area has experienced prescribed and natural fires and supports grazing. As a consequence, invasion of cheatgrass in native shrub areas and juniper encroachment have occurred in this study area during the last few decades.

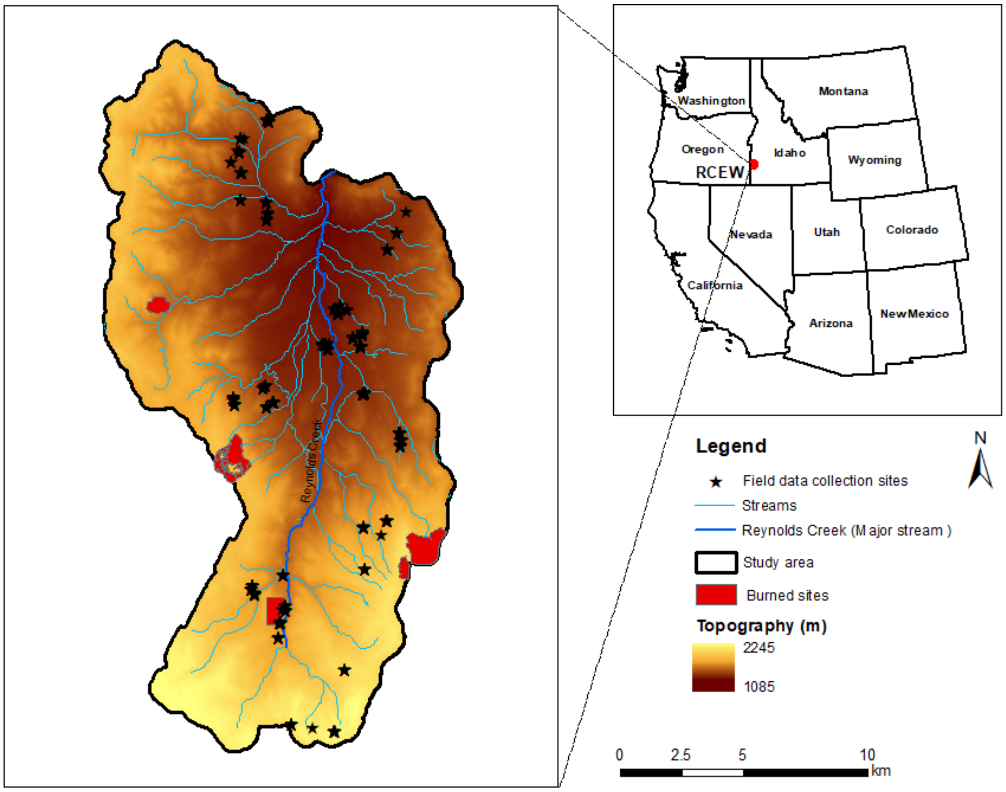

Washington
Montana
Oregon
Idaho
Wyoming
RCEW
Nevada
Utah
Colorado
California
Arizona
NewMexico
Reynolds Creek
Legend
N
Field data collection sites
Streams
Reynolds Creek (Major stream)
Study area
Burned sites
Topography (m)
2245
1085
0
2.5
5
10
km

Figure 3.1 Reynolds Creek Experimental Watershed, SW Idaho with the topographic gradient and stream network. The black stars represent 10 m x 10 m field plots across the watershed.

Field data

Reference field data were collected at 10 m x 10 m plots randomly selected over the study area (Figure. 3.1). We established 5 transects at 1 m, 3 m, 5 m, 7 m, and 9 m in each plot. We collected canopy heights of all the shrubs, plant area index (PAI) and images at 2 m intervals along each transect totaling 20 measurements per plot using a ceptometer (AccuPAR LP-80, Decagon Devices Inc., Pullman, WA, USA) and a camera (Nikon COOLPIX AW120) respectively. Plot scale PAI was calculated by averaging the 20 measurements (Glenn et al., 2017). The collected images were extensively analyzed using the “Samplepoint” freeware program to estimate the species abundance presented within each plot (v1.59, Booth, Cox, & Berryman, 2006). Each photo approximately covered 2 m^2 on the ground. We placed 100 equally spaced grid points within each photo and identified the material under each grid point as vegetation species, dead wood, litter or as bare ground which were later used to calculate the percent cover of each material in those plots.

Environmental data

The topographic variables of altitude, slope, aspect, and topographic wetness index were calculated using the small footprint airborne lidar derived 3m digital elevation model. The slope data were categorized into 10° groups between 0 – 90°. The aspect data were categorized into two major directions as north (+/- 90° – 180° from south) and south (0° - +/- 90° from south). We did not use the flat aspect category as there were not many pixels in this category due to the complex topography of the study area. Stream networks and soil types were retrieved from the Reynolds Creek Critical Zone Observatory database (GIS Server, Reynolds Creek Critical Zone Observatory, 2015). Euclidian distance from streams to each pixel was calculated and used as the distance to water. In addition, we investigated four different fires that occurred in the study area over the past 15 years including the Koke fire (natural fire, 2014), Whiskey mountain fire (prescribed fire, 2005), Break fire (prescribed fire, 2002) and Rabbit creek fire (natural fire, 1996). The burned areas varied from 0.5 km^2 to 2 km^2.

Airborne lidar data

Small-footprint waveform lidar data were acquired August 2014 using the NASA Airborne Snow Observatory (ASO) Riegl LMS-Q1560 (RIEGL Laser Measurement Systems GmbH, Horn, Austria) dual laser scanner (1064 nm wavelength). The mean above ground level of ASO was 1000 m (700 – 1300 m due to terrain conditions) resulting in a footprint of 20 – 60 cm. The study area was scanned at a pulse repetition rate of 400 kHz per laser and the backscattered signal was sampled at 1 ns per sample. All the lidar waveforms in each flight line were Gaussian decomposed following the workflow of Ilangakoon et al., 2018.We derived spatial coordinates, incident angle, pulse width, amplitude, and scattering cross section of all echoes in each waveform. The resulting average point density across the watershed was 10-14 pts/m^2 with positional accuracies of 0.14 m in vertical and 0.11 m in horizontal directions.

Derivation of bare ground

Finding the bare ground was critical in this study as it was the base to calculate all three selected functional traits. Most ground derivation methods initialize with the last return, the lowest elevation points at defined pixel sizes, regardless of the type of lidar (waveform or discrete return). However, low-stature vegetation tends to widen the waveform without a separate vegetation pulse. Thus, defining ground using the last return may remove some vegetation responses especially in this ecosystem. Thus, in our method, we used pulse width deviation to initialize the bare ground search. We used a 3.2 ns pulse width threshold to separate bare ground from vegetation based on Ilangakoon et al. (2018). The initial ground surface was generated at 3 m spatial resolution using minimums of last returns with pulse widths < 3.2 ns using thin plate smoothing. From the initial ground plane, the noisy below ground points (points below 1 m within 0.5 m radii) were removed. Then, the mean of the last returns with pulse width < 3.2 ns were used to generate the final ground plane at 3 m resolution. We selected 3m in order to have at least a few ground returns in dense shrub/tree areas. Though overlapping flight lines provide more returns per given area, mismatching between flight lines can be spurious and sometimes greater than 75 cm, which is well above most shrub heights. Thus, we initialized the point cloud adjustments at flight line level with stepwise increment adding flight line to the process to cover whole study area. The final vertical and horizontal errors of airborne lidar data were 0.14 m and 0.11m respectively.

ASO Functional traits

In this study, we introduce a new method for PAI calculation and canopy height derivation using mutli-angle waveform pulse widths and scattering cross sections. These are derived from the lidar. The workflow for the PAI and the overall workflow to derive the functional traits is displayed below (Figure. 3.2).

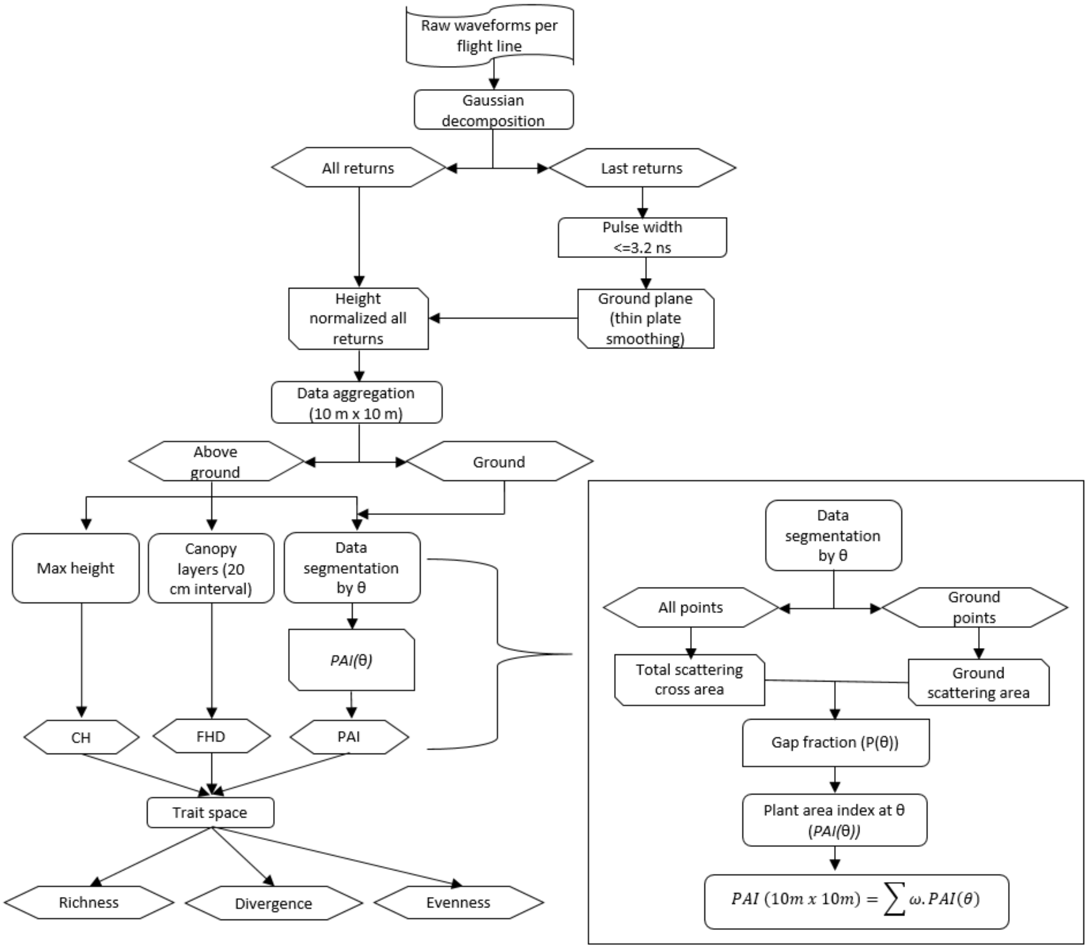

Raw waveforms per flight line
Gaussian decomposition
All returns
Last returns
Pulse width <=3.2 ns
Ground plane (thin plate smoothing)
Height normalized all returns
Data aggregation (10 m x 10 m)
Above ground
Ground
Max height
Canopy layers (20 cm interval)
Data segmentation by θ
PAI(θ)
CH
FHD
PAI
Trait space
Richness
Divergence
Evenness
Data segmentation by θ
All points
Ground points
Total scattering cross area
Ground scattering area
Gap fraction (P(θ))
Plant area index at θ (PAI(θ))
$PAI\ (10m\ x\ 10m) = \sum \omega.PAI(\theta)$

Figure 3.2 Summary workflow for deriving plant area index from small footprint waveform lidar.

Once the ground plane is defined (section 2.5), all lidar returns were normalized using the ground plane. Then, the data were aggregated into 10 x 10 m spatial pixels. This scale was used with the assumption we captured both trees and shrubs, and small variations of functional diversity in shrub dominated areas. At each 10 m pixel, ground and above ground points were separated from all flight lines. To calculate the PAI at different incident angles, the ground and above ground data were separated based on the incident angles. We used groups of incident angles from -30° to +30° (field of view of the laser system) at 5° intervals. Once the data were separated, gap fraction at each incident angle (P(θ)) was calculated as the ratio between sum of scattering cross sections from ground returns to the sum of scattering cross sections to the total returns at a given angle range. The plant area index at θ was then calculated using the following equations (1):

$$PAI(\theta) = L\Omega(\theta) = -\frac{ln(P(\theta))cos\theta}{G(\theta)} \qquad (1)$$

Where, P(θ), G(θ), **Ω**, L are the gap fraction at incident angle θ, extinction coefficient, clumping index, and the true leaf area index (Norman and Campbell, 1989). As we do not separate woody and leaf materials, and do not correct for the clumping effect, the above equation provides the plant area index (PAI). The extinction coefficient G(θ) can be calculated using

$$G(\theta) = \frac{(x^2 + tan^2\,\theta\,)^{0.5}\,\cos\theta}{x+1.774\,(x+1.182)^{-0.733}} \qquad (2)$$

where x is the shape parameter. We used x = 1(spherical leaf distribution) as we did not have inclination measurements. To calculate the PAI of the pixel from all data from all incident angles, we used a weighted sum method counting the differences in number of laser points at each incident angle range.

$$PAI = L\Omega = \omega_1.L\Omega(\theta_1) + \omega_2.L\Omega(\theta_2) + \cdots + \omega_n.L\Omega(\theta_n)$$

$$= -\begin{pmatrix} \omega_1.cos\theta_1 \frac{\ln P(\theta_1)}{G(\theta_1)} \\ + \omega_2.cos\theta_2 \frac{\ln P(\theta_2)}{G(\theta_2)} \\ + \cdots + \omega_n.cos\theta_n \frac{\ln P(\theta_n)}{G(\theta_n)} \end{pmatrix} \quad (3)$$

To calculate the foliage height density (FHD), above ground lidar points were fragmented into 20 cm layers. For each layer, the ratio between numbers of points in each layer to the total number of points was calculated. The total foliage height diversity is then calculated using the following equation (MacArthur & MacArthur 1961).

$$FHD = -\sum p_i\, .\, ln\, p_i \quad (4)$$

Where, p_i is the ratio of foliage in i^{th} canopy layer. Canopy height is the maximum height from all above ground returns from each 10 m x 10 m pixel area.

Once the three traits were calculated at 10 m x 10 m pixel resolution for the whole study area, the trait values within the study area were normalized to vary between to 0 and 1.

GEDI waveform simulation and functional traits

GEDI full waveform lidar signals were simulated for the nominal two year period using the GEDI simulator (Hancock et al., 2019). Discrete return ASO lidar point clouds were used to represent the vertical distribution of surface material to simulate the GEDI waveforms. The GEDI simulator shows that the difference between the height metrics (e.g. RH50) derived from simulated and observed large footprint waveforms (e.g. LVIS waveforms) is minimal when using Riegl laser systems that produce point clouds with pulse density greater 3 beams/m^2 (Hancock et al., 2019). In our study, we used a point cloud with average point density of 14 pts/m^2 derived from a Riegl QLMS 1560 dual laser scanner (Ilangakoon et al., 2018). Hence, we assume that the simulated GEDI waveforms closely represent the actual GEDI waveforms. To account for data loss due to clouds, we used a conservative (based on our study area) estimate of ~50% for the mean global cloud cover to derive the GEDI track patterns under cloudy conditions. A GEDI waveform was modelled as a Gaussian shaped laser pulse with a FWHM of 15.6 ns (an effective footprint size of 22 m) at a sampling rate of 15 cm. We added Gaussian noise to the simulations to match the expected performance of the GEDI power beams by night. The noisy simulated waveforms were smoothed using a Gaussian function and the mean noise was subtracted with a threshold equal to 3.5 times the standard deviation (Hofton, Minster, & Blair, 2000). The structural parameters of canopy height (CH), foliage height diversity (FHD), and plant area index (PAI) were extracted using the GEDI metric calculator. The GEDI metric calculator uses the Gaussian decomposition to define the ground and to generate the relative height (RH) metrics. The GEDI estimate of canopy height was considered as the 98th percentile (RH98) of the cumulative energy of the waveform (Drake et al., 2002). The FHD metric was used as the foliage height diversity. The sum of PAIs calculated based on the Gaussian decomposed data (gLAI 10 – gLAI 40) were used as the PAI (Tang et al., 2012).

In addition, we generated another set of GEDI waveforms using point clouds clipped from field plots to assess the correlation between field observed and GEDI derived vegetation structure. Though we used the same simulation process as described above, we modified the point cloud input to the simulator to match the GEDI footprint size (25 m). To do that, we created 25 m footprints surrounding each 10 m x 10 m plots and clipped the point clouds within each footprint. The vegetation points beyond the 10m x 10 m plot boundary were removed. By removing the vegetation points beyond the plot boundaries, this allowed us to segregate the GEDI signal from vegetation within the field plots. Hence, we can directly compare the GEDI metrics from the simulated waveforms at the plot scale to the field observed vegetation heights and PAI.

Functional diversity

We calculated the functional diversity around each pixel at a rectangular neighborhood with a side length of 500 m. Initial mapping of functional traits showed that 500 m is the minimum spatial resolution that could produce a wall to wall map using GEDI footprints for the study area. The average number of GEDI footprints per 500 m pixels was 20. The neighborhoods were selected using a moving window approach following Schneider, et al. (2017). PAI, CH, and FHD of all the pixels within the rectangular neighborhood were selected and the functional diversity indices of richness, divergence, and evenness were calculated following Villéger, Mason, & Mouillot. (2008) and Schneider et al. (2017). The resulting values were assigned as the center pixel functional diversity.

The functional richness is the convex hull volume covered by the PAI, CH, and FHD of the pixels within the selected neighborhood mapped in the 3-D trait space. Functional divergence (FDiv) is calculated as:

$$\Delta|d| = \sum_{i=1}^{S} \frac{1}{S} \left| dG_i - \overline{dG} \right| \quad (\text{Schneider et al.}, 2017), \qquad (5)$$

$$FDiv = \frac{\overline{dG}}{\Delta|d| + \overline{dG}} \quad (\text{Schneider et al.}, 2017), \qquad (6)$$

Where, S is the number of pixels used to map the functional divergence, dGi is the Euclidean distance between the ith pixel and the centre of gravity of 3D trait space defined by CH, FHD, and PAI as trait axes, and $\overline{dG}$ is the mean distance of all pixels to the centre of gravity.

The calculation of Functional evenness (FEve) requires partial weighted evenness (PEW) measurements. To calculate the PEW, the euclidiance distances between nodes were calculated using the minimum spanning tree method in MATLAB (Prim, 1957). In this study, a node is a pixel defined in the 3D trait space by its PAI, FHD, and CH trait axes. The minimum spanning tree was drawn connecting all the pixels in the functional space using the minimum possible total edge weight.

$$PEW_i = \frac{EW_l}{\sum_{i=1}^{S-1} EW_l} \tag{7}$$

$$FEve = \frac{\sum_{l=1}^{S-1} \min\left(PEW_l\ , \frac{1}{S-1}\right) - \frac{1}{S-1}}{1 - \frac{1}{S-1}} \tag{8}$$

Where, EW_l is the euclidian distance of branch *l* in the minimum spanning tree, and S-1 is the number of branches. Here, S is the number of pixels used in the pixel space (500 m rectangular space in this case). The subscript i represents the 500 m rectangular neighborhood space used to calculate the functional diversity.

Statistical analysis

To analyze the effects of abiotic factors on functional diversity, we randomly selected 200 data points across the watershed that capture the full range of functional diversity and the environmental factors. We resampled the environmental factors (elevation, slope, aspect, topographic wetness index, distance to the nearest stream, and soil type) to the same 500 m spatial resolution of the functional diversity. We used the average of the continuous values (e.g. elevation). For categorical variables (e.g. soil type), we used the major category within 500 m space. We set the minimum distance between data points greater than 520 m to avoid the mutual inclusion of niche spaces (500 m). A generalized linear model was applied to the scaled environment variables. The relative importance of each factor in the linear model was used (R package relaimpo, calc.relimp) to assess the capability of all factors together and each factor separately to explain the variance of each diversity index derived from ASO and GEDI. We further used ANOVA type 1 to show the statistical significance of each environmental variable to explain the functional diversity.

Results

We derived three functional traits (CH, FHD, and PAI) from both ASO and GEDI full waveform signals. Our field observations showed a range of functional trait values for PAI (0.19 – 1.88) and CH (0.31 m – 2.52) with mean values of 0.76 and 1.20 m, respectively. Field observed functional traits showed a high correlation (R2 = 51 - 77%) with ASO derived functional traits and a moderate correlation (R2 = 35 - 45%) with those derived from simulated GEDI waveforms (Figure. 3.3). The RMSE of GEDI derived canopy heights and PAI were 0.60 m and 1.39, respectively. However, the bias in height estimates from GEDI in this study is less than the bias observed in a previous study of a forested environment (2.35 m – 3.83 m) (Qi et al., 2019). GEDI shows difficulty in estimating vegetation heights less than 2 m (red circle in Figure. 3.3c). We further observed that both ASO and GEDI derived PAI were significantly correlated to their CH and FHD values (Figure. 3.4).

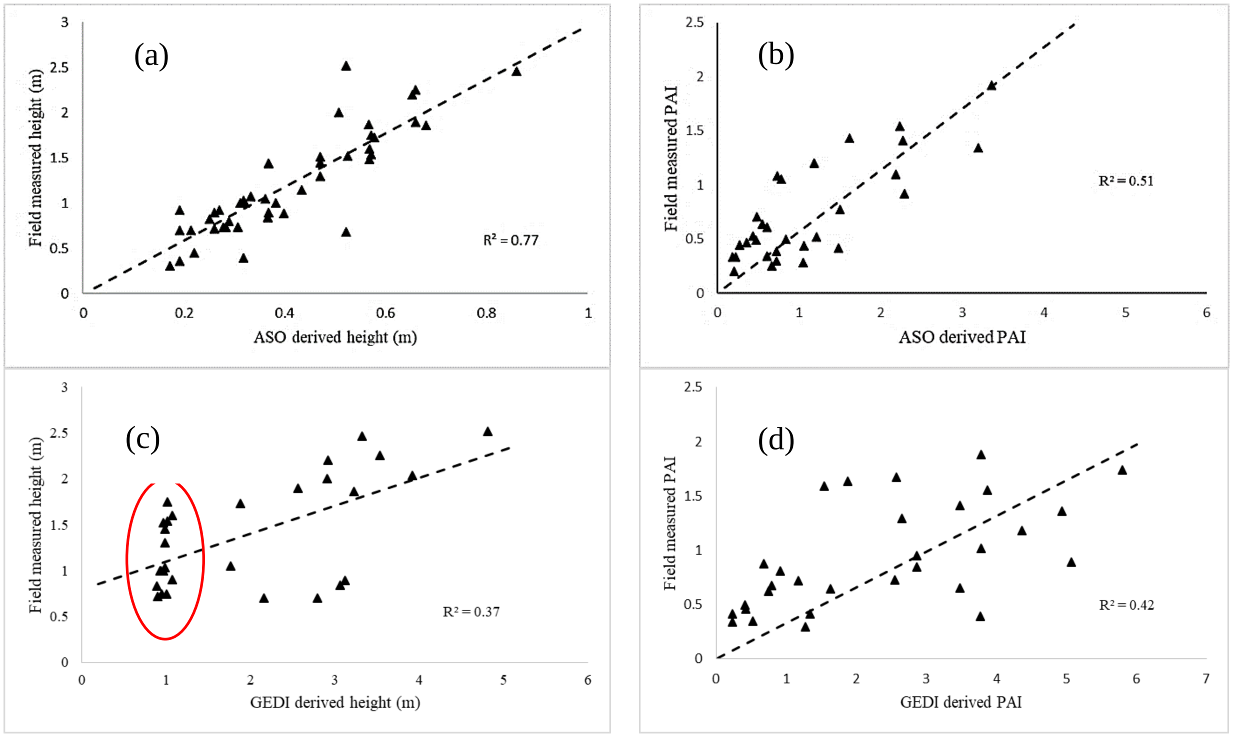


Figure 3.3 Correlation between a) field measured max vegetation heights and ASO derived vegetation heights, b) field measured plant area index (PAI) and ASO derived PAI, c) field measured max vegetation heights and GEDI derived vegetation heights, and d). field measured PAI and GEDI derived PAI .

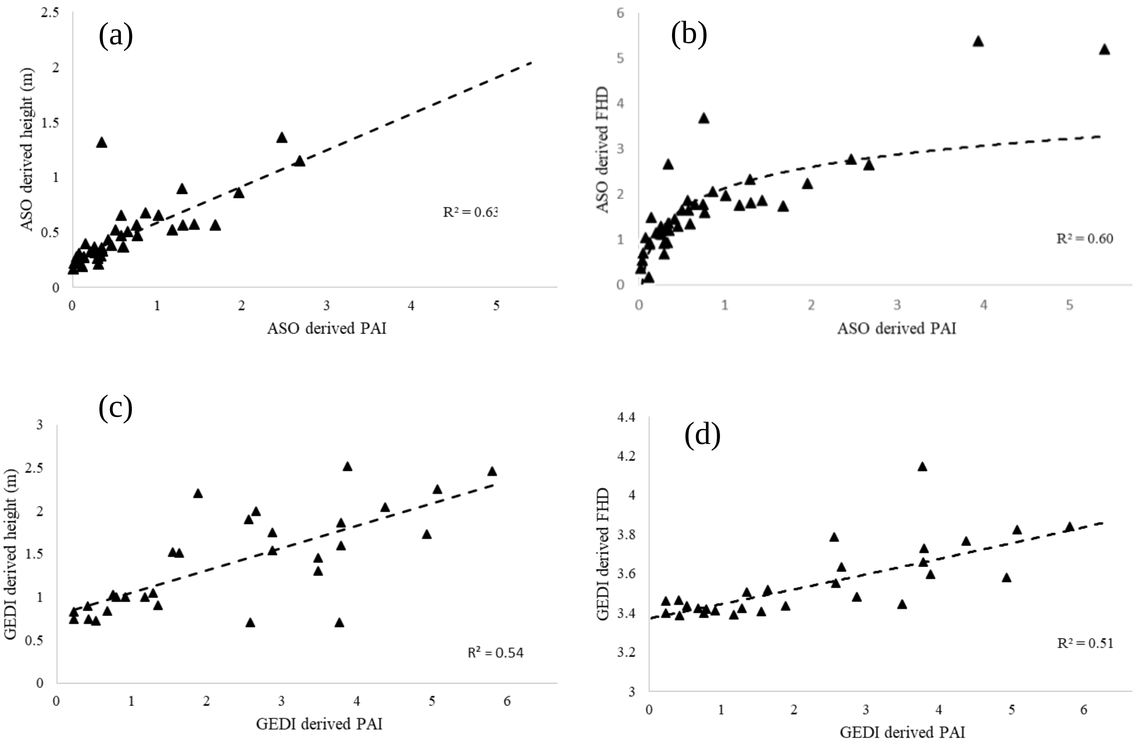


Figure 3.4 Correlation between a) field measured max vegetation heights and ASO derived PAI, b) ASO derived FHD and ASO derived PAI, c) field measured max vegetation heights and GEDI derived PAI, and d). GEDI derived FHD and GEDI derived PAI .

Figure. 3.5 displays the distribution of morphological traits of PAI, FHD, and CH of the study area at 10 m pixel size for ASO and at footprint scale (25 m) for GEDI. Morphological traits vary from high (dark blue) to low (brown). The ASO derived PAI and FHD varied between 0 – 6.87 and 0– 4.9 while those from GEDI varied between 0.0 – 8.3 and 3.3 – 5.2 (Figure. 3.5a, b, and d, and e). Due to the vegetation height range (0 – 29 m) of the study area, the canopy height variability within shrub dominant regions (mostly between 0- 1 m) is not clearly visible in the watershed scale map, hence, the log canopy height map is presented for both ASO and GEDI (Figure. 3.5c and f). The highest PAI, FHD and CH (dark blue regions) are in the southern portion of the study area. These densely vegetated regions consist of a number of tree and shrub species. The north and central portions of the study area show a range of ecosystem characteristics ranging from low to high vegetation cover (~6% to over 60 %), however, the height range is lower than in the southern portion of the watershed. The fine resolution of traits from ASO, especially FHD and log canopy height, results in larger distributions across the watershed in contrast to the coarser resolution traits from GEDI. However, the tree-shrub ecotone transition from south to north is clearly visible in all trait maps except in the GEDI derived PAI (Figure. 3.5d). Central and eastern portions of the watershed are comprised of low, but consistent trait values (yellow- brown areas in all sub figures except Figure. 3.5d). These areas are dominated by agriculture and grazing. Among all traits studied, FHD and log canopy height distinguish not only tree and shrub dominancy, but also variations in shrub cover and density (referring to the variations of green color in the shrub dominant central and north regions of the watershed in Figures. 3.5b, c, e, and f). Our field measurements of shrub cover and density confirmed the relationships between FHD and shrub cover and density, and had strong correlations with both ASO and GEDI derived FHD.

Moreover, we evaluated the functional diversity (functional richness, functional evenness, and functional divergence) around 500 m neighborhoods of each pixel center throughout the watershed with both ASO and GEDI derived morphological traits. Figure. 3.6 displays the diversity distribution of the study area from ASO and GEDI respectively. Both ASO and GEDI showed very low functional richness values and varied between 0.0 – 0.11 and 0.0 – 0.2, respectively (Figure. 3.6a and d). The evenness and divergence showed relatively higher values compared to the richness. The ASO evenness and divergence ranged between 0.33- 0.99 and 0.55 – 0.87 (Figure. 3.6b and c) while those of GEDI varied between 0.58- 0.85 and 0.57 – 1.0 (Figure. 3.6e and f) correspondingly. The tree-dominant southern portion of the study area is characterized by high richness with low evenness whereas the shrub-dominant central and northern portions of the watershed are characterized by the opposite. Highest richness was observed at densely vegetated tree dominant regions while the lowest was at sparse shrub dominant regions with shrub cover less than 30%. Highest evenness was observed mostly in the eastern and some northwest areas. These areas are used for grazing and are shrub and grass dominant. Evenness and richness differentiate the tree-shrub ecotones and relative densities within shrub dominant areas. Further, relatively low evenness is reported along the stream network of the watershed compared to its surroundings (refer to Figure. 3.1 for the stream network). A variable functional diversity was observed throughout the watershed regardless of the tree-shrub ecotone distribution. Both GEDI and ASO diversity maps provide similar patterns across the watershed. The ASO and GEDI showed 54%, 28% and 13% correlation between functional richness, functional evenness, and functional divergence, respectively (Figure. 3.7). The ASO data showed 64% inverse correlation between evenness and richness.

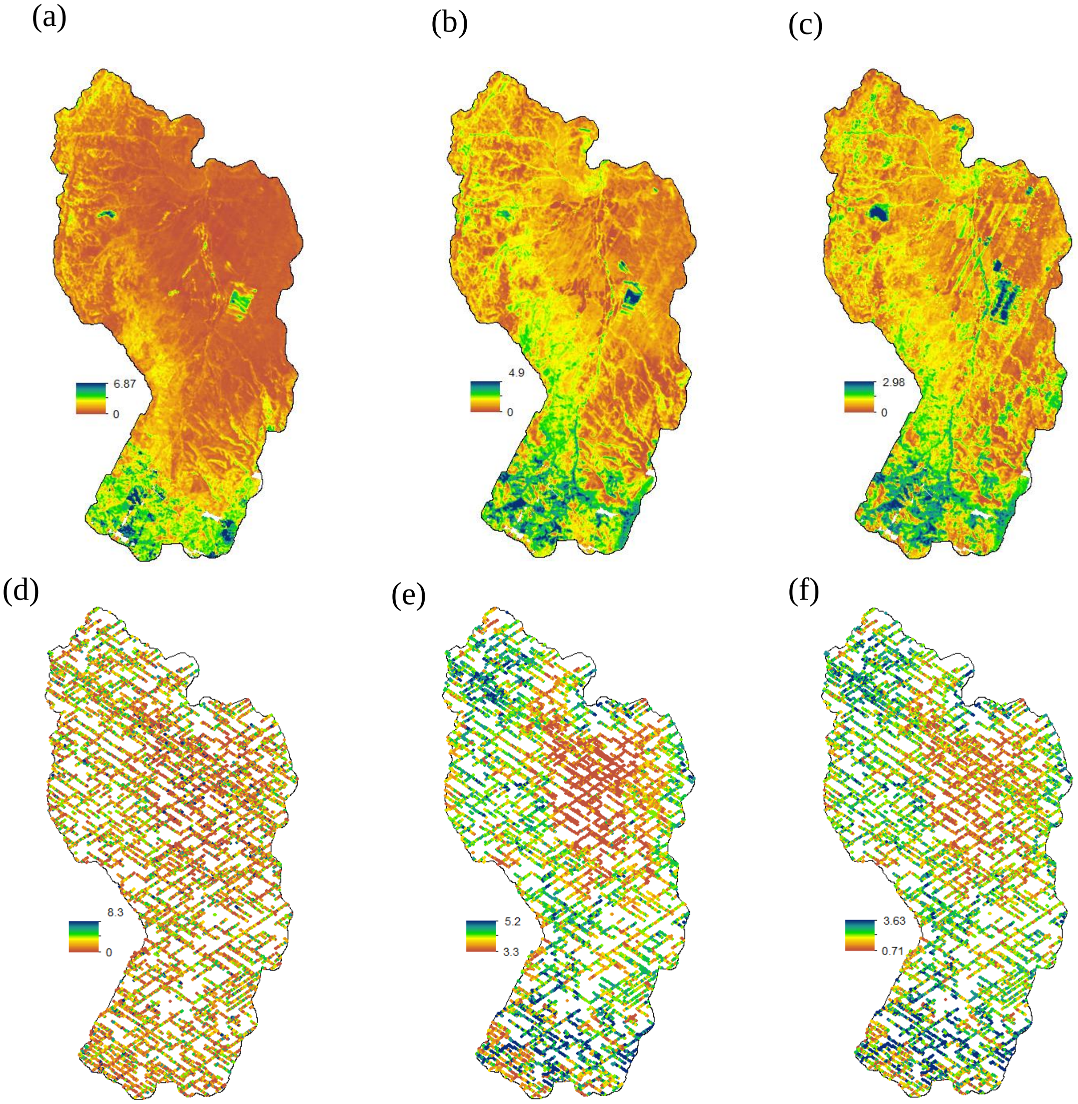


Figure 3.5 ASO and GEDI derived functional traits distribution across the RCEW. Top row: a). ASO derived PAI, b). ASO derived FHD, c). ASO derived CH. Bottom row: d). GEDI derived PAI, e). GEDI derived FHD, f). GEDI derived CH. The ASO based maps were derived at 10 m spatial resolution. GEDI data are displayed at footprint scale. The ASO and GEDI canopy heights are displayed as log canopy heights to enhance the visualization of canopy height distribution across the study area.

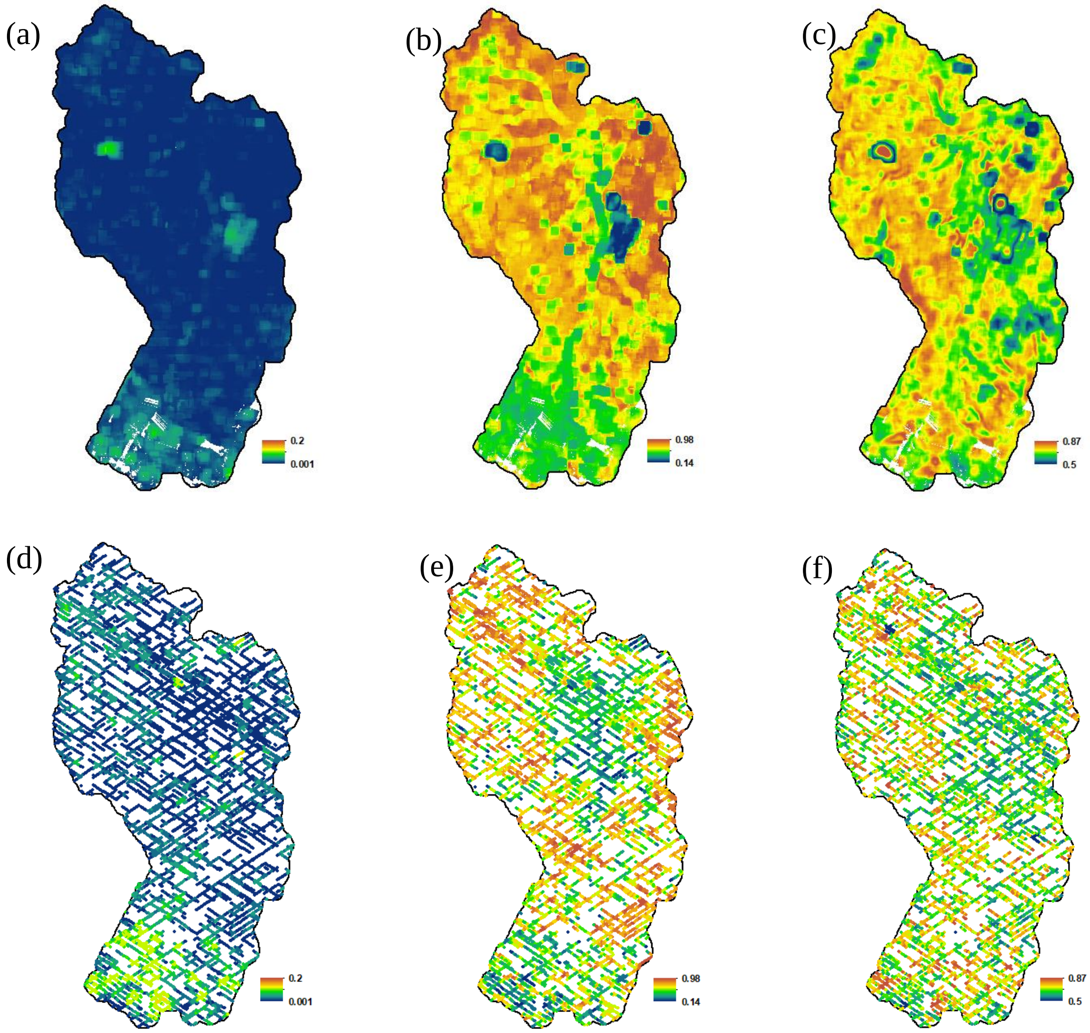
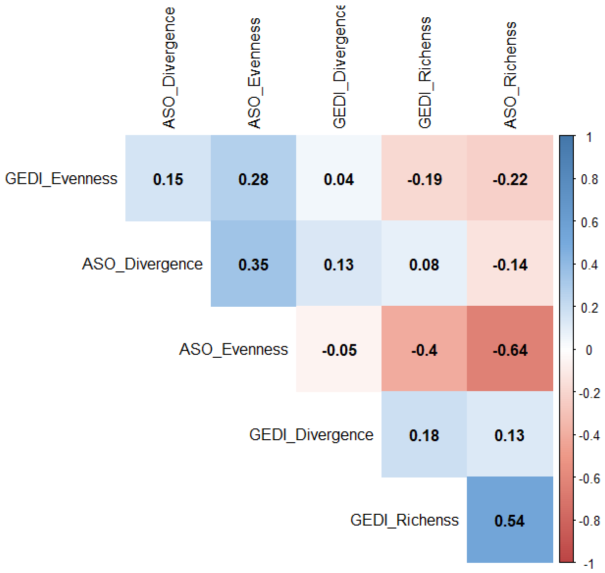

(a)
0.2
0.001
(b)
0.98
0.14
(c)
0.87
0.5
(d)
0.2
0.001
(e)
0.98
0.14
(f)
0.87
0.5

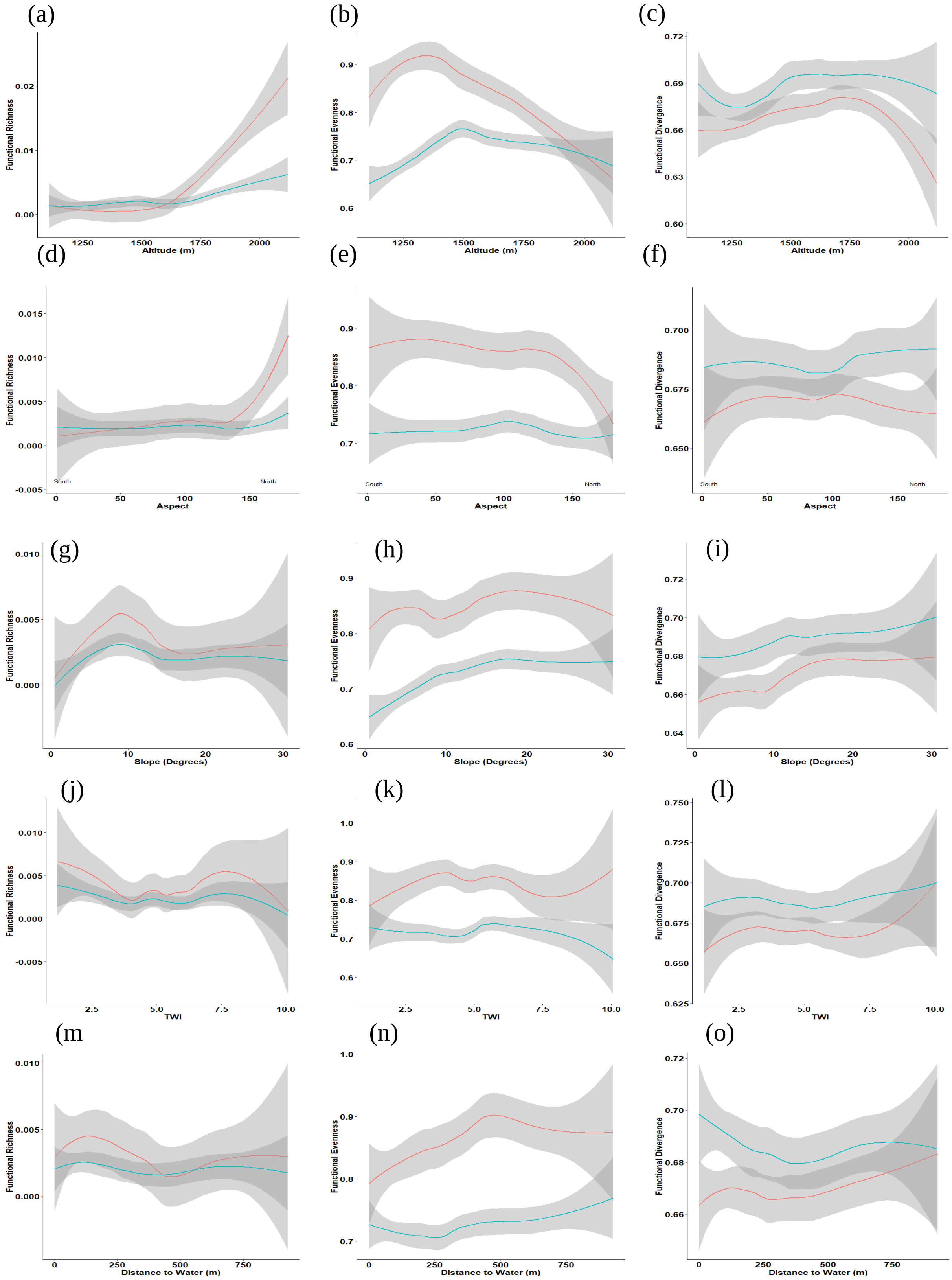
Figure. 3.6 Functional diversity derived using 500 m spatial neighborhood from ASO (top row- richness, evenness, divergence), and GEDI (bottom row- richness, evenness, divergence). Functional richness, functional evenness and functional divergence of RCEW derived from functional traits distribution across 500 m neighborhood to each 10 m x 10 m pixel in ASO and 25 m x 25 m footprints in GEDI.



	ASO_Divergence	ASO_Evenness	GEDI_Divergenc	GEDI_Richenss	ASO_Richenss
GEDI_Evenness	0.15	0.28	0.04	-0.19	-0.22
ASO_Divergence		0.35	0.13	0.08	-0.14
ASO_Evenness			-0.05	-0.4	-0.64
GEDI_Divergence				0.18	0.13
GEDI_Richenss					0.54

Figure. 3.7 Correlation between ASO and GEDI based diversity indices at 500 m spatial neighborhood.

Functional diversity indices demonstrated a range of trends with environmental gradients (Figure. 3.8). Both the ASO and GEDI derived diversity showed consistent trends with the selected abiotic variables. At altitudes below 1500 m, the richness showed little to no change whereas above 1500 m, a drastic increase with ASO and a mild increase with GEDI was observed (Figure. 3.8a). With functional evenness, a consistent but opposite trend was demonstrated with the altitude (Figure. 3.8b). Both ASO and GEDI evenness showed an increase up to 1500 m and then continuously decreased at higher altitudes. The ASO divergence also demonstrated a general increase with elevation up to 1700 m and then decreased dramatically, however, GEDI divergence did not show a clear trend with altitude (Figure. 3.8c). A very similar trend between aspect and functional richness was observed (Figure. 3.8d). While the richness increased with aspect, the evenness and divergence did not show a clear trend with aspect (Figure. 3.8e and f). However, ASO evenness showed a decrease only at north aspects as with altitude (the red curve in Figure. 3.8e). The ASO and GEDI richness increased with slope up to 10° and then showed almost an independent behavior at higher slopes (Figure. 3.8g). Both ASO and GEDI evenness and divergence exhibited an increase with slope (Figure. 3.8h and i). Among other factors examined, only ASO evenness showed an increase with distance to water up 500 m and showed no change thereafter (Figure. 3.8n). Our generalized linear models developed for diversity indices with abiotic factors demonstrate the variance of each diversity explained by the abiotic factors (Figure. 3.9). With all the environmental factors combined (elevation, slope, aspect, soil, distance to water and topographic wetness index) 24%, 16% and only 4% of the variance of ASO richness, evenness and divergence, can be explained. For GEDI, these abiotic factors could explain 20%, 26%, and 15% of the variance of richness, evenness, and divergence. Within that, the results further show that functional richness is strongly controlled by elevation (77% in ASO and 80% in GEDI) (Figure. 3.9a and d). The influence of all factors were generally less than 10%. The functional divergence is mainly influenced by slope (80% in ASO, 75% in GEDI) (Figure. 3.9c and f). In addition, elevation showed some influence on functional divergence (ASO- 16%, GEDI- 19%). In contrast, while the ASO evenness is mainly influenced by elevation (41%) and soil (48%), the GEDI evenness is mainly influenced by slope (88%) (Figure. 3.9b and e). All other factors showed less than 10% importance in explaining the evenness derived from both ASO and GEDI.


(a)
Functional Richness
0.00
0.01
0.02
1250
1500
1750
2000
Altitude (m)
(b)
Functional Evenness
0.6
0.7
0.8
0.9
1250
1500
1750
2000
Altitude (m)
(c)
Functional Divergence
0.60
0.63
0.66
0.69
0.72
1250
1500
1750
2000
Altitude (m)
(d)
Functional Richness
-0.005
0.000
0.005
0.010
0.015
South
North
0
50
100
150
Aspect
(e)
Functional Evenness
0.7
0.8
0.9
South
North
0
50
100
150
Aspect
(f)
Functional Divergence
0.650
0.675
0.700
South
North
0
50
100
150
Aspect
(g)
Functional Richness
0.000
0.005
0.010
0
10
20
30
Slope (Degrees)
(h)
Functional Evenness
0.6
0.7
0.8
0.9
0
10
20
30
Slope (Degrees)
(i)
Functional Divergence
0.64
0.66
0.68
0.70
0.72
0
10
20
30
Slope (Degrees)
(j)
Functional Richness
-0.005
0.000
0.005
0.010
2.5
5.0
7.5
10.0
TWI
(k)
Functional Evenness
0.6
0.7
0.8
0.9
1.0
2.5
5.0
7.5
10.0
TWI
(l)
Functional Divergence
0.625
0.650
0.675
0.700
0.725
0.750
2.5
5.0
7.5
10.0
TWI
(m)
Functional Richness
0.000
0.005
0.010
0
250
500
750
Distance to Water (m)
(n)
Functional Evenness
0.7
0.8
0.9
1.0
0
250
500
750
Distance to Water (m)
(o)
Functional Divergence
0.66
0.68
0.70
0.72
0
250
500
750
Distance to Water (m)

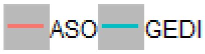


Figure. 3.8 Functional diversity – environmental gradient trends of RCEW. The red and blue curves represent the mean variation of diversity indices from ASO and GEDI respectively and the surrounding gray area represents the standard deviation. From top to bottom variation of functional diversity with altitude (a,b,c), aspect (d,e,f), slope (g,h,i), topographic wetness index (TWI) (j,k,l), and distnace to water (m,n,o) are displayed.

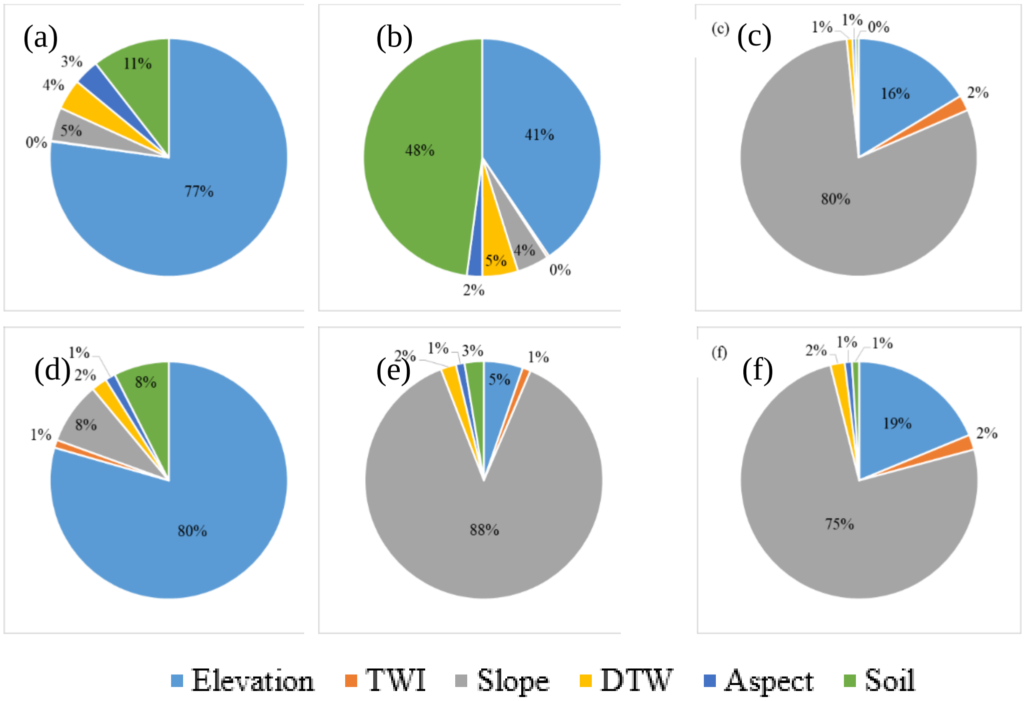


Figure. 3.9 Variance of functional diversity explained by each abiotic factor; a). ASO functional richness, b). ASO functional evenness, c). ASO functional divergence, d). GEDI functional richness, e). GEDI functional evenness, f). GEDI functional divergence. TWI – topographic wetness index, DTW – distance to the nearest stream in meters.

We further investigated functional traits and diversity shifts in response to fire disturbances. Figure. 3.10 and Figure. 3.11 are the functional traits and diversity distributions derived from ASO within disturbed and surrounding undisturbed areas of the fires. The ASO traits data are at 10 m spatial resolution to compare with the field observations. Frequency distributions of all three traits show a shift towards lower values in disturbed areas than the undisturbed areas. The PAI shows a shift of trait values towards the lower end than the other two traits regardless of whether a fire is prescribed or natural (Figures. 3.10b, e, h, and k). Further, a shift of traits is clearly visible in recently burned areas than the oldest Rabbit Creek burn scar revealing the trends of trait recovery. The shift of functional traits clearly demarcates a sharp fire boundary from the prescribed fires. In the natural fires (Kirk and Rabbit), the burned region can be identified, however, due to the diffuse nature of natural disturbance events, the boundaries of the disturbance cannot be demarcated. Functional richness showed a minor increase compared to the surrounding of all disturbed areas except the very recent Koke fire (Figure. 3.11d, g, and j). The functional evenness has also increased while the functional divergence decreased in the fire-disturbed areas compared to their surroundings (except Whiskey fire divergence- Figure. 3.11f).

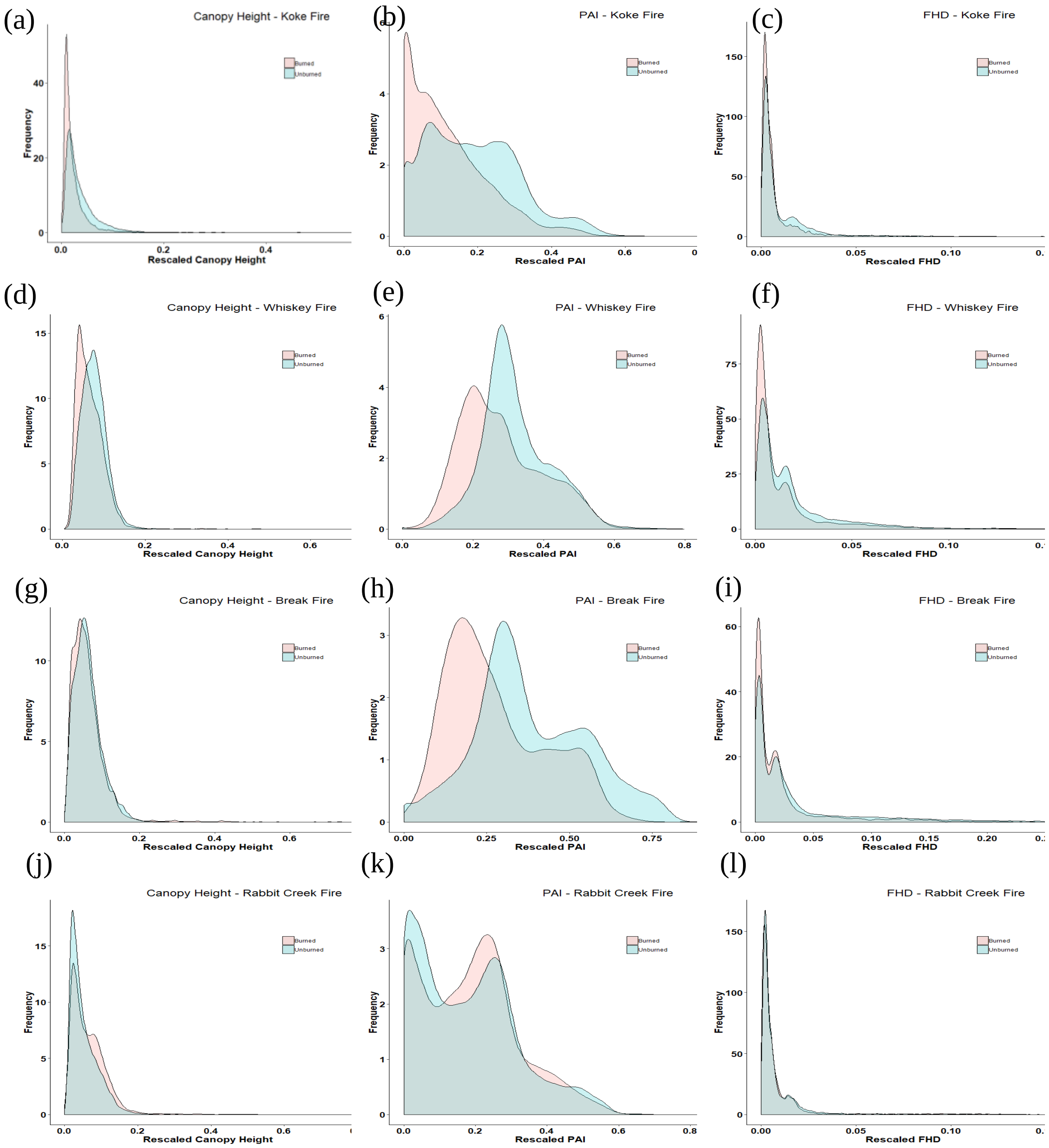


Figure. 3.10 Frequency distribution of functional traits within the fire disturbed and surrounding undisturbed areas of RCEW. The pink represents the functional traits from burned areas while the green represents the functional traits of undisturbed areas.

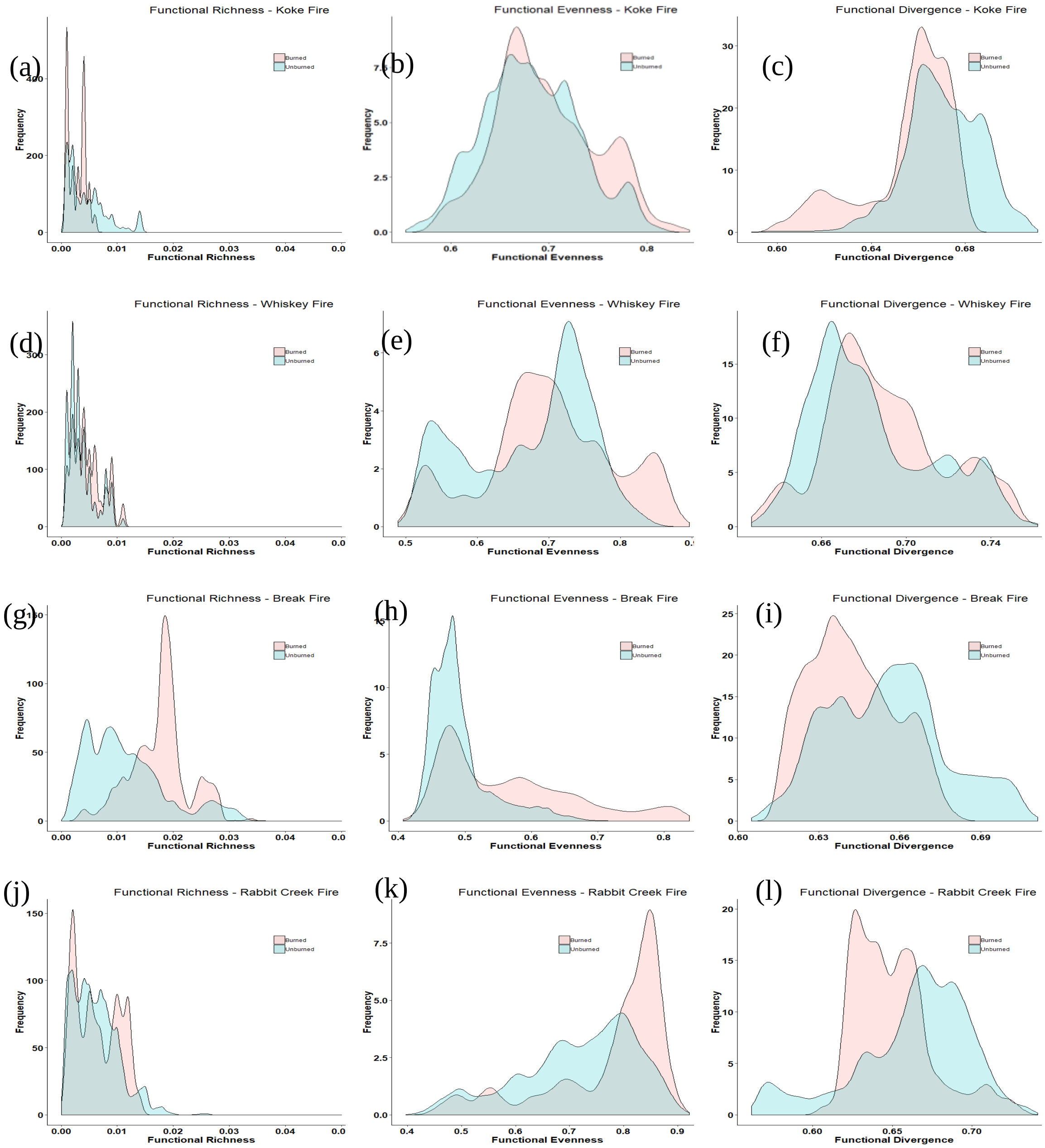


Figure. 3.10 Frequency distribution of functional diversity indices of burned and unburned areas of RCEW. The pink represents the functional diversity from burned areas while the green represents the functional diversity of undisturbed areas.

We used boxplots instead of frequency distribution curves to show the functional traits and diversity of burned and unburned areas from GEDI, as the number of GEDI pixels in the burned areas was low (Figure. S.1 and Figure. S.2). Neither functional traits nor diversity showed a clear relationship with fire disturbance, likely due to the sparseness and low number of GEDI footprints (2-42) within 500 m neighborhood (Figure. S.1 and S.2). However, at least three of the four fires showed a decrease in the distribution of functional traits after fire (Figure. S.1). The GEDI functional diversity did not show clear shifts between disturbed and undisturbed areas. Finally, we calculated the functional diversity of the four fires at a range of spatial neighborhood sizes with 30 m increments to enhance our understanding of the scale dependency on functional diversity (Figure. 3.12). The time series of fires (over 15 years) also allowed us to evaluate the functional recovery post fire. Our analysis showed that the functional richness of the disturbed and undisturbed areas are similar when we use smaller neighborhoods and increase with the increment of the spatial neighborhood (Figure. 3.12a). The difference of functional evenness and divergence between disturbed and undisturbed areas is very high with smaller spatial neighborhoods and decreases with an increase of the spatial neighborhood (Figure. 3.12b, and c). Further, we observed that functional divergence differences decrease when increasing the time since disturbance, and thus showing a trend of functional recovery (Figure. 12c).

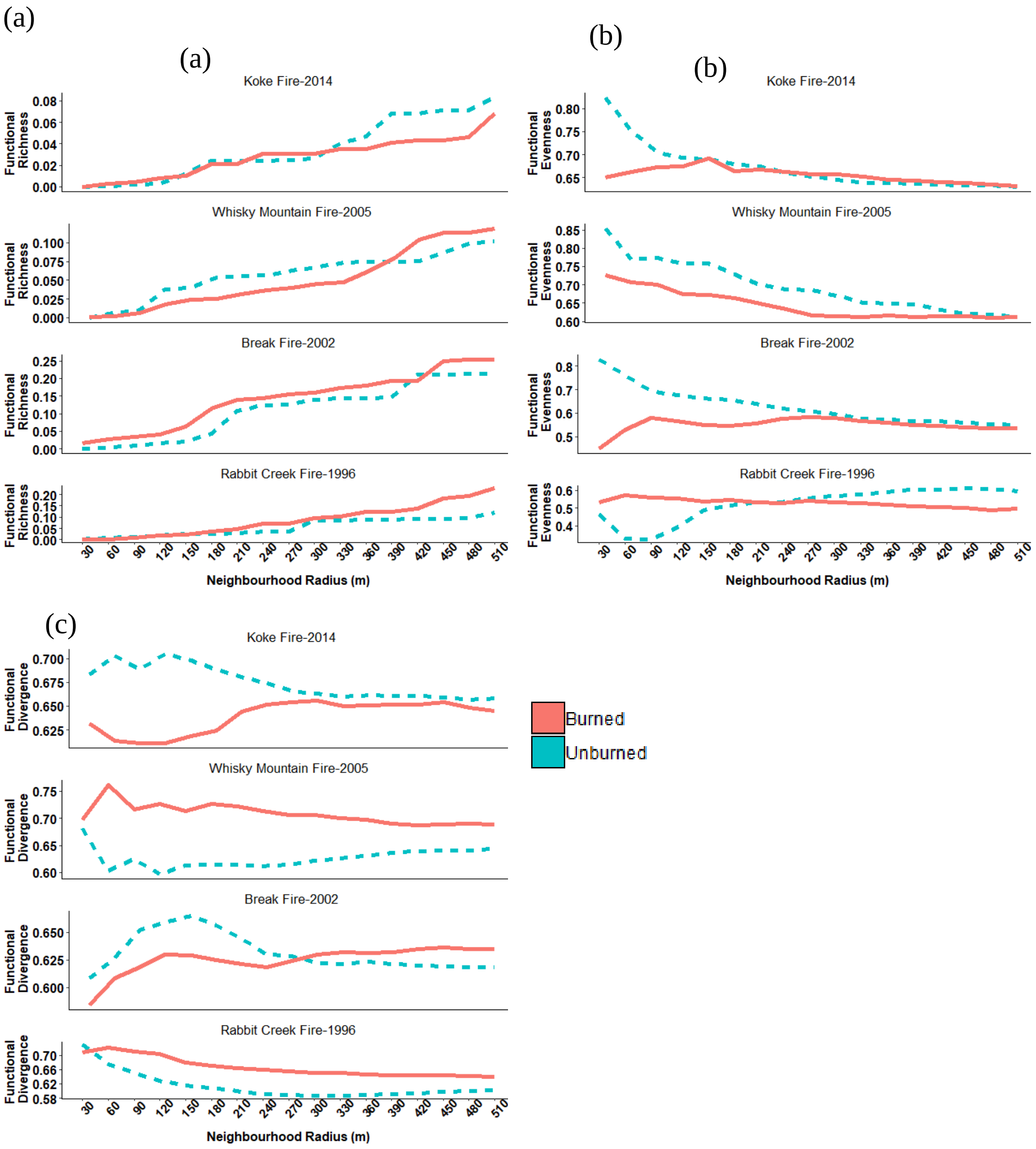


Figure. 3.11 Functional diversity variability with search neighborhood radius at four different burned regions. Pink represents the burned areas and the blue represents the surrounding unburned areas.

Discussion

Mapping functional traits

Mapping functional traits and diversity at continuous spatial scales in semi-arid ecosystems and identifying their abiotic controls allow us to understand the ecosystem function – processes relationships and trends. We demonstrate the potential of GEDI to map functional traits and diversity of semi-arid ecosystems. The moderate correlation between GEDI and field estimates of canopy height and PAI show the limitation of GEDI in these heterogeneous semi-arid ecosystems. Though a low correlation is expected from large footprint lidar systems over short canopies where remote sensing is highly challenging, our results show higher correlations than previous studies (e.g. Qi & Dubayah, 2016; Qi et al., 2019). The wider pulse widths of GEDI (15.6 ns) result primarily in single peaks over short height vegetation, combining responses from both vegetation and the ground. Previous studies from large footprint laser systems including simulated GEDI show that the RMSE of height estimation is 2.35 m – 3.83 m. Our results confirm that GEDI data can be used to estimate vegetation heights above 2 m with lower RMSE and bias than Qi et al. (2019), who also used simulated GEDI to estimate vegetation heights. More than 50% of our field plots contain shrubs, with an average shrub height greater than 50 cm and max shrub height greater than 1 m. In addition, the average shrub cover of the study plots varies from 6% - 65% allowing the simulated GEDI waveforms to capture and incorporate vegetation into the ground signal. These vegetation heights and high canopy density of some field plots may have led to the relatively high correlation of GEDI.

To calculate vegetation height, PAI and FHD, the separation of vegetation from ground returns or quantifying the vegetation contribution in the ground return is essential. We used the ratio between scattering cross section of lidar points above ground and the total lidar cross section coverage per plot as the PAI. In ecosystems with canopy height less than 2 m, taller vegetation will have better detection of vegetation returns especially from laser systems with wider pulse widths such as ASO (3 ns) and GEDI (15.6 ns). Hence, a high correlation can result between any variables that require vegetation – ground separation to the canopy height (PAI and FHD in this case). In addition, the canopy cover exponentially increases when the shrub height increases, especially in Wyoming big sagebrush (*Artemisia tridentata* ssp. *wyomingensis*) and basin big sagebrush (*A. tridentata* ssp. *tridentata*) (Frandsen, 1983) and has a high correlation with leaf area index ($R^2 = 0.73$) (Olsoy, Mitchell, Levia, Clark, & Glenn, 2016). Utilizing other techniques such Partial Curve-Fitting (PCF) (Zhuang & Mountrakis, 2015) or deconvolution may have a high potential of separating the vegetation from the ground peak, and hence may provide a better estimation of vegetation structure of shrub dominant ecosystems from GEDI. However, we observed a consistent pattern of functional trait distributions with both the ASO and GEDI across the study area revealing GEDI's feasibility to track trends and patterns of functional traits at regional and global scales. All of the functional traits from both ASO and GEDI clearly demarcate the tree-shrub ecotones. ASO lidar further shows the variations of functional traits within tree and shrub ecotones as well as along the watershed's stream network. The consistency between ASO and GEDI functional trait maps further emphasizes the great potential of

GEDI data to help map functional traits in heterogeneous semi-arid ecosystems across the globe.

Functional diversity in semi-arid ecosystems

Our results emphasize that the trends and patterns of functional diversity depend on the relative distribution of the environmental factors considered here (primarily elevation, soil, slope, and aspect). We observed different functional evenness and richness patterns over shrub versus tree ecotones, whereas functional divergence did not depict a trend based on the ecotone. Functional richness is the niche extent in the trait space. The high functional richness in this study is characterized by the extreme values of all traits within the considered niche space (500 m radius). We observed high functional richness at high elevations and especially with north aspects. These areas in the southern portion of the study area receive the highest precipitation throughout the year, mostly as snow, which helps maintain perennial streamflow. In addition, this region shows a mean annual temperature range from 4 – 16 C° providing favorable conditions for vegetation growth. Due to the high dependency with elevation gradient, we did not include precipitation or temperature as abiotic factors. However, future studies including these climatic variables and their temporal variability would help us to understand the temporal dynamics of the functional diversity. Further, most of the southern portion of the watershed is comprised of north facing slopes. Seyfried, Harris, Marks, & Jacob (2000) show that soils of the Harmehl-Gabica association, which are dominant at elevations higher than 1370 m MSL, are associated with deep soil profiles with thick surficial A horizons and contain high organic matter in north facing slopes. Due to the availability of water, soil nutrients, and solar radiation, these areas of the watershed provide favorable conditions for conifers, aspen and mountain sagebrush. These species have diverse morphological characteristics. Though functional richness may decrease with elevation in many other regions due to increased aridity and a decrease in temperatures at high elevations (Durán et al., 2019; Schneider et al., 2017; Wieczynski et al., 2019), the deep soil profile and low aridity at high altitudes in this study area leads to a high functional richness. However, the highest functional richness observed in the study is 0.2. The high topographic gradients over short distances limits the spatial extension of this functionally rich zone. Hence, the vegetation in this southern portion of the study area may have higher competition for resources leading to a lower functional evenness.

In contrast, the lower elevations in the central and north portions of the study area are relatively dry with lower and intermittent precipitation. The northern watershed consists of shallow, rocky soils with mesic soil temperatures favoring sagebrush species. This unique and important plant community in the northern Great Basin ecoregion is referred to as the sagebrush-steppe and is co-dominated by big sagebrush and several perennial grasses and forbs. The functional evenness is the distribution of trait abundance in the occupied niche space. Our functional trait maps show that the trait variability in these regions are minimal. Rather than the abundance, the distribution of structurally similar vegetation in the niche space is reflected by the evenness. This further explains the negative correlation between functional richness and evenness. Sparsely distributed, structurally similar shrubs can effectively utilize the entire range of resources available. Hence, we observe functionally even, but functionally less rich landscapes in the lower elevations of RCEW. This lower functional richness may also be a reflection of environmental tolerance (beta niches) to disturbance such as grazing (Tilman & Downing, 1994).

In this case, the remotely sensed functional diversity can indicate how the niche is grouped in trait space. Though we observe similar morphological traits at lower elevation and diverse morphological traits at higher elevations, the local grouping or the patchiness of the vegetation is mainly governed by the soils and slopes. Within the same soil association, the slope changes the soil thickness within close proximities. Hence, the different densities of vegetation can be observed in the trait space leading to very dynamic divergence characteristics in the study area, in all ecotones. The variable functional diversity in the study area may resemble the maximum use of available resources.

Disturbance and time-since-disturbance show a significant influence on driving the functional diversity in the study area. Further, the functional diversity in the fire regions are highly dependent on the spatial neighborhood size as well as the size of fire. The small-sized fires (< 2 km^2) with few GEDI shots limited the use of GEDI to infer fire effects at coarser resolution (figure S3). Similar functional richness over disturbed and undisturbed areas from ASO data may be due to the standing dead or burned wood in the disturbed area. Integrating other remote sensing techniques that can differentiate woody materials from leafy vegetation could resolve this (e.g. Li et al., 2018). However, the increase of functional richness of both disturbed and undisturbed areas with the expansion of neighborhood size is due to the potential merging of different traits associations. The intermittent steep slopes in the functional richness curves could be the transition zones of functionally different niches. These intermittent steps could further indicate the relevant scales of spatial organization.

Functional divergence and functional evenness show a clear separation between disturbed and undisturbed areas. Our results emphasize that fire in this region can lead to a less even and more divergent ecosystem, and are compatible with the results of Chillo, Anand, & Ojeda (2011). The differences disappear at large spatial extents though, where functional evenness and divergence of burned and unburned areas converge. This convergence emphasizes the importance of using fine resolution remote sensing data such as ASO to identify fine-scale disturbance effects. At global scales, this could be potentially accomplished fusing GEDI data with ICESat-2 and Synthetic Aperture Radar (SAR) (Bae et al., 2019; Qi, Lee et al., 2019, Qi et al, 2019). In addition, a higher functional divergence is observed in recent disturbances and this divergence decreases with time since fire. However, even after 15 years, these disturbed areas have not functionally recovered. Disturbance results in a functionally different and disconnected landscape from the surrounding. The slow rate of recovery largely affects native habitat, especially for sage grouse (a threatened species), and potential changes in ecosystem processes. The ideal habitat for sage grouse requires connected mosaics of sagebrush shrublands which allow them safe migration, secure shelter and food resources (Stiver et al., 2015). In addition, disturbance-driven fragmentation strongly affects biodiversity and resource distributions (e.g. ecological functions and processes), especially along edges of the disturbance (Collinge, 1996). Recovery from disturbance in the face of global change represents a substantial challenge to agencies that manage these lands. Our study shows that assessing functional diversity can help identify areas for restoration or other management activities to consider for treatment.

Conclusion

Mapping functional diversity across space and time is pivotal to capturing and monitoring the stability and resiliency of ecosystems under changing climate conditions. In this study, we assessed the trends and patterns of functional diversity and their abiotic controls of a semi-arid ecosystem using small footprint and simulated large footprint (GEDI) waveform lidar data. Our results show that both the fine resolution airborne and coarse resolution GEDI can capture functional richness, evenness and divergence of the semi-arid ecosystem. We further show that functional diversity is mainly driven by the local elevation gradient, elevation breakpoint, slope, and soil type in this study area. Fire disturbance results in a more even and less divergent ecosystem that has implications for biodiversity and habitat quality. While we found a positive trend of post-fire functional recovery, the ecosystem has not fully recovered even after 15 years from disturbance. Overall, GEDI's capability to capture trends and patterns of functional diversity in this semi-arid ecosystem demonstrates its exciting potential to identify critical biophysical and ecological shifts in similar ecosystems. This information will help monitor the changes in carbon-cycle dynamics, habitats and biodiversity across the globe in semi-arid ecosystems.

CHAPTER FOUR: ESTIMATING ABOVEGROUND BIOMASS IN A SEMI-ARID ECOSYSTEM FROM LARGE FOOTPRINT LIDAR DATA: INSIGHTS FOR GEDI

Abstract

Aboveground biomass (AGB) is a crucial parameter for assessing carbon stocks and ecosystem health including biodiversity and habitat quality. Large footprint full waveform lidar data are widely used to estimate aboveground biomass at regional scales and is often restricted to forest ecosystems. The potential of large footprint lidar data to estimate biomass in semi-arid ecosystems is not yet being explored due to challenges of overlapping vegetation and ground signals. The objectives of this study are to isolate the vegetation-only contribution in the signals from large footprint lidar in order to estimate aboveground biomass, and to quantify the associated uncertainty in these estimates, for a mountainous, low-stature vegetation ecosystem. This study shows that relative height metrics (RH) corrected for ground-effects from simulated waveforms of the Global Ecosystem Dynamics Investigation (GEDI) system can be used to estimate the aboveground biomass of a semi-arid ecosystem. The linear model with waveform derived RH explained 42% of variance of airborne lidar derived reference biomass with RMSE of 351 kg/ha (16% RMSE). Our model estimated the aboveground biomass of 82% of the study area with less than 50% uncertainty. Uncertainties greater than 50% were observed in cultivated areas with vegetation heights lower than 0.5 m. Our study shows the potential of large footprint lidar data to estimate biomass of semi-arid ecosystems that cover a vast majority of the terrestrial landsurface. Biomass estimates from GEDI will

help constrain the higher uncertainties in inter-annual variability of global carbon dynamics caused by semi-arid ecosystems. In addition, this new capability to estimate AGB at global scales can provide early warnings as to whether a semi-arid ecosystem is a net carbon sink or source.

Introduction

Aboveground biomass (AGB) plays an important role in ecosystem processes such as carbon and water cycling, habitats and biodiversity (Urbazaev et al., 2018). Since biomass is closely related to both the size and spatial distribution of vegetation, accurate estimates of the vegetation structure are essential (Qi & Dubayah, 2016; Qi et al., 2019). Lidar remote sensing is a dominant technique for assessing the status and dynamics of AGB across spatial extents (Goetz and Dubayah, 2011). Lidar data are used to extract canopy height, canopy cover, crown volume and other vegetation structure metrics (LAI, FHD, succession etc.) with direct relationships with biomass (Drake et al., 2003; Dubayah et al., 2010; Duncanson et al., 2015; Hurtt et al., 2004). However, such studies are mostly limited to small areas where airborne lidar are available. To clearly understand the dynamics of carbon cycling, biodiversity, and habitat quality, global coverage of lidar data are essential (Milenković et al., 2017).

NASA’s Global Ecosystem Dynamics Investigation (GEDI) mission is now collecting large footprint (~ 25 m) waveform lidar signals across the globe from the International Space Station. The GEDI mission is expected to collect about 10 billion waveform lidar signals digitized at 15 cm intervals (1 ns) over its nominal 2-year mission period (Goetz & Dubayah, 2011). The waveform lidar signals provide additional capabilities to derive vegetation structure compared to discrete lidar as it digitizes the total amount of lidar return energy at high vertical resolution (~1 ns = 15 cm). These backscattered waveform signals are sensitive to the spatial distribution of the scattering component, bare ground and vegetation. The features derived from the waveform lidar signals are used to retrieve detailed vegetation structure and biomass. Among others, the relative height metrics (RH) calculated from the cumulative waveform energy are widely used in ecological applications. (Qi et al., 2019; Silva et al., 2018). For example, the height at 98th percentile of the cumulative energy profile (RH98) is used to estimate canopy height (Qi t al., 2019).

However, the use of large footprint lidar signals and their features for vegetation structure or biomass estimates are not developed in low stature vegetation such as in semi-arid ecosystems. These ecosystems cover approximately 40% of the global terrestrial land surface and are critical for global carbon dynamics (Ahlstrom et al., 2015; Poulter et al., 2015). However, extracting biomass signals from large footprint lidar in semi-arid ecosystems is challenging due to the vegetation heights (< 2 m) that are close to the range resolution of the lidar system (Zhuang & Mountrakis, 2014). This similarity of vegetation height and lidar range resolution tends to generate single backscattered pulses combining both the vegetation and ground responses. This ground-vegetation assimilation is further enhanced when the ground slope is greater than 10° (Dong, Tang, Min, Veroustraete, 2019) (Figure 4.1). The slope induced backscattered waveforms alter the vegetation height estimates, leading to higher uncertainties in biomass estimates (Silva et al., 2018). Several studies have investigated the influence of ground slope on the backscattered waveform to understand the entire signal that is convolved with the ground effects and the vegetation (Gardner 1992; Abshire, McGarry, Pacini, Blair, & Elman

1994; Carabajal and Harding 2001; Filin, 2003). Dong et al., 2019 showed that the deviation of the backscattered pulse due to the ground slope can be calculated using the tangent of the ground slope and the sensor's field of view. Further, by limiting the size of the laser footprint on the ground to 25-30 m, one can minimize the effects on ground slope in the backscattered waveforms. (Pang, Lefsky, Sun, & Ranson 2011). Pang, Lefsky, Sun, & Ranson (2011) further showed that the vegetation signal can be completely assimilated in the ground single when vegetation height is less than the size of the ground pulse deviated by sloped terrain. Though a separate vegetation signal cannot be retrieved in such cases, the upper part of the waveform is related to the canopy echo. If the ground slope is known, the slope effect in the waveform can be calculated and removed (Dong, Tang, Min, Veroustraete, 2019; Wang et al., 2019). As a result, improved vegetation metrics and biomass estimates can be achieved from large footprint lidar waveforms collected in mountainous terrains (Wang et al., 2019).

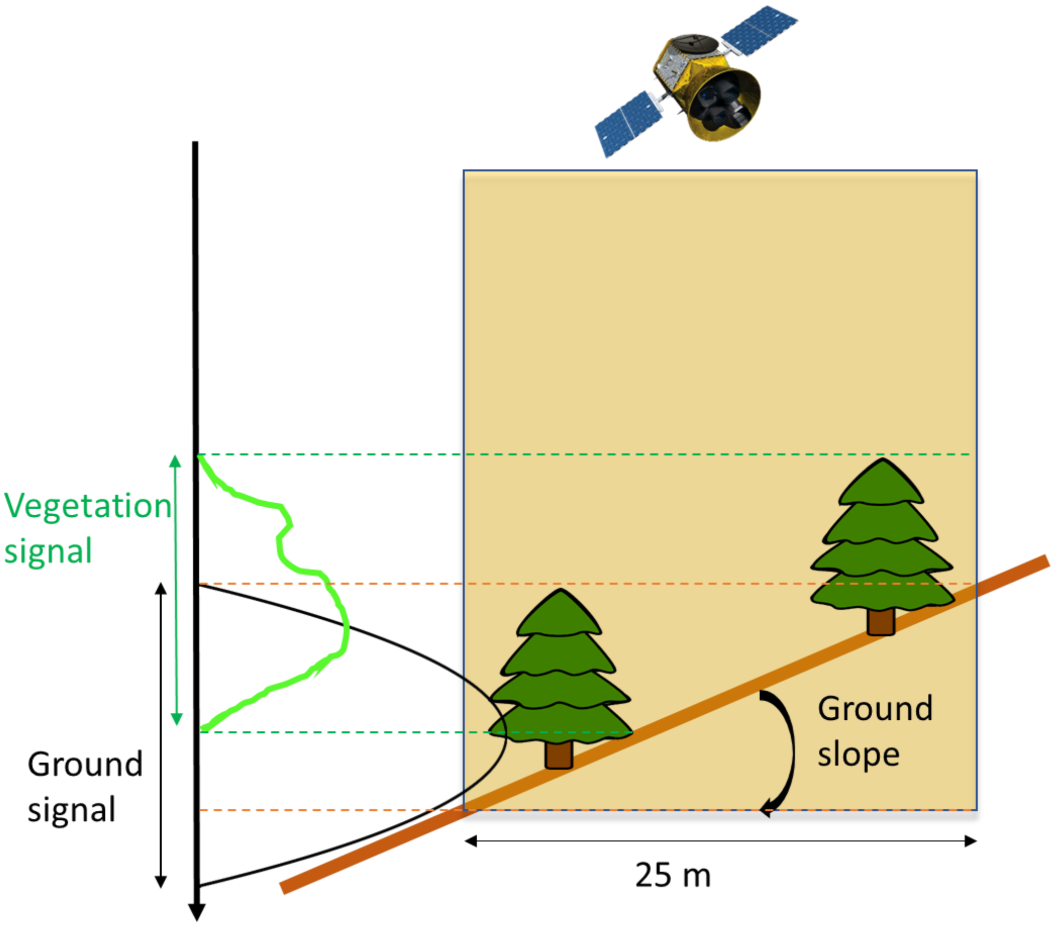


Figure. 4.1 **Assimilation of vegetation signal with ground signal in large footprint lidar data over sloped terrain.**

In semi-arid ecosystems, the vast majority of large footprint lidar signals are single pulse waveforms that include both the vegetation and ground (including slope) response. Thus, the aim of this study is to investigate the effectiveness of large footprint lidar to estimate aboveground biomass in a mountainous, semi-arid ecosystem. We first corrected the GEDI derived RHs using metrics derived from the reference ground waveforms. The reference ground waveforms were generated using the respective ground slopes. Then, we identified the GEDI RHs that are potentially incorporated with the vegetation response. A linear model was developed to estimate GEDI biomass using the selected RH metrics. Finally, an uncertainty analysis was performed to quantify the biomass uncertainty at both GEDI footprint and 1 km scales in a low stature, semi-arid ecosystem.

Methods

Study area

The study was carried out in the Reynolds Creek Experimental Watershed (RCEW). The study area covers approximately 270 km^2 within the Great Basin ecoregion in the Western US (Figure. 4.2). This region has a range of topography (1100 m – 2200 m) and vegetation communities. While many varieties of grass, forbs, and shrubs dominate the low elevations, trees of Aspen (*Populus tremuloides*), juniper (*Juniperus occidentalis*), and Douglas fir (*Pseudotsuga menziesii*) mark the high elevations. Further, low stature sagebrush (*Artemisia tridentata*), and bitterbrush (*Purshia tridentata*) of varying densities and cover are found throughout RCEW. In addition, riparian vegetation with cottonwood and willow are found within valleys, and along streams across the watershed. The study area is further characterized by a mean annual temperature and precipitation that vary between 4.6–9.2 ° C and 230-959 mm, respectively. This area has experienced prescribed and natural fires and supports grazing. As a consequence, invasion of cheatgrass in native shrub areas and juniper encroachment have occurred in this study area during the last few decades.

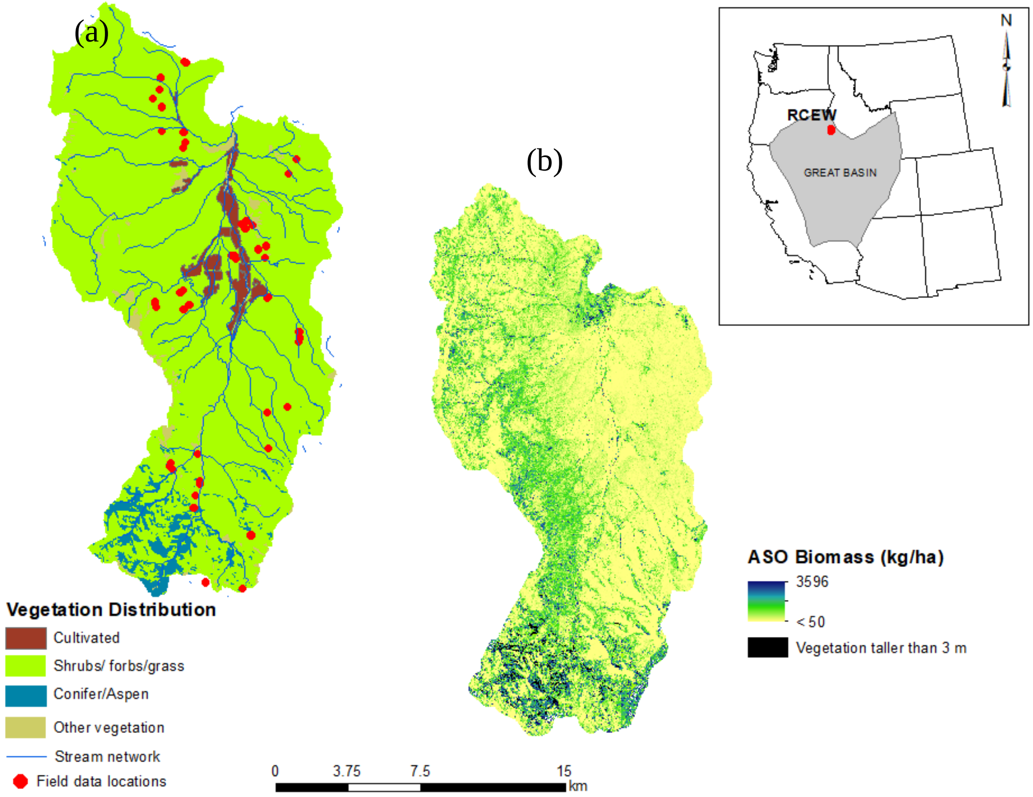

(a)
(b)
RCEW
GREAT BASIN
N
Vegetation Distribution
Cultivated
Shrubs/ forbs/grass
Conifer/Aspen
Other vegetation
Stream network
Field data locations
0
3.75
7.5
15
km
ASO Biomass (kg/ha)
3596
< 50
Vegetation taller than 3 m

Figure. 4.2 a) Reynolds Creek Experimental Watershed (RCEW) study area vegetation. b) ASO derived reference biomass distribution. The red dots in a) represents the spatial distribution of field plots.

Field data

Vegetation heights and cover data were collected at 43, 10 m x 10 m randomly selected plots over the study area during the summers of 2014 and 2015 (Figure 1). The plot size of 10 m x 10 m was assigned focusing on a separate study (Ilangakoon et al., 2018). The ground slopes within plots were between 1.5° to 13°. Within each plot, we established 5 transects at 1 m, 3 m, 5 m, 7 m, and 9 m. We collected canopy heights of all shrubs and took images at 2 m intervals along each transect totaling 20 measurements per plot using a Nikon COOLPIX AW120 camera. The canopy heights of all shrubs were then used to calculate the plot scale mean and max vegetation heights. The collected images were analyzed using the “Samplepoint” freeware program to estimate vegetation cover within each plot (v1.59, Booth et al., 2006).

Airborne lidar data (ASO)

Small-footprint waveform lidar data were acquired in August 2014 using the NASA Airborne Snow Observatory’s Riegl LMS-Q1560 (RIEGL Laser Measurement Systems GmbH, Horn, Austria) dual laser scanner (ASO hereafter). The mean above ground level of ASO was 1000 m (700 – 1300 m due to terrain conditions) and generated footprints of 20 – 60 cm. The study area was scanned at a pulse repetition rate of 400 kHz per laser and the backscattered signal was sampled at 1 ns per sample. All lidar waveforms in each flight line were Gaussian decomposed following the Ilangakoon et al., (2018) workflow. We derived spatial coordinates, incident angle, pulse width, amplitude, and scattering cross section of all echoes in each waveform. The resulting average point density across the watershed was 10-14 pts/m^2. The ASO derived point cloud was used to create the reference biomass data and to simulate GEDI waveforms. The use of the same ASO point cloud for the GEDI simulation helped to avoid system bias and reduce uncertainty. Further, this allowed us to assume that the GEDI system observed the same vegetation structure as that of ASO, enabling a direct comparison.

GEDI simulation

GEDI full waveform lidar signals were simulated using the GEDI simulator (Hancock et al., 2019). Discrete return ASO lidar point clouds were used to represent the true vertical distribution of the surface. We used a Riegl system derived point cloud (point density of 10-14 pts/m^2) for the simulation. We assumed our simulated GEDI waveforms closely represent actual GEDI waveforms based on a previous study where Riegl laser system with point densities > 3 beams/m^2 were used in a similar manner (Hancock et al., 2019). We used an assumption of ~50% cloud cover (based on our study area which has 46 – 63% mean cloudy days per annum) to derive the GEDI track patterns. The percent cloudy days of the study area was identified using the mean annual cloud cover dataset for the period of 2000-2014 (Wilson & Jetz (2016). A GEDI waveform was modelled as a Gaussian shaped laser pulse (FWHM = 15.6 ns) that has an effective footprint size of 22 m with a sampling rate of 1 ns (~ 15 cm). We added noise to the simulations to match the expected performance of the GEDI power beams. Simulated waveforms were smoothed by a Gaussian of FWHM of 11 ns (75% of GEDI system pulse) (Hofton et al., 2000). The mean noise was subtracted, and a threshold was set equal to 3.5 times the standard deviation. The ground was identified by Gaussian fitting the denoised waveforms (Hofton et al., 2000). The noise filtered waveforms were then used to derive the relative height metrics (RH metrics) at 2% intervals from the cumulative waveform energy profile (Hancock et al., 2019).

In addition, we generated another set of GEDI waveforms and RH metrics using point clouds clipped from field plots. To match the GEDI footprint size (25 m), we created 25 m footprints surrounding each 10 m x 10 m plot. Then the vegetation points beyond the 10 m x 10 m plot boundary were removed. This enabled us to directly compare the GEDI RH metrics with the field observed vegetation heights.

Reference biomass

We derived the reference biomass using ASO lidar point clouds because we did not have field biomass data over the GEDI footprints. We derived percent vegetation cover and maximum vegetation heights at 5 m spatial resolution from the ASO point clouds using the BCAL LiDAR Tools (Streutker and Glenn, 2006, https://github.com/bcal-lidar). Percent vegetation cover in this case is the percentage of above ground lidar points (aka vegetation points) to the total lidar points within each 5 m pixel space. We used a 25 cm crown threshold in the BCAL lidar tools to obtain the optimal vegetation cover for the study area from the airborne lidar data (Li et al., 2015). The percent vegetation cover was then used to calculate the biomass at 5 m spatial resolution (the middle image in figure 1b) for the study area using equation 1;

$$Biomass \ = -2391 + 797.6\ (PVC) + \epsilon \tag{1}$$

where biomass is measured in grams (g) and the PVC is the percent vegetation cover at 5 m (Li et al, 2015). This relationship was developed with a separate airborne lidar dataset with calibration and validation using field and terrestrial laser scanner data for the same study area (Li et al., 2015). In this study, the 5m biomass map was resampled to 25 m resolution and used as the reference biomass map (ASO biomass hereafter). We further resampled the ASO biomass into 1 km spatial resolution to compare with the 1 km GEDI biomass.

GEDI biomass model development

To remove the slope effect in GEDI waveforms, we simulated ground only waveforms considering the ground slope at each GEDI footprint (Wang et al., 2019). The ground slopes were extracted using the slope map derived from the ASO point cloud. The Gaussian pulse of flat ground (slope = 0°) can be modelled using the equation below;

$$y = A * \ exp^{\left(\frac{-x^2}{2\sigma^2}\right)} \tag{2}$$

where y is the Gaussian pulse and the A, x, and σ are Amplitude, time, and the standard deviation of the pulse respectively. The broadening of the pulse on a sloped terrain can be explained by;

$$\sigma_1 = \sigma + \ \beta * \gamma * \tan\theta \tag{3}$$

where γ is the footprint size (25 m in this case), σ1 is the standard deviation of the pulse on the sloped terrain, and θ is the terrain slope. β in the above equation is the coefficient for the conversion from half of the broadened waveform to the Gaussian shape factor (i.e., standard deviation). As in most cases, the laser pulse levels off at 2 times the standard deviation. Thus, we used 0.5 for the β. With the corrected pulse deviation, the waveform on sloped terrain can be modelled as;

$$y = A * \ exp^{\left(\frac{-x^2}{2(\sigma+\ \beta*\gamma*\tan\theta)^2}\right)} \tag{4}$$

We used the same signal start as GEDI waveforms for these ground only waveforms. We then calculated the RH metrics at 2% intervals from the ground only waveforms (GRH). The GRH metrics were subtracted from the RH metrics derived from simulated GEDI waveforms to obtain the slope corrected RH metrics. The slope corrected RH metrics were then used for the biomass model development as described below.

We first performed a correlation analysis between slope corrected RH metrics and field estimated vegetation heights. By doing this, we aimed to find the best RH metrics that can represent the true vegetation structure, and hence biomass. The RH metrics that showed higher correlation (> 20%) with field observed vegetation heights were used to find the best model for biomass estimates. We randomly selected 10% of the total simulated GEDI footprints across the study area (630 GEDI footprints). GEDI footprints with vegetation heights greater than 3 m were excluded from the biomass model calibration and validation in order to find the best model for low stature vegetation. The ASO reference biomass data at selected GEDI footprints were extracted using ArcMap. Forward and backward elimination was used to select the best slope corrected GEDI RH predictors for biomass estimates. We used “leapSeq” in the R package “leaps” as it uses 10-fold cross validation for each model run with both forward and backward elimination. The model performance was evaluated using the R^2, adjusted R^2, root mean square error (RMSE) and the bias. We limited our models to use a maximum of five predictors. A linear model was developed using the best predictors to estimate the mean biomass at each GEDI footprint (~25 m) of the study area. In addition, we estimated 2.5% lower and 97.5% upper boundaries of biomass within each GEDI footprint. The biomass at the 95% credible interval (CI95) was then calculated using equation 5 (Urbazaev et al., 2018). The percent uncertainty of biomass in each GEDI footprint was calculated using equation 6.

$$CI95 = (Biomass\ at\ 97.5\% - Biomass\ at\ 2.5\%)/2 \qquad (5)$$

$$Footprint\ Uncertainty = \frac{Biomass\ at\ 95\%\ credible\ interval}{Mean\ footprint\ biomass} * 100 \qquad (6)$$

The footprint level biomass data were resampled to 1 km spatial resolution to develop a seamless biomass map for the study area. We included all GEDI footprints to generate the seamless 1 km map for the study area because we have few footprints from low stature vegetation in high altitude areas. In addition, the uncertainty analysis detailed the impact of the inclusion of GEDI footprints with vegetation heights > 3m. The 1 km spatial resolution was selected to be compatible with GEDI's expected Level 4 biomass products (Coyle, Stysley, Poulios, Clarke, & Kay, 2015). The 1 km GEDI biomass product was directly compared with the 1 km ASO biomass product. The uncertainty of GEDI biomass at 1 km resolution was reported as the standard deviation. The processing workflow to estimate footprint scale and 1 km scale GEDI biomass using the simulated GEDI lidar data is shown in Figure. 4.3.

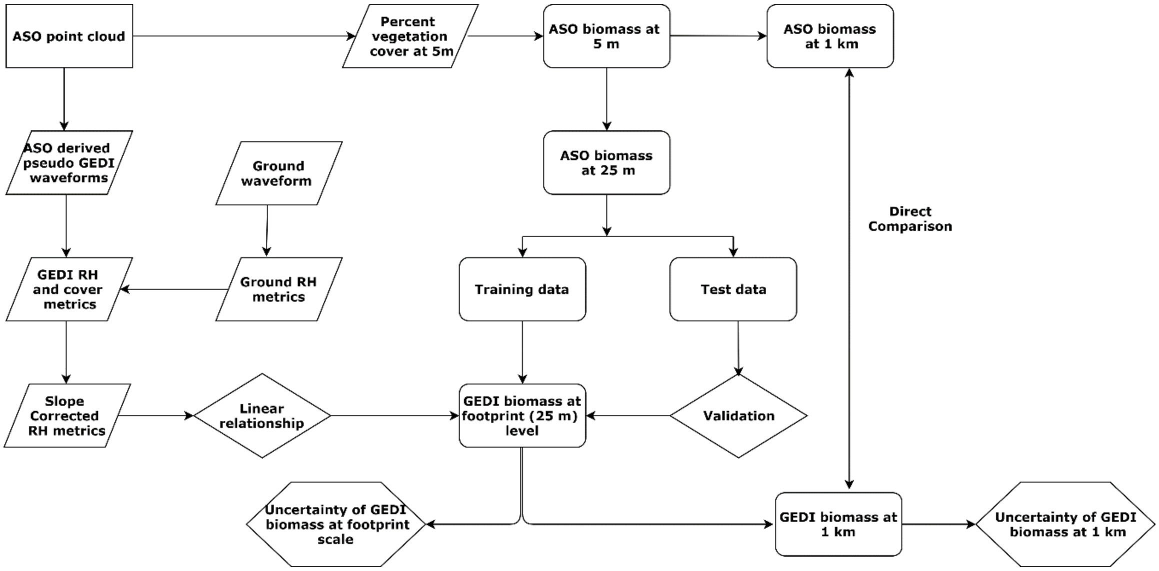

ASO point cloud
Percent vegetation cover at 5m
ASO biomass at 5 m
ASO biomass at 1 km
ASO derived pseudo GEDI waveforms
Ground waveform
ASO biomass at 25 m
Direct Comparison
GEDI RH and cover metrics
Ground RH metrics
Training data
Test data
Slope Corrected RH metrics
Linear relationship
GEDI biomass at footprint (25 m) level
Validation
Uncertainty of GEDI biomass at footprint scale
GEDI biomass at 1 km
Uncertainty of GEDI biomass at 1 km

Figure. 4.3 Processing workflow of GEDI biomass and biomass uncertainty estimates.

Results

Using the ASO point cloud data, we derived maximum vegetation height, percent vegetation cover and the reference biomass maps for the study area. The ASO percent cover and max vegetation height of the study area varied between 0.14% - 97% and 0.07 m – 29.2 m respectively. The southern portion of the study area depicts taller vegetation (> 3 m) and denser canopies (percent vegetation cover > 50 %) (Figure. 4.4). The rest of the study area is dominated by shorter vegetation (< 3 m) with a range of canopy cover (6 – 65%). Intermittent dense canopies with taller vegetation are observed throughout the watershed (blue area within yellow regions in Figure. 4.4). The ASO derived biomass of the study area varied between -96 kg/ha to 3022 kg/ha. The negative biomass occurred in cultivated areas where vegetation heights are lower than 0.5 m.

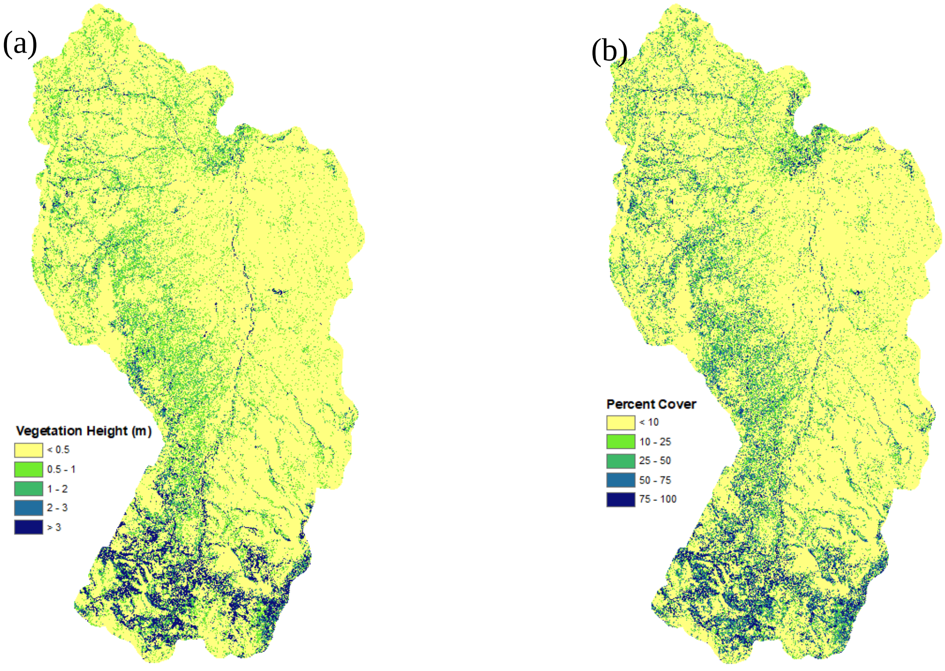


Figure. 4.4 ASO point cloud derived a) maximum vegetation heights and, b) percent vegetation cover maps at 5 m resolution.

All GEDI RH metrics between RH16 and RH68 showed correlations greater than 23% with field observed max vegetation heights (23% - 35%). A decreased correlation was observed below RH16 and above RH68 (Figure. 4.5). At RH16, the GEDI heights varied between -2.45 m to 1.48 m and at RH68, the GEDI heights varied between 0.07 m to 2.12 m. The maximum vegetation heights of our plots ranged between 0.3 m to 2.5 m. This showed that our field observed vegetation heights are within RH16 – RH68. The percent vegetation cover of the field plots varied between 49% - 78%.

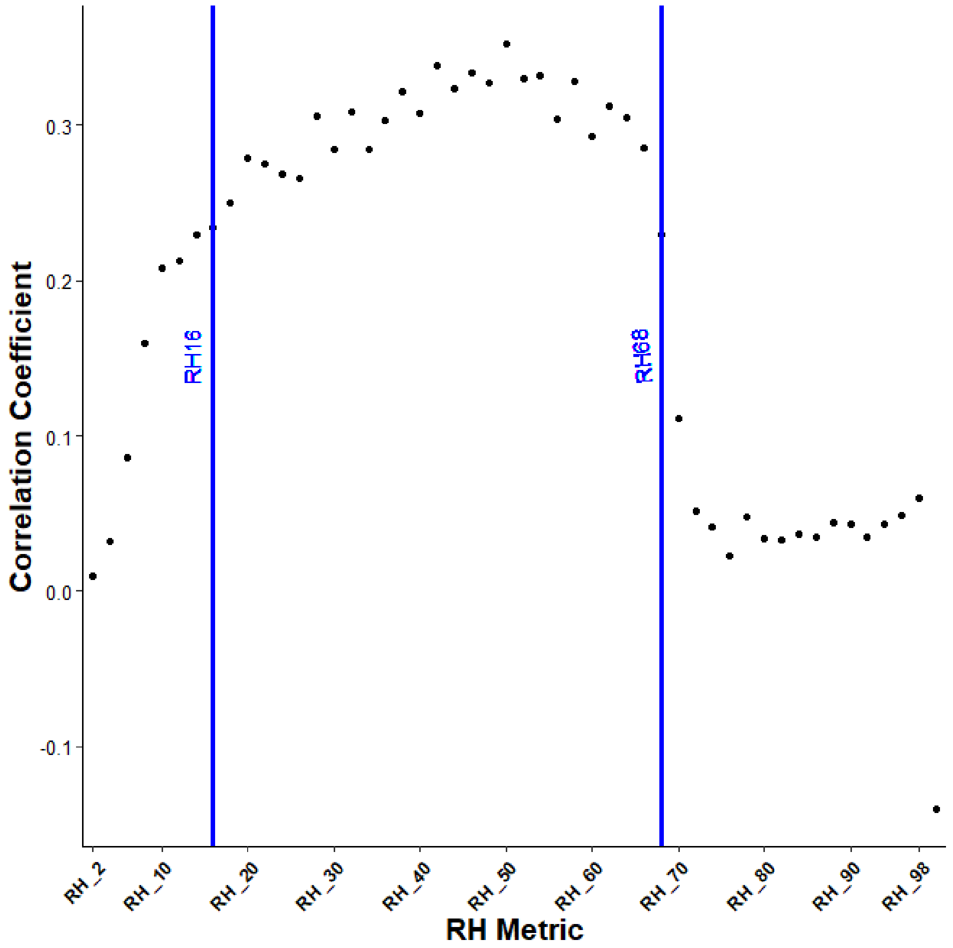


Figure. 4.5 Correlation coefficients between slope corrected GEDI RH metrics and field observed max vegetation heights.

The slope corrected RH16 and RH68 were the best predictors to estimate biomass from the GEDI data. The final model with best predictors is shown in the equation 7.

$$GEDI\ Footprint\ Biomass = -66 - 306\ (RH16) + 96\ (RH68) \tag{7}$$

Where the GEDI footprint biomass was measured in kg/ha. The RH16 and RH68 were the heights at 16th and 68th percentiles, respectively, of the cumulative waveform energy. The two predictor variables were significant at the 0.001 level with a p value of 2.2e-16. The model explained 42% (Adjusted $R^2 = 0.41$) variance of the biomass estimates with a mean RMSE of 351 kg/ha (mean bias = -2.06 kg/ha) (Figure. 4.6). This RMSE is approximately 16% of average shrub biomass (~13900 kg/ha) in Western US ecosystems (Martin et al., 1981; Li et al., 2017)

slope

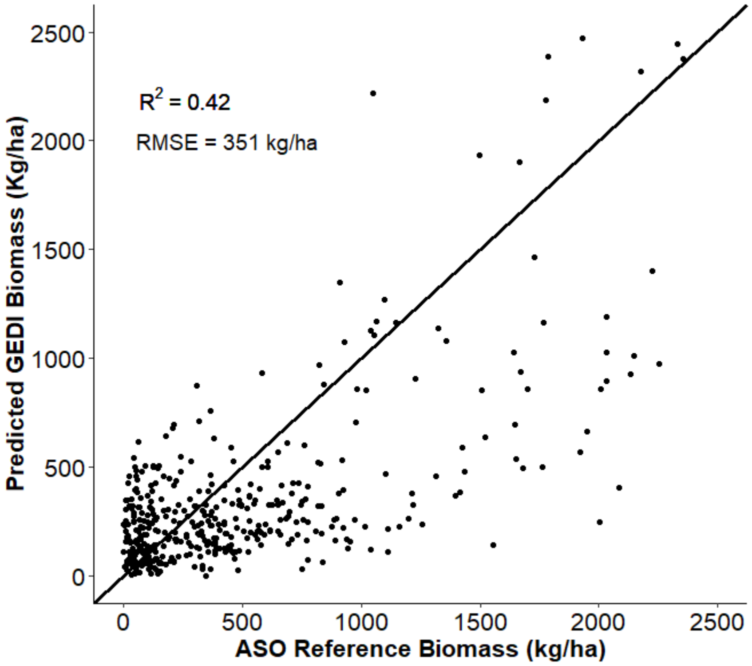


Figure. 4.6 Reference ASO biomass versus predicted GEDI footprint biomass (kg/ha) with the 1:1 line.

Figure. 4.7 shows the ASO biomass and predicted GEDI biomass at 25 m resolution. While the ASO biomass varied between -96 kg/ha to 3022 kg/ha, the GEDI biomass varied between -211 kg/ha to 3596 kg/ha. Both ASO and GEDI showed negative biomass in cultivated regions where vegetation heights are less than 0.5 m (refer Figure. 4.2 for cultivated regions). The GEDI footprint biomass at the 95% credible interval (CI95) showed a range between 24 kg/ha to 399 Kg/ha. The highest CI95 biomass was reported in tree dominant regions (yellow - orange areas, on Figure. 4.8a) while the lowest was reported in shrub dominant regions (blue areas, in the Figure. 4.8b). However, all the areas with vegetation heights greater than 1 m (tree and shrubs) showed uncertainties less than 10% (blue dots in Figure. 4.8c). This was approximately 26% of all footprints studied. From all the GEDI footprints studied, 64% showed uncertainty less than 25% and 82% showed uncertainty less than 50%. The areas with shorter vegetation (grass, shorts shrubs, and cultivated areas) showed the highest uncertainties, sometimes greater than 100%. These high uncertainty areas made up approximately 9% of all footprints. Our predicted footprint biomass showed a right skewedness with 87% of footprints with biomass less than 500 kg/ha (Figure. S.5). Excluding pixels with uncertainty greater than 100%, the uncertainty also showed a right skewedness (Figure. S.5).

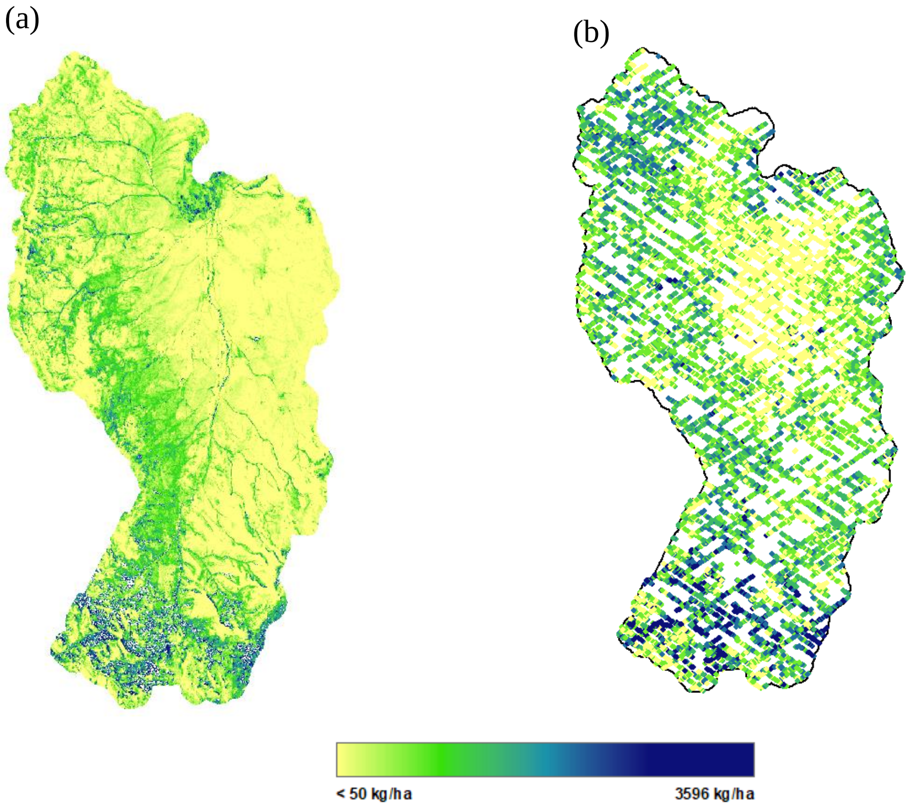


Figure. 4.7 a) ASO reference biomass with vegetation above 3m masked and b) predicted GEDI footprint biomass.

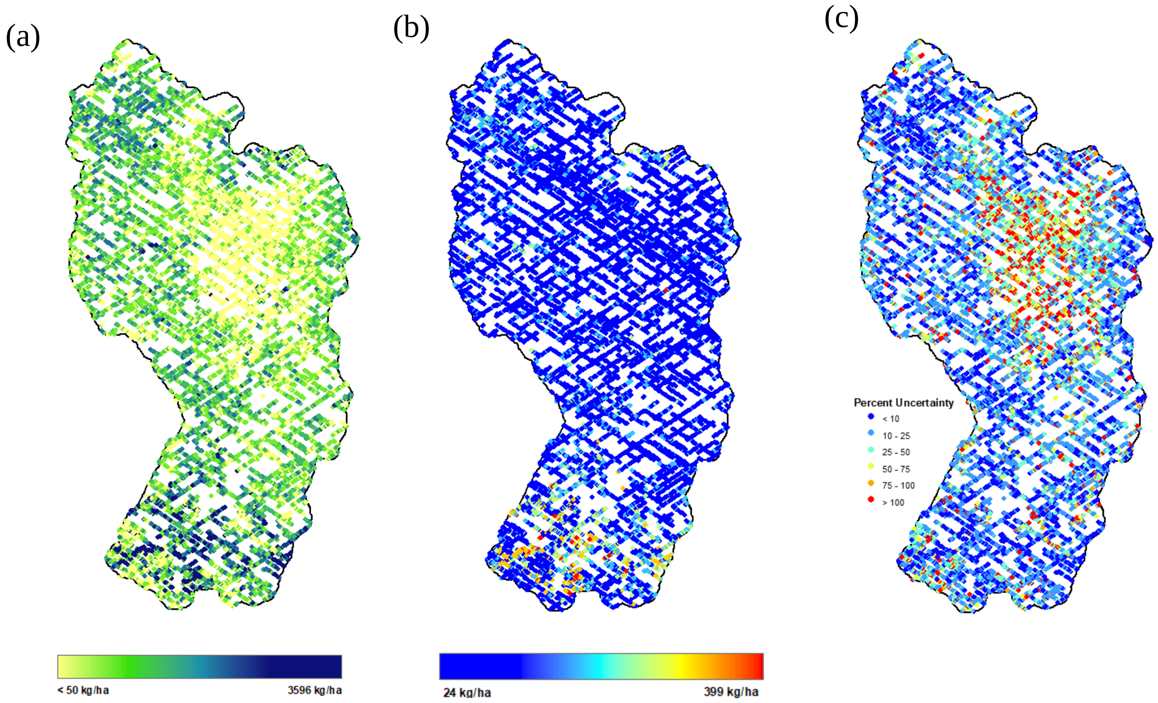


Figure. 4.8 a) GEDI footprint level biomass; and b) prediction uncertainty from the 95% credible interval; and c) percent uncertainty.

At 1 km scale, the ASO and GEDI biomass ranged from 7 kg/ha to 1500 kg/ha and 39 kg/ha to 1364 kg/ha respectively (Figure. 4.9). According to our results, GEDI underestimated biomass in tree dominant regions and overestimated biomass in the shrub dominant regions (Figure. 4.9b). The minimum differences were observed in areas with vegetation heights above 1 m and the canopy cover between 25 – 75 %. GEDI showed a higher underestimation of biomass compared to ASO in dense canopies. The largest difference of footprint biomass (3992 kg/ha) within a 1 km pixel was observed in tree-shrub co-dominant areas. (Figure. 4.10b). Relatively uniform biomass was observed in the shrub-dominant central and northern portions of the study area. The highest standard deviation at 1 m scale was reported in the southern portion of the study areas (1240 kg/ha) (Figure. 4.10c). The lowest reported standard deviation within a 1 km pixel was 40 kg/ha and was from a shrub dominant region.

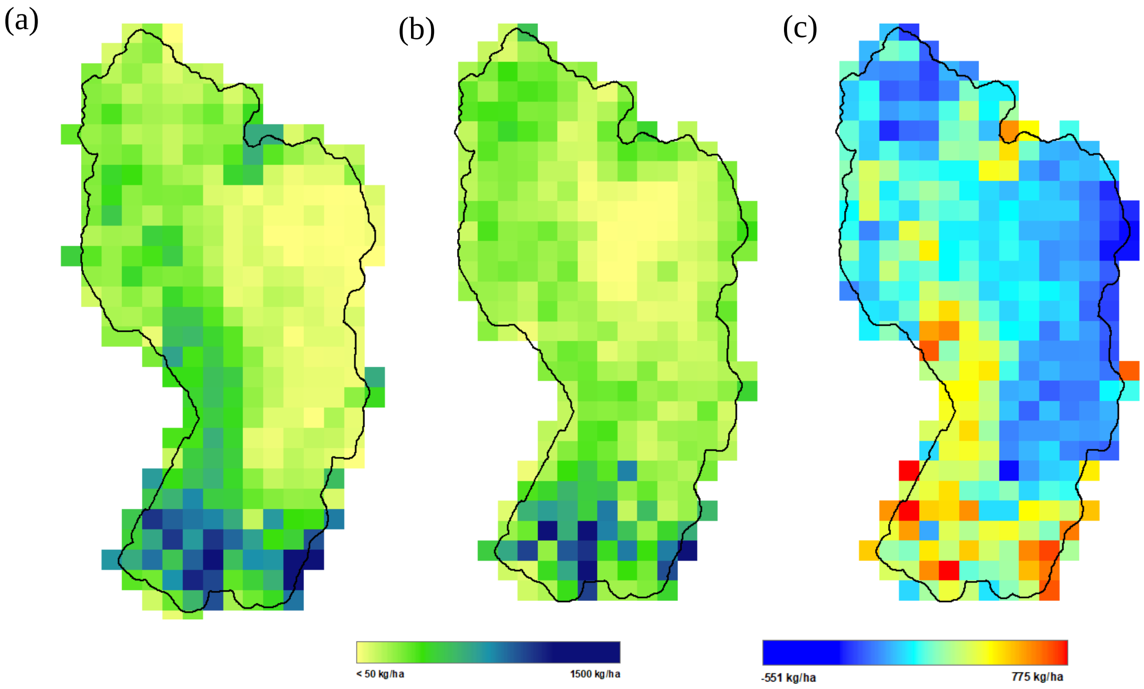

(a)
(b)
(c)
< 50 kg/ha
1500 kg/ha
-551 kg/ha
775 kg/ha

Figure. 4.9 ASO and GEDI biomass comparison including a) reference ASO biomass resampled to 1 km; b) predicted GEDI biomass resampled to 1 km; and c) biomass differences between reference and GEDI at 1 km.

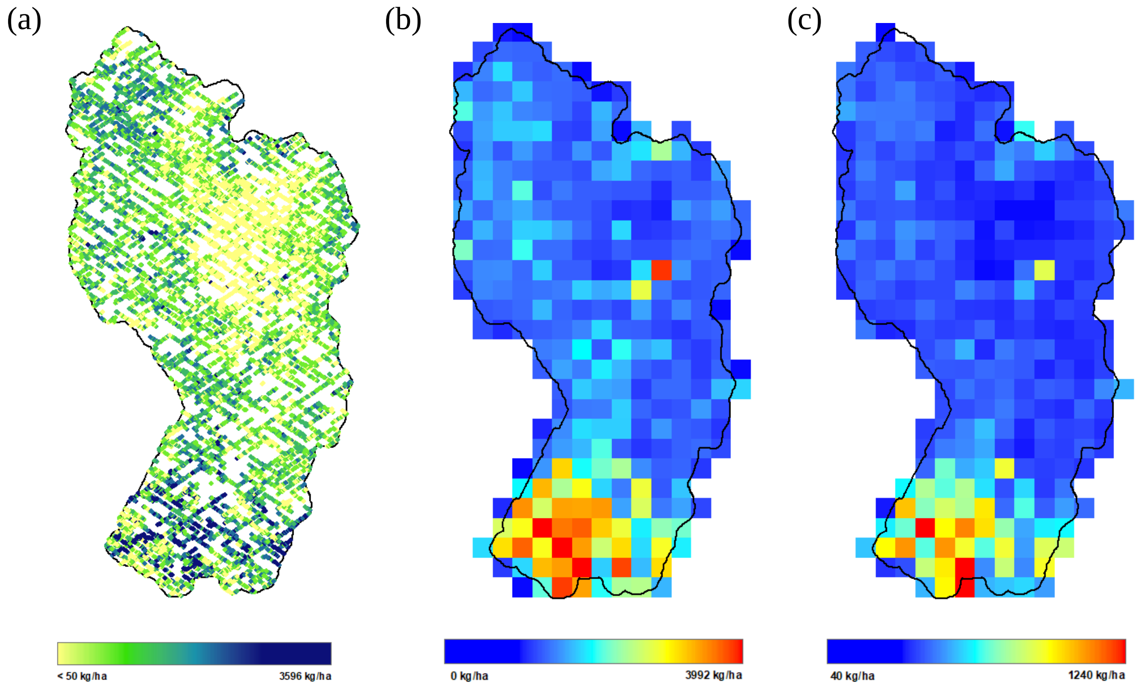


Figure. 4.10 Uncertainty of biomass with upscaling to 1 km including a) predicted GEDI footprint level biomass; b) range of biomass within each 1 km pixel and c) standard deviation of biomass at 1 km pixels derived from the 25 m footprint biomass.

Discussion

GEDI is expected to deliver 1 km biomass products globally using approximately 10 billion footprint lidar waveforms collected across the globe during its nominal two-year period. To generate biomass from the GEDI waveforms, a set of biomass – lidar models are being calibrated based on continental regions and plant functional types using a set of global training data (Duncanson et al., 2019). The calibrated models will be used to estimate biomass at the footprint scale, which will be later resampled to 1km spatial resolution. Dryland ecosystems (of which semi-arid is a type of) cover over 40% of the terrestrial land surface. Therefore, it is critically important to understand the potential opportunities and limitations of GEDI to estimate biomass in these ecosystems. This study provides a framework to estimate semi-arid ecosystem biomass from large footprint GEDI lidar data using the slope corrected waveform features (RH metrics). The slope correction followed by the correlation analysis with field measured heights helped isolate the RH metrics that may contain vegetation responses in semi-arid ecosystems. We used all amplitudes from the waveform signal start to signal end to derive the RH metrics. The RH metrics below the ground position showed negative values. The RH metrics above the ground position depicted positive values. If the bare ground position was located incorrectly, the RH metric that represents the canopy bottom can show a negative value. The both positive and negative values of our RH16 across the watershed showed its association with canopy bottom and sometimes with bare ground. The negative coefficient of RH16 in the final biomass model may be due to this ground and canopy confusion at lower RH metrics. The RH68 is closely associated with the maximum vegetation heights of the field plots and thus we can assume it represents the top of the canopy. We can also assume the selection of RH16 and RH68 reflects the height range where the vegetation response is distributed within the GEDI waveform (Figure. 4.11).

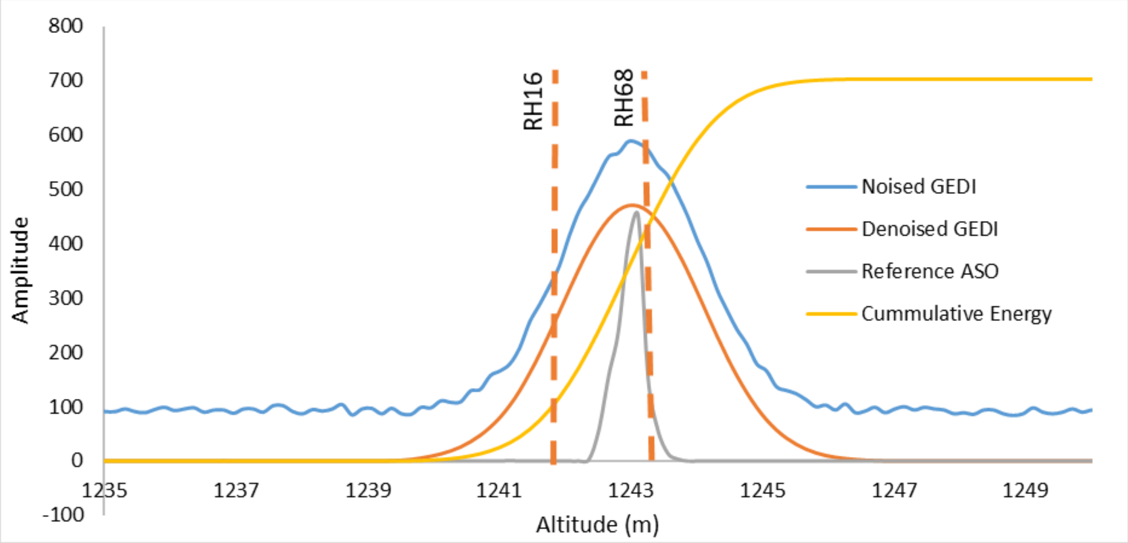


Figure. 4.11 A sample simulated GEDI waveform with the reference ASO waveform and the cumulative energy profile. The reference ASO waveform is contained within the GEDI RH16 and RH68.

In this study, we compared the simulated GEDI derived biomass with the airborne lidar derived biomass as we did not have adequate number of field biomass samples. Airborne lidar data has been widely used to accurately estimate the biomass in a range of ecosystems including this study area (Ferraz et al., 2016; Li, Glenn, Olsoy, Mitchell, & Shrestha, 2015). Airborne lidar derived biomass models have also been used to calibrate the spaceborne lidar derived biomass models (Narine et al., 2019). Our slope corrected GEDI biomass model explained only about 42 % of the ASO biomass. This low R^2 could be mainly due to uncertainties of the reference biomass (not calculated in this study). The reference biomass model was explicitly developed for shrubs and thus likely did not represent the tree biomass. Further, the model we used to generate the reference biomass map was developed using a different airborne lidar dataset (Li et al., 2015). The model developed by Li et al. (2015) explained 54% ($R^2 = 0.54$) of the variance of terrestrial laser scanner derived biomass. This inherent low R^2 and the differences in point densities, lidar sensor capabilities and characteristics between our ASO data and the data used in Li et al. (2015) may have biased our results. However, our modeled RMSE is about 16% of the average biomass of the dominant shrub type (*Artemisia tridentate* spp.) at 100% crown coverage (data from Martin (1981). According to a study performed in a similar study site, our RMSE is roughly 15% of total biomass (2100-2630 kg/ha) (Li et al., 2017). The low RMSE compared to other GEDI calibration studies (e.g., Armston et al., 2019) could be due to differences in vegetation cover, composition, and the year of the data collection between the different studies.

We assessed the variance of GEDI biomass with ground slope, vegetation heights and vegetation cover to confirm our slope corrected biomass estimates are not biased due to the ground slope. Our results revealed that 71% of the variance of the GEDI biomass can be explained by vegetation cover, height and ground slope. Among them, less than 17% was explained by the ground slope (figure S.6). Those three factors collectively explained 62% variance of CI95 biomass. From that, only a 0.012% was explained by the ground slope (Figure S.6). The variance analysis confirmed that the greatest influence on biomass estimates is from vegetation height and cover. Silva et al. (2018) showed that taller and denser vegetation can decrease lidar derived biomass uncertainty. However, in our study, the uncertainty decreased with vegetation heights up to 10 m and then increased (Figure. 4.11a). As expected, the percent uncertainty decreased with an increase in percent vegetation cover (Figure. 4.12b). Though the vegetation is short, if the vegetation cover is above 30%, biomass can be estimated with uncertainties as low as 25%. This is consistent with the findings of Glenn et al., (2016), that used Multiple Altimeter Beam Experimental Lidar (MABEL) instrument and Landsat 8's Operational Land Imager (OLI) to estimate the vegetation cover and biomass of this study area. The higher variance in higher vegetation cover may be due to the difficulty of the waveform signal to penetrate to the ground.

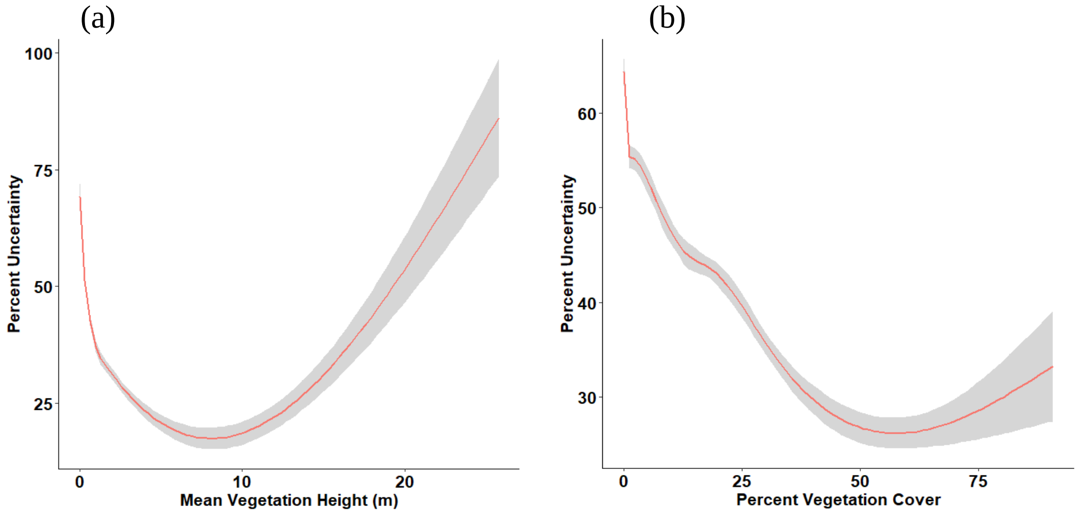


Figure. 4.12 Correlation of percent uncertainty with mean footprint vegetation height and percent vegetation cover.

According to our footprint scale uncertainty analysis, the highest uncertainty was observed in the cultivated regions with shorter vegetation (< 0.5 m). There are a few footprints scattered across the watershed with uncertainties higher than 100%. These footprints may be associated with grazing or grass dominated areas (e.g. short vegetation). Overall, our slope corrected GEDI metrics were able to estimate the biomass of more than 80% of the study area with less than 50% uncertainty. The overestimation of GEDI biomass at 1 km scales in shrub dominant regions can be attributed to the overestimation of vegetation heights from GEDI (Chapter 3). The underestimation of biomass in tree dominant sites could be due to multiple reasons. Our model selection focused on RH metrics that best describe the shrub structure (the RH metrics for trees may be different). The uncertainty in detecting canopy tops could also underestimate the tree biomass. In addition, the footprint distribution in the tree dominant region may not provide a representative sample of actual tree distribution. The highest standard deviation of biomass was reported in tree-shrub co-existing regions. This is mainly due to occurrence of the low biomass of shrub-dominant and high biomass of tree-dominant footprints within a 1 km space. Our footprint density investigation revealed no trends between footprint density and biomass standard deviation at the 1 km scale (Figure. 4.13). The highest number of footprints per 1 km observed was 110 and the lowest was 2.

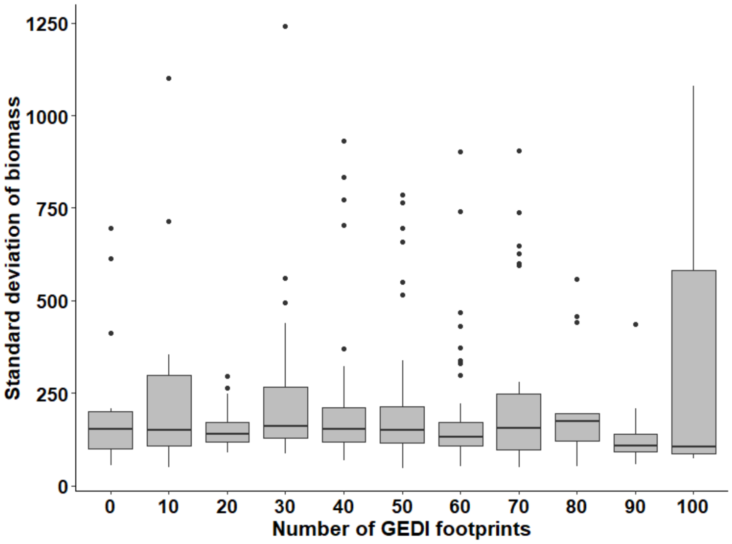


Figure. 4.13 GEDI footprint density and standard deviation of biomass at 1 km pixels.

Overall, our study highlights the exciting potential to use large footprint lidar to estimate biomass in a semi-arid ecosystem. Once the slope influence is corrected, the RH metrics from large waveform lidar can predict biomass of ecosystems with vegetation heights lower than 2 m. A future study with field-estimated biomass (rather than airborne lidar data) may improve our understanding of the errors associated with large footprint waveform lidar. Estimating biomass and associated uncertainties further helps to quantify fuel loads, post disturbance biomass dynamics and their spatial distributions across the landscape at scales relevant to fire management and land conservation. Further, biomass estimates across spatial extents provides better understanding of habitat quality.

Conclusion

The ability to estimate biomass in mountainous, semi-arid ecosystems using large footprint lidar data would significantly improve our capacity to monitor biomass and carbon dynamics at local to global scales. In order to do this accurately, future studies should focus on the thresholds in which GEDI performs well across a range of dryland ecosystems, and across spatial scales. We need to consider the GEDI footprint density and distribution, topography, and vegetation characteristics (cover and height), as these factors govern the accuracy of estimates. Ultimately, biomass datasets of field, terrestrial laser scanning, and airborne lidar data that provide calibration and validation of GEDI (or other spaceborne large footprint lidar data) are needed. These robust calibration and validation datasets across dryland ecosystems will provide uncertainty estimates that can be used at local scales as well as in global models of biomass and carbon dynamics.

CHAPTER FIVE: CONCLUSION

In chapter two, I developed a novel methodology to distinguish semi-arid shrubs from bare ground and co-dominant tree communities using full waveform lidar derived structural and biophysical variables. I found that structural variables are more important than the biophysical variables in distinguishing the vegetation from each other and from bare ground. Further, I found that the differentiation of PFTs are scale dependent. My main contribution in this study is finding a novel approach to distinguish semi-arid shrub vegetation from bare ground. Vegetation distributions in many semi-arid ecosystems are topographically controlled and low-height vegetation often coexists with taller tree communities. The topographic and species complexity in these ecosystems makes classification using optical remote sensing data challenging. Though high-resolution terrestrial laser scanning (TLS) can be used to distinguish shrubs and grass from bare ground, the use of TLS is limited to a maximum of a few hectares. Nonetheless, the adverse effects of climate driven changes on biodiversity, habitat quality and ecosystem productivity are discernable at regional to global scale measurements. The ability to differentiate vegetation from bare ground using airborne lidar waveforms is a significant improvement for ecosystems where vegetation height is lower than the height uncertainty in discrete return lidar (often +/- 15 cm). The lack of importance of the number of peaks for PFT classification demonstrates the capability of waveform features to differentiate PFTs even within a lower pulse density scenario. Overall, in this study I showed the exciting promise to constrain PFTs in heterogeneous semi-arid ecosystems, providing new opportunities for automated inventorying and monitoring, and estimating biomass and carbon from waveform lidar data. The results of this chapter are published in Remote Sensing of Environment (Ilangakoon et al., 2018) and are used as a basis for the remaining chapters of my dissertation.

In chapter three, I mapped functional diversity (functional richness, functional evenness, and functional divergence) using structural traits derived from airborne and spaceborne full waveform lidar. In this study, I showed that the spatial patterns of functional diversity can be explained by elevation gradient, slope and soil type. In particular, I found an elevation breakpoint (1500 m – 1700 m) in which diversity indices change drastically. I showed the potential of the newly launched GEDI to capture the trends and patterns of functional diversity in this semi-arid ecosystem. The disturbance-functional diversity relationships I developed showed that fires can lead this ecosystem into more even and less divergent states. According to my results, disturbed areas show a trend of post-fire functional recovery; however, they were not completely recovered even after 15 years since disturbance. My contribution from this chapter is multifold. Changes in vegetation structure and other ecosystem properties from disturbance can cause shifts in carbon pools and fluxes that may have far-reaching consequences. In addition, habitat quality and availability for local fauna are likely to be affected accordingly. Here I provided a detailed framework to map functional diversity in a spatially continuous way at both fine and coarse scales. Spatially continuous functional diversity maps can reveal several important biophysical and ecological shifts occurring in semi-arid ecosystems that help monitor the changes in carbon-cycle dynamics, habitats and biodiversity. In this chapter, I further discussed the key environmental factors that govern functional distributions. My methodology can be used to elucidate post-fire recovery for mitigation strategies and decision making.

In my fourth chapter, I estimated aboveground biomass from large footprint waveform data. I showed that the GEDI system has the potential to estimate AGB in this semi-arid ecosystem with less than 50% uncertainty over 80% of the area. This finding is remarkable given that our study area is dominated by low-height shrubs. I found it was necessary to remove slope effects in the GEDI lidar metrics. I showed that both vegetation height and percent vegetation cover influence the waveform shape and hence the waveform derived metrics. I found greatest uncertainties in upscaling in areas where trees and shrubs co-exist in close proximity whereas homogeneous vegetation, regardless of shrub or tree dominance, had the lowest uncertainties in AGB estimation. My main contribution in this study is, expanding the mission capabilities of large footprint lidar data to include semi-arid ecosystems. With near global coverage of GEDI, we have the potential to constrain the higher uncertainties in inter-annual variability of global carbon dynamics caused by semi-arid ecosystems. In addition, the capability to estimate AGB at global scales can provide early warnings whether the ecosystem is a net carbon sink or source.

My work advances the understanding of complexity and dynamics of semi-arid vegetation structure, function, and diversity. The methodologies and findings can be used to assess ecosystem services and processes. This information may be of interest to agencies such as United States Department of Agriculture (USDA) and Bureau of Land management (BLM) in the US and other agencies across the globe who have direct involvement in understanding and managing semi-arid ecosystems. For example, information about potential postfire diversity trends can be used to implement best management practices (e.g. artificial seeding) to protect habitats and maintain biodiversity. Further, detailed parametrization of key environmental factors (elevation, slope, and soil) and plant functional types in studies of semi-arid ecosystem demography and process modeling can help to constrain model uncertainty. In sum, a better understanding of ecosystem functions and processes can be obtained with the data and methods presented in this work.

The methodology I developed to distinguish short height vegetation from bare ground and other tree communities can be applicable to other similar ecosystems. Laser pulse deviation is a function of travel time (footprint size), system noise, atmospheric noise, and the emitted pulse width as well as the target characteristics. Hence, care must be taken to consider deployed system noise, emitting pulse width, and lidar footprint size and atmospheric conditions at the time of data collection before making ground-vegetation separation thresholds. In this study, the influence of precipitation and temperature were not considered as potential confounding factors due to the sparseness of those measurements. Further, the point measurements of precipitation and temperature showed a clear linear relationship with elevation (figure S.4) Hence, the use of elevation data incorporated the effects of precipitation and temperature.

There are a number of future studies that can be undertaken based on my findings. In this study, I calculated the functional diversity at a 500 m spatial neighborhood size. Though it provided consistent functional diversity measurements between airborne and GEDI systems, the lower number of GEDI footprints within small-sized fire boundaries (< 1 km^2) limited the use of GEDI to infer disturbance effects. Further, higher uncertainties of biomass estimates resulted in areas with diverse vegetation structure across fine scales. Future studies with fusion of GEDI with other data sources that also can provide accurate estimates of vegetation structure such as ICESat-2 and TanDEM-X may provide the capability to study small-sized disturbances and to constrain the uncertainties of AGB estimates at global scales. Building upon the methods and techniques introduced in this study, mission capabilities of NASA's future missions such as NISAR can be improved to estimate vegetation structure, biomass / carbon dynamics beyond the woody plants and forests. In this study, I evaluated the functional diversity distribution across the study area and explained the overall confounding environmental factors. Previous studies in the same and similar ecosystems have shown that aspect is one of the major abiotic factors that control lower elevation vegetation distributions, with diminished effects at higher elevations (McNamara et al., 2018; Poulos., 2016). Hence, future studies may consider elevation breakpoints to primarily divide the ecosystem and perform detailed functional diversity studies on low and high elevation sites separately. This would allow understanding of the impact of other factors that drive functional diversity and hence, ecosystem processes.

Peer-reviewed published manuscripts:

[1] Ilangakoon, N. T., Glenn, N. F., Dashti, H., Painter, T. H., Mikesell, T. D., Spaete, L. P., & Shannon, K. (2018). Constraining plant functional types in a semi-arid ecosystem with waveform lidar. Remote Sensing of Environment, 209, 497-509.

[2] H. Dashti, N. F. Glenn, S. Ustin, J. J. Mitchell, Y. Qi, N. T. Ilangakoon, A. N. Flores, J. L. Silván-Cárdenas, K. Zhao, L. P. Spaete, and M. de Graaff, “Empirical Methods for Remote Sensing of Nitrogen in Drylands May Lead to Unreliable Interpretation of Ecosystem Function,” IEEE Trans. Geosci. Remote Sens., pp. 1–12, 2019.

[3] H. Dashti, A. Poley, N. F. Glenn, N. Ilangakoon, L. Spaete, D. Roberts, J. Enterkine, A. N. Flores, S. L. Ustin`, and J. J. Mitchell, “Regional Scale Dryland Vegetation Classification with an Integrated Lidar-Hyperspectral Approach,” Remote Sensing , vol. 11, no. 18. 2019.

Manuscript in review:

[1] Ilangakoon, N. T., Glenn, N. F., Dashti, H., Schneider, F. D., Hancock, S., Spaete, L., & Goulden, T. Spaceborne lidar reveals trends and patterns of functional diversity in semi-arid ecosystems, In Review. 2020

[2] Ilangakoon, N, Glenn, N.F., Dashti, H., Li, Aihua., Hancock, S., Aboveground biomass in a semi-arid ecosystem from large footprint lidar data: insights for GEDI, In Prep. 2020

[3] Hamid Dashti, Karun Pandit, Nancy F. Glenn, Douglas J. Shinneman, Gerald N. Flerchinger, Andrew A. Hudak, Marie Anne de Graaf, Alejandro Flores, Susan Ustin, Nayani Ilangakoon, Aaron W. Fellows, “Performance of the Ecosystem Demography model (EDv2.2) in simulating photosynthesis capacity and activity along an elevation gradient in a dryland study area” In Review, 2019.

Book chapter:

[1] J. J. Mitchell, N. F. Glenn, K. M. Dahlin, N. T. Ilangakoon, H. Dashti, and M. C. Maloney, "Integrating Hyperspectral and LiDAR Data in the Study of Vegetation.," in Hyperspectral Remote Sensing of Vegetation (Volume I), II Ed., P. S. Thenkabail, J. G. Lyon, and A. Huete, Eds. London, Ney York: CRC Press-Taylor and Francis group, 2018, p. 449.

Datasets:

[1] Ilangakoon, N., N. Glenn, and L. Spaete. 2017. LiDAR Data, DEM, and Maximum Vegetation Height Product from Southern Idaho, 2014. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1532

[2] N. F. Glenn, L. P. Spaete, R. Shrestha, A. Li, N. Ilangakoon, J. MitchelL, U. S. L, Y. Qi, H. Dashti, and K. Finan, "Shrubland Species Cover, Biometric, Carbon and Nitrogen Data, Southern Idaho, 2014." ORNL Distributed Active Archive Center, 2017.

[3] H. Dashti, N. F. Glenn, L. P. Spaete, and N. Ilangakoon, "Hyperspectral Imagery from AVIRIS-NG for Sites in ID and CA, USA, 2014 and 2015." ORNL Distributed Active Archive Center, 2018.

REFERENCES

Abdalati, W., Zwally, H. J., Bindschadler, R., Csatho, B., Farrell, S. L., Fricker, H. A., Webb, C. (2010). The ICESat-2 Laser Altimetry Mission. Proceedings of the IEEE, 98(5), 735–751. https://doi.org/10.1109/JPROC.2009.2034765

Abed, F. M., Mills, J. P., & Miller, P. E. (2012). Echo Amplitude Normalization of Full-Waveform Airborne Laser Scanning Data Based on Robust Incidence Angle Estimation. IEEE Transactions on Geoscience and Remote Sensing, 50(7), 2910–2918. https://doi.org/10.1109/TGRS.2011.2175232

Abelleira Martínez, O.J., Fremier, A.K., Günter, S., Ramos Bendaña, Z., Vierling, L., Galbraith, S.M., Bosque- Pérez, N.A. and Ordoñez, J.C. (2016), Scaling up functional traits for ecosystem services with remote sensing: concepts and methods. Ecol Evol, 6: 4359-4371. doi:10.1002/ece3.2201

Abshire, J. B., McGarry, J. F., Pacini, L. K., Blair, J. B., & Elman, G. C. (1994). Laser altimetry simulator. Version 3.0: User's guide.

Ahlstrom, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Zeng, N. (2015). The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science, 348(6237), 895–899. https://doi.org/10.1126/science.aaa1668

Alexander, C., Deák, B., Kania, A., Mücke, W., & Heilmeier, H. (2015). Classification of vegetation in an open landscape using full-waveform airborne laser scanner data. International Journal of Applied Earth Observation and Geoinformation, 41, 76–87. https://doi.org/10.1016/j.jag.2015.04.014

Armston, J.D., Tang, H., Silva, C.E., Kellner, J.R., Duncanson, L., Hofton, M.A., Hancock, S., Blair, J.B. and Dubayah, R., 2019. Post-launch Evaluation of GEDI Footprint-Level Cover and Biomass Algorithms. AGUFM, 2019, pp.B31A-07.

Arnell, N. W., Lowe, J. A., Challinor, A. J., & Osborn, T. J. (2019). Global and regional impacts of climate change at different levels of global temperature increase. Climatic Change, 155(3), 377-391.

Bae, S., Levick, S. R., Heidrich, L., Magdon, P., Leutner, B. F., Wöllauer, S., … Müller, J. (2019). Radar vision in the mapping of forest biodiversity from space. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-12737-x

Bagaram, M. B., Giuliarelli, D., Chirici, G., Giannetti, F., & Barbati, A. (2018). UAV Remote Sensing for Biodiversity Monitoring: Are Forest Canopy Gaps Good Covariates? Remote Sensing, 10(9). https://doi.org/10.3390/rs10091397

Bardgett, R. D., & van der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. Nature, 515(7528), 505–511. https://doi.org/10.1038/nature13855

Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32.

Booth, D. T., Cox, S. E., & Berryman, R. D. (2006). Point sampling digital imagery with 'SamplePoint'. Environmental Monitoring and Assessment, 123(1-3), 97-108

Braun, D., Damm, A., Hein, L., Petchey, O. L., & Schaepman, M. E. (2018). Spatio-temporal trends and trade-offs in ecosystem services: An Earth observation based assessment for Switzerland between 2004 and 2014. *Ecological Indicators*, *89*(September 2017), 828–839. https://doi.org/10.1016/j.ecolind.2017.10.016

Braghiere, R. K., Quaife, T., Black, E., He, L., & Chen, J. M. (2019). Underestimation of Global Photosynthesis in Earth System Models Due to Representation of Vegetation Structure. *Global Biogeochemical Cycles*, 2018GB006135. https://doi.org/10.1029/2018GB006135

Bruggisser, M., Roncat, A., Schaepman, M. E., & Morsdorf, F. (2017). Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification. Remote Sensing of Environment, 196, 28–41. https://doi.org/10.1016/j.rse.2017.04.025

Bruniquel-Pinel, V., & Gastellu-Etchegorry, J. P. (1998). Sensitivity of Texture of High Resolution Images of Forest to Biophysical and Acquisition Parameters.

Remote Sensing of Environment, 65(1), 61–85. https://doi.org/10.1016/S0034-4257(98)00009-1

Canfield, R.H. 1941. Application of the line interception method in sampling range vegetation. J. Forestry 39:388-394.

Cao, L., Coops, N. C., Innes, J. L., Dai, J., Ruan, H., & She, G. (2016). Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data. International Journal of Applied Earth Observation and Geoinformation, 49, 39–51. https://doi.org/10.1016/j.jag.2016.01.007.

Carabajal, C. C., & Harding, D. J. (2001). Evaluation of geoscience laser altimeter system (GLAS) waveforms for vegetated landscapes using airborne laser altimeter scanning data. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34(3/W4), 125-130.

Chen, W., Li, X., Wang, Y., Chen, G., & Liu, S. (2014). Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. Remote Sensing of Environment, 152, 291–301. https://doi.org/10.1016/j.rse.2014.07.004

Chen, J., Luo, Y., Xia, J., Wilcox, K. R., Cao, J., Zhou, X., Jiang, L., Niu, S., Estera, K.Y., Huang, R. and Wu, F., (2017). Warming Effects on Ecosystem Carbon Fluxes Are Modulated by Plant Functional Types. Ecosystems, 20(3), 515-526. https://link.springer.com/article/10.1007/s10021- 016-0035-6.

Chillo, V., Anand, M., & Ojeda, R. A. (2011). Assessing the use of functional diversity as a measure of ecological resilience in arid rangelands. Ecosystems, 14(7), 1168-1177.

Collinge, S. K. (1996). Ecological consequences of habitat fragmentation: implications for landscape architecture and planning. Landscape and urban planning, 36(1), 59-77.

Conti, G., & Díaz, S. (2013). Plant functional diversity and carbon storage - an empirical test in semi-arid forest ecosystems. Journal of Ecology, 101(1), 18–28. https://doi.org/10.1111/1365-2745.12012

Coomes, D. A., Dalponte, M., Jucker, T., Asner, G. P., Banin, L. F., Burslem, D. F. R. P., Qie, L. (2017). Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. Remote Sensing of Environment, 194, 77–88. https://doi.org/10.1016/j.rse.2017.03.017

Coyle, D. B., Stysley, P. R., Poulios, D., Clarke, G. B., & Kay, R. B. (2015, September). Laser transmitter development for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar. In Lidar Remote Sensing for Environmental Monitoring XV (Vol. 9612, p. 961208). International Society for Optics and Photonics.

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.12575

Dashti Hamid, Karun Pandit, Nancy F. Glenn, Douglas J. Shinneman, Gerald N. Flerchinger, Andrew A. Hudak, Marie Anne de Graaf, Alejandro Flores, Susan Ustin, Nayani Ilangakoon, Aaron W. Fellows, “Performance of the Ecosystem Demography model (EDv2.2) in simulating photosynthesis capacity and activity along an elevation gradient in a dryland study area” In Review, 2019

Diaz-Uriarte, R., & de Andrés, S. A. (2005). Variable selection from random forests: application to gene expression data. arXiv preprint q-bio/0503025.

Díaz, S., Hodgson, J. G., Thompson, K., Cabido, M., Cornelissen, J. H., Jalili, A., & Band, S. R. (2004). The plant traits that drive ecosystems: evidence from three continents. Journal of vegetation science, 15(3), 295-304.

Diaz, S., Lavorel, S., de Bello, F., Quetier, F., Grigulis, K., & Robson, T. M. (2007). Incorporating plant functional diversity effects in ecosystem service assessments. Proceedings of the National Academy of Sciences, 104(52), 20684–20689. https://doi.org/10.1073/pnas.0704716104

Dietze M.C., A. Fox, J.Betancourt, M.Hooten, C.Jarnevich, T. Keitt, M. A. Kenney, C. Laney, L. Larsen, H.W. Loescher, C.Lunch, B.Pijanowski, J.T. Randerson, E.Read, A.Tredennick, K.C. Weathers, E. P. White. 2017. Iterative ecological forecasting: Needs, opportunities, and challenges. NEON Workshop Report. DOI: 10.6084/m9.figshare.4715317

D'Odorico, P., Bhattachan, A., Davis, K. F., Ravi, S., & Runyan, C. W. (2013). Global desertification: drivers and feedbacks. Advances in water resources, 51, 326-344.

Dong, L., Tang, S., Min, M., & Veroustraete, F. (2019). Estimation of Forest Canopy Height in Hilly Areas Using Lidar Waveform Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(5), 1559–1571. https://doi.org/10.1109/JSTARS.2019.2908682

Drake, J. B., Dubayah, R. O., Clark, D. B., Knox, R. G., Blair, J. B., Hofton, M. A., & Prince, S. (2002). Estimation of tropical forest structural characteristics using large-footprint lidar. Remote Sensing of Environment, 79(2-3), 305-319.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., & Hofton, M. (2003). Above- ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. Global ecology and biogeography, 12(2), 147-159.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. Journal of Geophysical Research: Biogeosciences, 115(G2).

Duncanson, L. I., Dubayah, R. O., Cook, B. D., Rosette, J., & Parker, G. (2015). The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation. Remote Sensing of Environment, 168, 102-112.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Carter, S., Chave, J., Herold, M., Crowther, T.W. and Falkowski, M., 2019. The importance of consistent global forest aboveground biomass product validation. Surveys in geophysics, 40(4), pp.979-999.

Durán, S. M., Martin, R. E., Díaz, S., Maitner, B. S., Malhi, Y., Salinas, N., … Enquist, B. J. (2019). Informing trait-based ecology by assessing remotely sensed functional diversity across a broad tropical temperature gradient. *Science Advances*, *5*(12), eaaw8114. https://doi.org/10.1126/sciadv.aaw8114

Endres, A. (2016). Relative importance analysis of Landsat, waveform LIDAR and PALSAR inputs for deciduous biomass estimation. European Journal of Remote Sensing, 795. https://doi.org/10.5721/EuJRS20164942

Ene, L. T., Næsset, E., Gobakken, T., Bollandsås, O. M., Mauya, E. W., & Zahabu, E. (2017). Large-scale estimation of change in aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. Remote Sensing of Environment, 188, 106–117. https://doi.org/10.1016/j.rse.2016.10.046

Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Gonçalves, G., Silva, C. A., ... & Pereira, L. (2016). Airborne lidar estimation of aboveground forest biomass in the absence of field inventory. Remote Sensing, 8(8), 653.

Fieber, K. D., Davenport, I. J., Ferryman, J. M., Gurney, R. J., Walker, J. P., & Hacker, J. M. (2013). Analysis of full-waveform LiDAR data for classification of an orange orchard scene. ISPRS Journal of Photogrammetry and Remote Sensing, 82, 63–82. https://doi.org/10.1016/j.isprsjprs.2013.05.002

Fisher, R. A., Koven, C. D., Anderegg, W. R., Christoffersen, B. O., Dietze, M. C., Farrior, C. E., ... & Lichstein, J. W. (2018). Vegetation demographics in Earth System Models: A review of progress and priorities. Global change biology, 24(1), 35-54. onlinelibrary.wiley.com/doi/10.1111/gcb.13910/full

Filin, S. (2003). Recovery of systematic biases in laser altimetry data using natural surfaces. Photogrammetric Engineering & Remote Sensing, 69(11), 1235-1242.

Frandsen, W. H. (1983). Modeling Big Sagebrush as a Fuel. Journal of Range Management, 36(5), 596. https://doi.org/10.2307/3898349

Funk, J. L., Larson, J. E., Ames, G. M., Butterfield, B. J., Cavender-Bares, J., Firn, J., Wright, J. (2017). Revisiting the Holy Grail: using plant functional traits to understand ecological processes: Plant functional traits. Biological Reviews, 92(2), 1156–1173. https://doi.org/10.1111/brv.12275

Fusco, E. J., Finn, J. T., Abatzoglou, J. T., Balch, J. K., Dadashi, S., & Bradley, B. A. (2019). Detection rates and biases of fire observations from MODIS and agency reports in the conterminous United States. Remote Sensing of Environment, 220, 30-40.

Gardner, C. S. (1992). Ranging performance of satellite laser altimeters.

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Cajgfinger, T., Gregoire, T., Grau, E., Ristorcelli, T. (2015). Discrete Anisotropic Radiative Transfer (DART 5) for Modeling Airborne and Satellite Spectroradiometer and LIDAR Acquisitions of Natural and Urban Landscapes. Remote Sensing, 7(2), 1667–1701. https://doi.org/10.3390/rs70201667

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Grau, E., Rubio, J., Cook, B. D., … Sun, G. (2016). Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. Remote Sensing of Environment, 184, 418–435. https://doi.org/10.1016/j.rse.2016.07.010

Glenn, N. F., Spaete, L. P., Sankey, T. T., Derryberry, D. R., Hardegree, S. P., & Mitchell, J. J. (2011). Errors in LiDAR-derived shrub height and crown area on sloped terrain. Journal of Arid Environments, 75(4), 377–382. https://doi.org/10.1016/j.jaridenv.2010.11.005

Glenn, N.F., L.P. Spaete, R. Shrestha, A. Li, N. Ilangakoon, J. Mitchell, S.L. Ustin, Y. Qi, H. Dashti, and K. Finan. 2017. Shrubland Species Cover, Biometric, Carbon and Nitrogen Data, Southern Idaho, 2014. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1503

Glenn, N.F., Neuenschwander, A., Vierling, L.A., Spaete, L., Li, A., Shinneman, D.J., Pilliod, D.S., Arkle, R.S. and McIlroy, S.K., 2016. Landsat 8 and ICESat-2: Performance and potential synergies for quantifying dryland ecosystem vegetation cover and biomass. Remote Sensing of Environment, 185, pp.233-242.

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Management, 2(3), 231–244. https://doi.org/10.4155/cmt.11.18

Gough, C. M., Curtis, P. S., Hardiman, B. S., Scheuermann, C. M., & Bond- Lamberty, B. (2016). Disturbance, complexity, and succession of net ecosystem production in North America's temperate deciduous forests. Ecosphere, 7(6), e01375.

Hancock, S., Armston, J., Li, Z., Gaulton, R., Lewis, P., Disney, M., Gaston, K. J. (2015). Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy. Remote Sensing of Environment, 164, 208–224. https://doi.org/10.1016/j.rse.2015.04.013

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L. I., Dubayah, R. (2019). The GEDI simulator: A large- footprint waveform lidar simulator for calibration and validation of spaceborne missions. Earth and Space Science. https://doi.org/10.1029/2018EA000506

Hancock, S., Disney, M., Muller, J.-P., Lewis, P., & Foster, M. (2011). A threshold insensitive method for locating the forest canopy top with waveform lidar. Remote Sensing of Environment, 115(12), 3286–3297. https://doi.org/10.1016/j.rse.2011.07.012

Hartley, A. J., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant functional type distributions and its impact on land surface models. Remote Sensing of Environment. https://doi.org/10.1016/j.rse.2017.07.037

Hassebo, Y. (2012). Active remote sensing: LiDAR SNR improvements. In Remote Sensing-Advanced Techniques and Platforms. InTech.

Heinzel, J., & Koch, B. (2011). Exploring full-waveform LiDAR parameters for tree species classification. International Journal of Applied Earth Observation and Geoinformation, 13(1), 152–160. https://doi.org/10.1016/j.jag.2010.09.010

Höfle, B., & Pfeifer, N. (2007). Correction of laser scanning intensity data: Data and model-driven approaches. ISPRS Journal of Photogrammetry and Remote Sensing, 62(6), 415–433. https://doi.org/10.1016/j.isprsjprs.2007.05.008

Hofton, M. A., Minster, J. B., & Blair, J. B. (2000). Decomposition of laser altimeter waveforms. IEEE Transactions on geoscience and remote sensing, 38(4), 1989-1996.

Holm, S., Nelson, R., & Ståhl, G. (2017). Hybrid three-phase estimators for large-area forest inventory using ground plots, airborne lidar, and space lidar. Remote Sensing of Environment, 197, 85–97. ttps://doi.org/10.1016/j.rse.2017.04.004

Hollaus, M., Aubrecht, C., Höfle, B., Steinnocher, K., & Wagner, W. (2011). Roughness Mapping on Various Vertical Scales Based on Full-Waveform Airborne Laser Scanning Data. Remote Sensing, 3(12), 503–523. https://doi.org/10.3390/rs3030503

Hooper, D., Buchmann, N., Degrange, V., Diaz, S., Gessner, M., Grime, P., … Spehn, E. (2006). Species diversity, functional diversity and ecosystem functioning. Biodiversity and Ecosystems Functioning: A Current Synthesis, 195-208 (2002).

Hopkinson, C., Chasmer, L. E., Sass, G., Creed, I. F., Sitar, M., Kalbfleisch, W., & Treitz, P. (2005). Vegetation class dependent errors in lidar ground elevation and canopy height estimates in a boreal wetland environment. Canadian Journal of Remote Sensing, 31(2), 191–206. https://doi.org/10.5589/m05-007

Hovi, A., Korhonen, L., Vauhkonen, J., & Korpela, I. (2016). LiDAR waveform features for tree species classification and their sensitivity to tree- and acquisition related parameters. Remote Sensing of Environment, 173, 224–237. https://doi.org/10.1016/j.rse.2015.08.019

Hovi, A., & Korpela, I. (2014). Real and simulated waveform-recording LiDAR data in juvenile boreal forest vegetation. Remote Sensing of Environment, 140, 665–678. https://doi.org/10.1016/j.rse.2013.10.003

Huang, J., Li, Y., Fu, C., Chen, F., Fu, Q., Dai, A., Shinoda, M., Ma, Z., Guo, W., Li, Z. and Zhang, L., (2017). Dryland climate change: Recent progress and challenges. Reviews of Geophysics. http://onlinelibrary.wiley.com/doi/10.1002/2016RG000550/full

Hurtt, G. C., Dubayah, R., Drake, J., Moorcroft, P. R., Pacala, S. W., Blair, J. B., & Fearon, M. G. (2004). Beyond potential vegetation: Combining lidar data and a height- structured model for carbon studies. Ecological Applications, 14(3), 873-883.

Ilangakoon, N. T., Glenn, N. F., Dashti, H., Painter, T. H., Mikesell, T. D., Spaete, L. P., & Shannon, K. (2018). Constraining plant functional types in a semi-arid ecosystem with waveform lidar. Remote Sensing of Environment, 209, 497-509.

Isbell, F., Craven, D., Connolly, J., Loreau, M., Schmid, B., Beierkuhnlein, C., … Eisenhauer, N. (2015). Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature*, *526*(7574), 574–577. https://doi.org/10.1038/nature15374

Jensen, J. R. (2005). Remote sensing and digital image processing. Introductory Digital Image Processing. A Remote Sensing Perspective, 1-34.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F. W., Asner, G. P., … Ustin, S. L. (2016). Monitoring plant functional diversity from space. *Nature Plants*, *2*(3). https://doi.org/10.1038/NPLANTS.2016.24

Jutzi, B., & Stilla, U. (2006). Range determination with waveform recording laser systems using a Wiener Filter. ISPRS Journal of Photogrammetry and Remote Sensing, 61(2), 95–107. https://doi.org/10.1016/j.isprsjprs.2006.09.001

Ko, C., Sohn, G., Remmel, T., & Miller, J. (2014). Hybrid Ensemble Classification of Tree Genera Using Airborne LiDAR Data. Remote Sensing, 6(11), 11225–11243. https://doi.org/10.3390/rs61111225

Ko, C., Sohn, G., Remmel, T., & Miller, J. (2016). Maximizing the Diversity of Ensemble Random Forests for Tree Genera Classification Using High Density LiDAR Data. Remote Sensing, 8(8), 646. https://doi.org/10.3390/rs8080646

Koetz, B., Morsdorf, F., Sun, G., Ranson, K. J., Itten, K., & Allgower, B. (2006). Inversion of a Lidar Waveform Model for Forest Biophysical Parameter Estimation. IEEE Geoscience and Remote Sensing Letters, 3(1), 49–53. https://doi.org/10.1109/LGRS.2005.856706

Korpela, I., Hovi, A., & Korhonen, L. (2013). Backscattering of individual LiDAR pulses from forest canopies explained by photogrammetrically derived vegetation structure. ISPRS Journal of Photogrammetry and Remote Sensing, 83, 81–93. https://doi.org/10.1016/j.isprsjprs.2013.06.002

Kulawardhana, R. W., Popescu, S. C., & Feagin, R. A. (2017). Airborne lidar remote sensing applications in non-forested short stature environments: a review. Annals of Forest Research, 60(1). https://doi.org/10.15287/afr.2016.719

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Prentice, I. C. (2005). A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system: DVGM FOR COUPLED CLIMATE STUDIES. Global Biogeochemical Cycles, 19(1). https://doi.org/10.1029/2003GB002199

Lavorel, S., McIntyre, S., Landsberg, J., & Forbes, T. D. A. (1997). Plant functional classifications: from general groups to specific groups based on response to disturbance. Trends in Ecology & Evolution, 12(12), 474–478. https://doi.org/10.1016/S0169-5347(97)01219-6

Li, A., Dhakal, S., Glenn, N.F., Spaete, L.P., Shinneman, D.J., Pilliod, D.S., Arkle, R.S. and McIlroy, S.K., (2017). Lidar aboveground vegetation biomass estimates in shrublands: Prediction, uncertainties and application to coarser scales. Remote Sensing, 9(9), p.903.

Li, Z., Strahler, A., Schaaf, C., Jupp, D., Schaefer, M., & Olofsson, P. (2018). Seasonal change of leaf and woody area profiles in a midlatitude deciduous forest canopy from classified dual-wavelength terrestrial lidar point clouds. Agricultural and Forest Meteorology, 262, 279–297. https://doi.org/10.1016/j.agrformet.2018.07.014

Li, A., Glenn, N. F., Olsoy, P. J., Mitchell, J. J., & Shrestha, R. (2015). Aboveground biomass estimates of sagebrush using terrestrial and airborne LiDAR data in a dryland ecosystem. Agricultural and Forest Meteorology, 213, 138–147. https://doi.org/10.1016/j.agrformet.2015.06.005

Lin, Y.-C. (2015). Normalization of Echo Features Derived from Full-Waveform Airborne Laser Scanning Data. Remote Sensing, 7(3), 2731–2751. https://doi.org/10.3390/rs70302731

Lin, Y.-C., Mills, J. P., & Smith-Voysey, S. (2010). Rigorous pulse detection from full-waveform airborne laser scanning data. International Journal of Remote Sensing, 31(5), 1303–1324. https://doi.org/10.1080/01431160903380599

Ludwig, J. A., & Reynolds, J. F. (1988). Diversity indices. Statistical Ecology: a primer on methods and computing, 85-101.

MacArthur, R. & MacArthur, J. On bird species diversity. Ecology 42, 595–599 (1961).

Milenković, M., Schnell, S., Holmgren, J., Ressl, C., Lindberg, E., Hollaus, M., Pfeifer, N., & Olsson, H. (2017). Influence of footprint size and geolocation error on the precision of forest biomass estimates from space-borne waveform LiDAR. Remote Sensing of Environment, 200, 74–88. https://doi.org/10.1016/j.rse.2017.08.014

Malenovsky, Z., Martin, E., Homolova, L., Gastelluetchegorry, J., Zuritamilla, R., Schaepman, M., Cudlin, P. (2008). Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution. Remote Sensing of Environment, 112(1), 1–18. https://doi.org/10.1016/j.rse.2006.02.028

Mallet, C., Bretar, F., Roux, M., Soergel, U., & Heipke, C. (2011). Relevance assessment of full- waveform lidar data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing, 66(6), S71–S84. https://doi.org/10.1016/j.isprsjprs.2011.09.008

Mallet, C., Lafarge, F., Roux, M., Soergel, U., Bretar, F., & Heipke, C. (2010). A Marked Point Process for Modeling Lidar Waveforms. IEEE Transactions on Image Processing, 19(12), 3204–3221. https://doi.org/10.1109/TIP.2010.2052825

MATLAB 2016b, The MathWorks Inc., Natick, MA, 2016

Marselis, S. M., Tang, H., Armston, J., Abernethy, K., Alonso, A., Barbier, N., Dubayah, R. (2019). Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. *Environmental Research Letters*, *14*(9), 094013. https://doi.org/10.1088/1748-9326/ab2dcd

Martin, R. E. (1981). Average biomass of four northwest shrubs by fuel size class and crown cover (Vol. 374). US Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station.

Medeiros, C.D., Scoffoni, C., John, G.P., Bartlett, M.K., Inman- Narahari, F., Ostertag, R., Cordell, S., Giardina, C. and Sack, L., 2019. An extensive suite of functional traits distinguishes Hawaiian wet and dry forests and enables prediction of species vital rates. Functional ecology, 33(4), pp.712-734.

Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J.H.C., Reich, P., Bahn, M. and Niinemets, Ü., 2018. A methodology to derive global maps of leaf traits using remote sensing and climate data. Remote sensing of environment, 218, pp.69-88.

Narine, L. L., Popescu, S., Neuenschwander, A., Zhou, T., Srinivasan, S., & Harbeck, K. (2019). Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. Remote Sensing of Environment, 224, 1-11.

Nautiyal, S., Bhaskar, K., & Khan, Y. I. (2015). Biodiversity of Semiarid Landscape. Springer International Publishing, Switzerland, 398pp.

Neuenschwander, A. L. (2009). Landcover classification of small-footprint, full-waveform lidar data. Journal of Applied Remote Sensing, 3(1), 033544. https://doi.org/10.1117/1.3229944

Norman, J.M., and G.S. Campbell. 1989. Canopy structure, pp 301-325. In Pearcy, R.W., J. Ehleringer, H.A. Mooney, and P.W. Rundel (Eds.). Plant Physiological Ecology: Field Methods and Instrumentation. Chapman and Hall, New York, NY

Olsoy, P. J., Mitchell, J. J., Levia, D. F., Clark, P. E., & Glenn, N. F. (2016). Estimation of big sagebrush leaf area index with terrestrial laser scanning. Ecological Indicators, 61, 815–821. https://doi.org/10.1016/j.ecolind.2015.10.034

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., Winstral, A. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. Remote Sensing of Environment, 184, 139–152. https://doi.org/10.1016/j.rse.2016.06.018

Pakeman, R. J., & Stockan, J. A. (2014). Drivers of carabid functional diversity: abiotic environment, plant functional traits, or plant functional diversity? Ecology, 95(5), 1213–1224. https://doi.org/10.1890/13-1059.1

Pang, Y., Lefsky, M., Sun, G., & Ranson, J. (2011). Impact of footprint diameter and off-nadir pointing on the precision of canopy height estimates from spaceborne lidar. Remote Sensing of Environment, 115(11), 2798-2809.

Poorter, L., & Markesteijn, L. (2008). Seedling traits determine drought tolerance of tropical tree species. Biotropica, 40(3), 321-331.

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., & Zimmermann, N. E. (2011). Plant functional type mapping for earth system models. Geoscientific Model Development, 4(4), 993–1010. https://doi.org/10.5194/gmd-4-993-2011

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Peylin, P. (2015). Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative. Geoscientific Model Development, 8(7), 2315–2328. https://doi.org/10.5194/gmd-8-2315-2015

Prim, R. C. Shortest connection networks and some generalizations. Bell Syst. Tech. J. 36, 1389–1401 (1957).

Qi, W., & Dubayah, R. O. (2016). Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping. Remote Sensing of Environment, 187, 253–266. https://doi.org/10.1016/j.rse.2016.10.018

Qi, W., Saarela, S., Armston, J., Ståhl, G., & Dubayah, R. (2019). Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data. Remote Sensing of Environment, 232, 111283.

Qi, W., Lee, S.-K., Hancock, S., Luthcke, S., Tang, H., Armston, J., & Dubayah, R. (2019). Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. Remote Sensing of Environment, 221, 621–634. https://doi.org/10.1016/j.rse.2018.11.035

Rango, A., Chopping, M., Ritchie, J., Havstad, K., Kustas, W., & Schmugge, T. (2000). Morphological Characteristics of Shrub Coppice Dunes in Desert Grasslands of Southern New Mexico derived from Scanning LIDAR. Remote Sensing of Environment, 74(1), 26–44. https://doi.org/10.1016/S0034-4257(00)00084-5

Ranson, K. J., & Sun, G. (2000). Modeling lidar returns from forest canopies. IEEE Transactions on Geoscience and Remote Sensing, 38(6), 2617–2626. https://doi.org/10.1109/36.885208

Reitberger, J., Krzystek, P., & Stilla, U. (2008). Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees. International Journal of Remote Sensing, 29(5), 1407–1431. https://doi.org/10.1080/01431160701736448

Riegl LMS-Q1560. (n.d.). Retrieved September 11, 2017, from http://www.riegl.com/nc/products/airborne-scanning/produktdetail/product/scanner/38/

Rödig, E., Knapp, N., Fischer, R., Bohn, F. J., Dubayah, R., Tang, H., & Huth, A. (2019). From small-scale forest structure to Amazon-wide carbon estimates. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-13063-y

Roncat, A., Bergauer, G., & Pfeifer, N. (2011). B-spline deconvolution for differential target cross-section determination in full-waveform laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing, 66(4), 418–428. https://doi.org/10.1016/j.isprsjprs.2011.02.002

Roncat, A., Pfeifer, N., & Briese, C. (2012). A linear approach for radiometric calibration of full- waveform Lidar data. In L. Bruzzone (Ed.) (p. 853708). https://doi.org/10.1117/12.970305

Schlesinger, W. H., Reynolds, J. F., Cunningham, G. L., Huenneke, L. F., Jarrell, W. M., Virginia, R. A., & Whitford, W. G. (1990). Biological feedbacks in global desertification. Science, 247(4946), 1043-1048.

Schimel, D. S., Asner, G. P., & Moorcroft, P. (2013). Observing changing ecological diversity in the Anthropocene. Frontiers in Ecology and the Environment, 11(3), 129–137. https://doi.org/10.1890/120111

Schimel, D., & Schneider, F. D. (2019). Flux towers in the sky: global ecology from space. *New Phytologist, 224*(2), 570–584. https://doi.org/10.1111/nph.15934

Schneider, F. D., Morsdorf, F., Schmid, B., Petchey, O. L., Hueni, A., Schimel, D. S., & Schaepman, M. E. (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. Nature Communications, 8(1). https://doi.org/10.1038/s41467-017-01530-3

Serbin, S. P., Wu, J., Ely, K. S., Kruger, E. L., Townsend, P. A., Meng, R., & Rogers, A. (2019). From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. New Phytologist, 224(4), 1557-1568.

Seyfried, M. S., Harris, R. C., Marks, D., & Jacob, B. (2000). A geographic database for watershed research, Reynolds Creek Experimental Watershed, Idaho, USA. Tech. Bull. NWRC 2000, 3, 26.

Silva, C. A., Saatchi, S., Garcia, M., Labriere, N., Klauberg, C., Ferraz, A., Meyer, V., Jeffery, K. J., Abernethy, K., White, L., Zhao, K., Lewis, S. L., & Hudak, A. T. (2018). Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study From Central Gabon. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(10), 3512–3526. https://doi.org/10.1109/JSTARS.2018.2816962

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Venevsky, S. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology, 9(2), 161–185. https://doi.org/10.1046/j.1365-2486.2003.00569.x

Stavros, E. N., Schimel, D., Pavlick, R., Serbin, S., Swann, A., Duncanson, L., Wennberg, P. (2017). ISS observations offer insights into plant function. Nature Ecology & Evolution, 1(7), 0194. https://doi.org/10.1038/s41559-017-0194

Stiver, S. J., Rinkes, E. T., Naugle, D. E., Makela, P. D., Nance, D. A., & Karl, J. W. (2015). Sage-Grouse Habitat Assessment Framework: Multiscale Habitat Assessment Tool. Bureau of Land Management and Western Association of Fish and Wildlife Agencies Technical Reference, 6710-1

Streutker, D. R., & Glenn, N. F. (2006). LiDAR measurement of sagebrush steppe vegetation heights. Remote Sensing of Environment, 102(1–2), 135–145. https://doi.org/10.1016/j.rse.2006.02.011

Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D. B., & Blair, B. (2012). Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. Remote Sensing of Environment, 124, 242-250.

Terrascan Manual (2016). TerrScan User’s Guide. Terrasolid.

Thomas, R. Q., Hurtt, G. C., Dubayah, R., & Schilz, M. H. (2008). Using lidar data and a height- structured ecosystem model to estimate forest carbon stocks and fluxes over mountainous terrain. Canadian Journal of Remote Sensing, 34(sup2), S351–S363. https://doi.org/10.5589/m08-036

Tilman, D., & Downing, J. A. (1994). Biodiversity and stability in grasslands. Nature, 367(6461), 363-365.

Urbazaev, M., Thiel, C., Cremer, F., Dubayah, R., Migliavacca, M., Reichstein, M., & Schmullius, C. (2018). Estimation of forest aboveground biomass and uncertainties by integration of field measurements, airborne LiDAR, and SAR and optical satellite data in Mexico. Carbon Balance and Management, 13(1). https://doi.org/10.1186/s13021-018-0093-5

Vaughn, N. R., Moskal, L. M., & Turnblom, E. C. (2012). Tree Species Detection Accuracies Using Discrete Point Lidar and Airborne Waveform Lidar. Remote Sensing, 4(2), 377–403. https://doi.org/10.3390/rs4020377

Villéger, S., Mason, N. W., & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. Ecology, 89(8), 2290-2301.

Violle, C., Reich, P. B., Pacala, S. W., Enquist, B. J., & Kattge, J. (2014). The emergence and promise of functional biogeography. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13690–13696. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4183284/

Wagner, W., Hollaus, M., Briese, C., & Ducic, V. (2008). 3D vegetation mapping using small- footprint full- waveform airborne laser scanners. International Journal of Remote Sensing, 29(5), 1433– 1452. https://doi.org/10.1080/01431160701736398

Wagner, Wolfgang. (2010). Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts. ISPRS Journal of Photogrammetry and Remote Sensing, 65(6), 505–513. https://doi.org/10.1016/j.isprsjprs.2010.06.007

Wagner, Wolfgang, Ullrich, A., Ducic, V., Melzer, T., & Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. ISPRS Journal of Photogrammetry and Remote Sensing, 60(2), 100–112. https://doi.org/10.1016/j.isprsjprs.2005.12.001

Wang, Y., Ni, W., Sun, G., Chi, H., Zhang, Z., & Guo, Z. (2019). Slope-adaptive waveform metrics of large footprint lidar for estimation of forest aboveground biomass. Remote Sensing of Environment, 224, 386-400.

Wieczynski, D. J., Boyle, B., Buzzard, V., Duran, S. M., Henderson, A. N., Hulshof, C. M., Kerkhoff, A. J., McCarthy, M. C., Michaletz, S. T., Swenson, N. G., Asner, G. P., Bentley, L. P., Enquist, B. J., & Savage, V. M. (2019). Climate shapes and shifts functional biodiversity in forests worldwide. Proceedings of the National Academy of Sciences, 116(2), 587–592. https://doi.org/10.1073/pnas.1813723116

Wilson, A.M., Jetz, W., (2016) Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. PLoS Biol 14(3): e1002415. doi:10.1371/journal. pbio.1002415.

Wright, I. J. 2004. The worldwide leaf economics spectrum. Nature 428: 621–827.

Wu, J., van Aardt, J. A. N., & Asner, G. P. (2011). A Comparison of Signal Deconvolution Algorithms Based on Small-Footprint LiDAR Waveform Simulation. IEEE Transactions on Geoscience and Remote Sensing, 49(6), 2402–2414. https://doi.org/10.1109/TGRS.2010.2103080

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., Xu, X. (2014). Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. Annals of Botany, 114(1), 1–16. https://doi.org/10.1093/aob/mcu077

Yao, W., Krzystek, P., & Heurich, M. (2012). Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. Remote Sensing of Environment, 123, 368–380. https://doi.org/10.1016/j.rse.2012.03.027

Yosef, G., Walko, R., Avisar, R., Tatarinov, F., Rotenberg, E., & Yakir, D. (2018). Large-scale semi-arid afforestation can enhance precipitation and carbon sequestration potential. Scientific reports, 8(1), 1-10.

Zahidi, I., Yusuf, B., Hamedianfar, A., Shafri, H. Z. M., & Mohamed, T. A. (2015). Object-based classification of QuickBird image and low point density LIDAR for tropical trees and shrubs mapping. European Journal of Remote Sensing, 48(1), 423–446. https://doi.org/10.5721/EuJRS20154824

Zhu, J., Jiang, L., & Zhang, Y. (2016). Relationships between functional diversity and aboveground biomass production in the Northern Tibetan alpine grasslands. Scientific reports, 6(1), 1-8.

Zhuang, W., & Mountrakis, G. (2015). Ground peak identification in dense shrub areas using large footprint waveform LiDAR and Landsat images. International Journal of Digital Earth, 8(10), 805–824. https://doi.org/10.1080/17538947.2014.942716

APPENDIX A

Supplementary figures for chapter 3

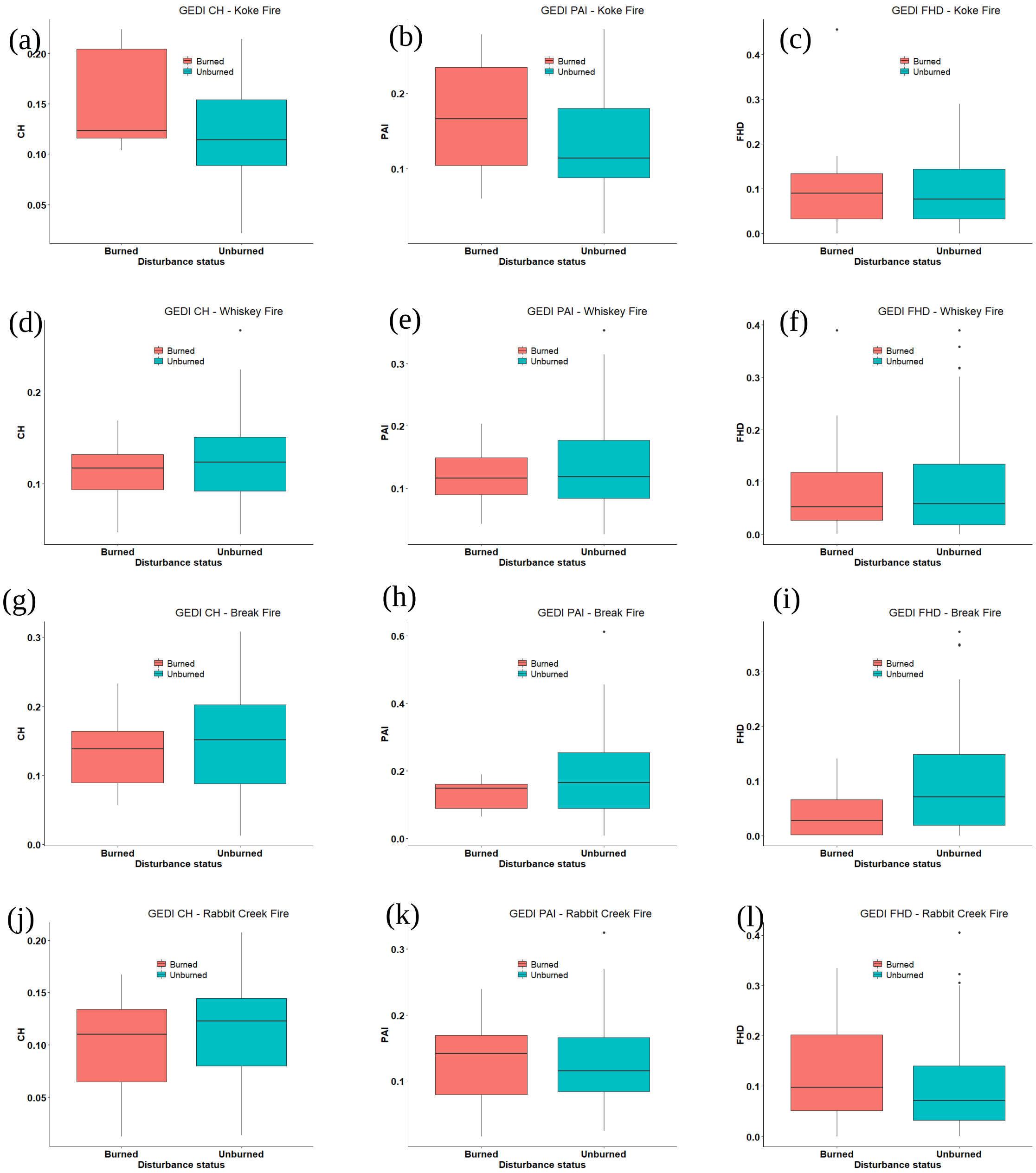


Figure S.1 Boxplots of CH, PAI, and FHD derived from GEDI of burned and surrounding unburned areas of Koke fire (a, b, c), Whiskey fire (d, e, f), Break fire (g, h, i), and Rabbit Creek fire (j, k, l).

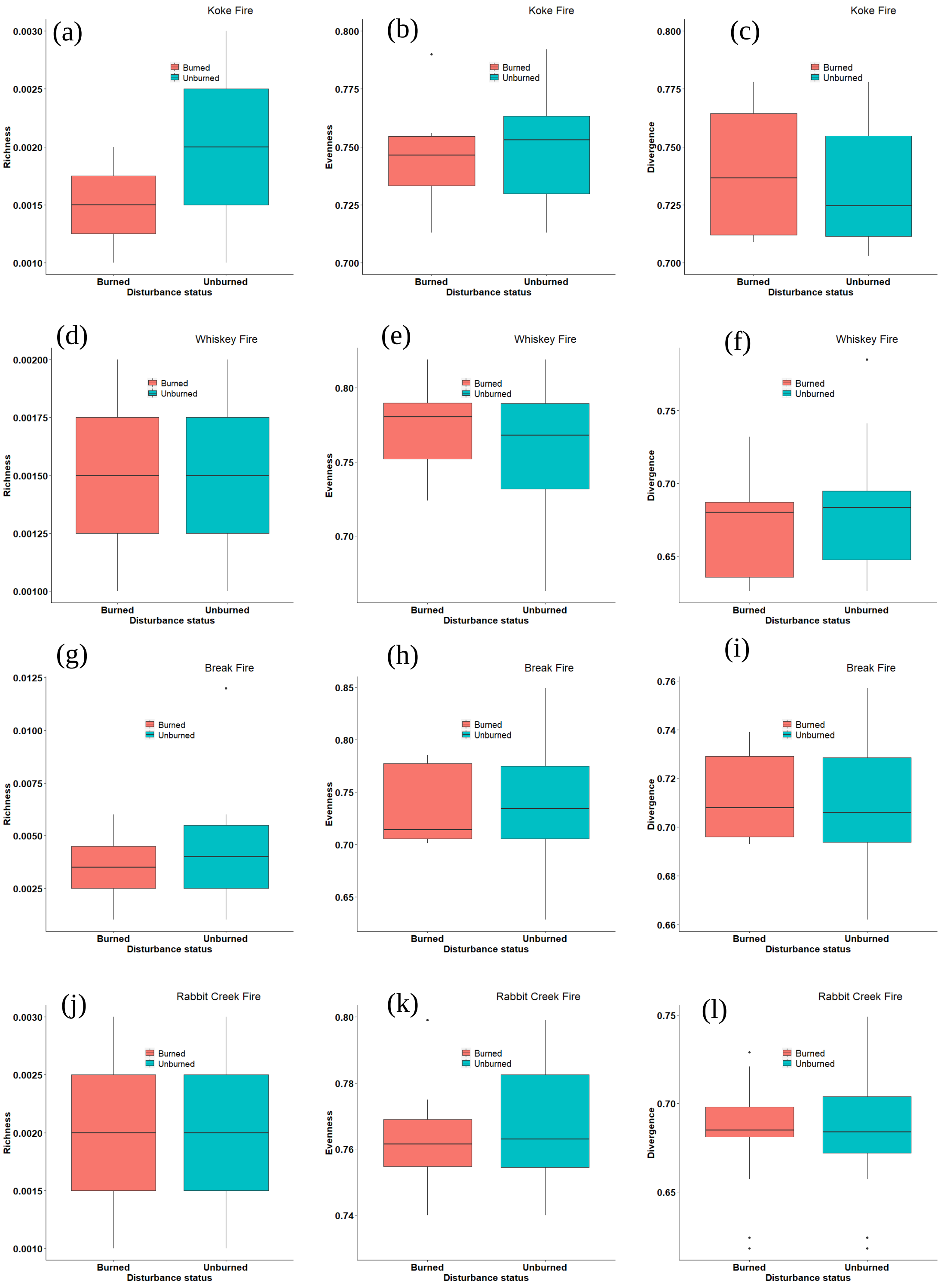


Figure S.2 Boxplots of functional richness, functional evenness, and functional divergence calculated using GEDI functional traits of burned and surrounding unburned areas of Koke fire (a, b, c), Whiskey fire (d, e, f), Break fire (g, h, i), and Rabbit Creek fire (j, k, l). The first column represents

functional richness while the second and third columns represent functional evenness and functional divergence respectively.

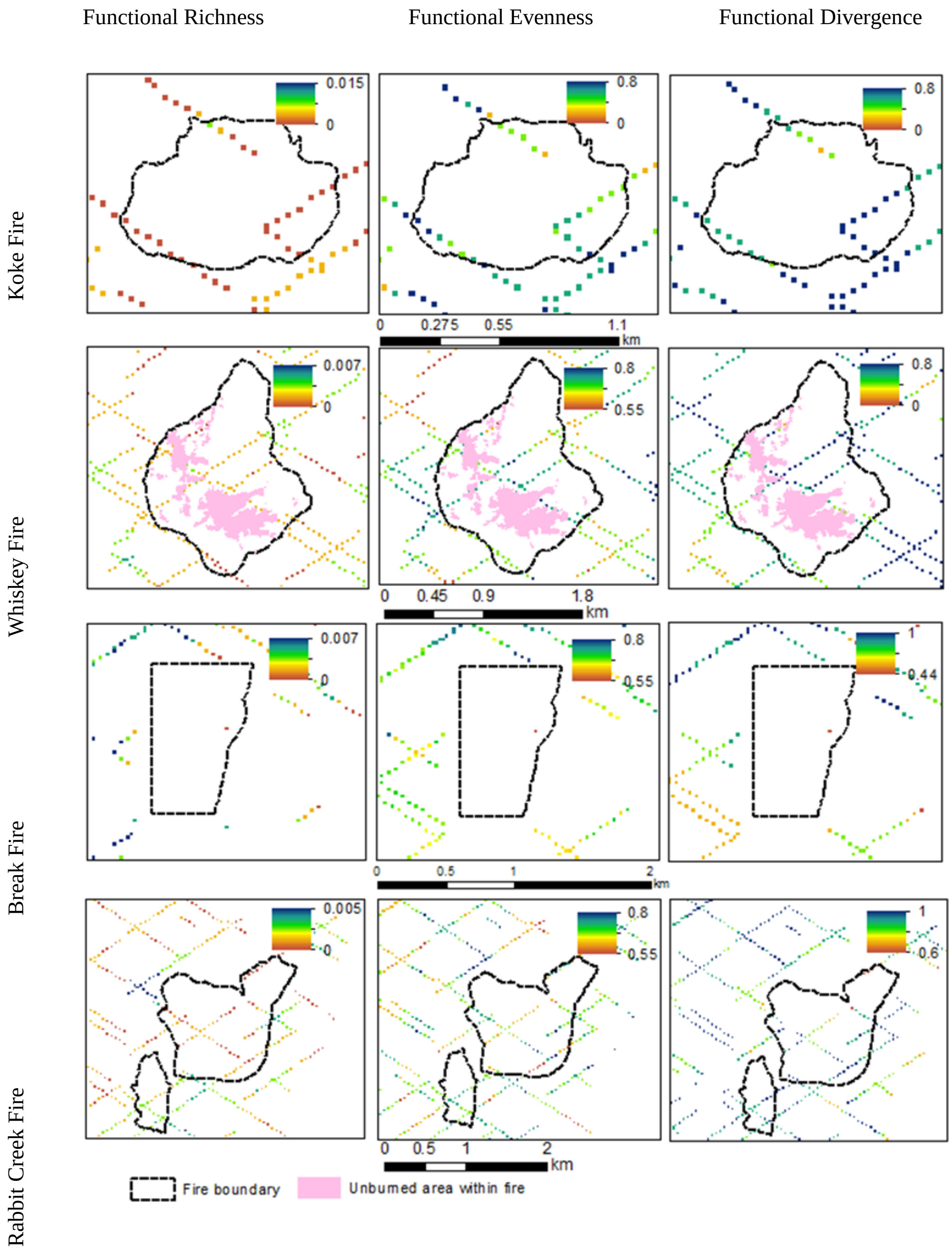

Functional Richness
Functional Evenness
Functional Divergence
Koke Fire
Whiskey Fire
Break Fire
Rabbit Creek Fire
0.015
0
0.8
0
0.8
0
0 0.275 0.55 1.1 km
0.007
0
0.8
0.55
0.8
0
0 0.45 0.9 1.8 km
0.007
0
0.8
0.55
1
0.44
0 0.5 1 2 km
0.005
0
0.8
0.55
1
0.6
0 0.5 1 2 km
Fire boundary
Unburned area within fire

Figure S.3 GEDI footprints colored by functional diversity of the four different fires studied. The first column represents functional richness while the second and third columns represent functional evenness and functional divergence respectively.

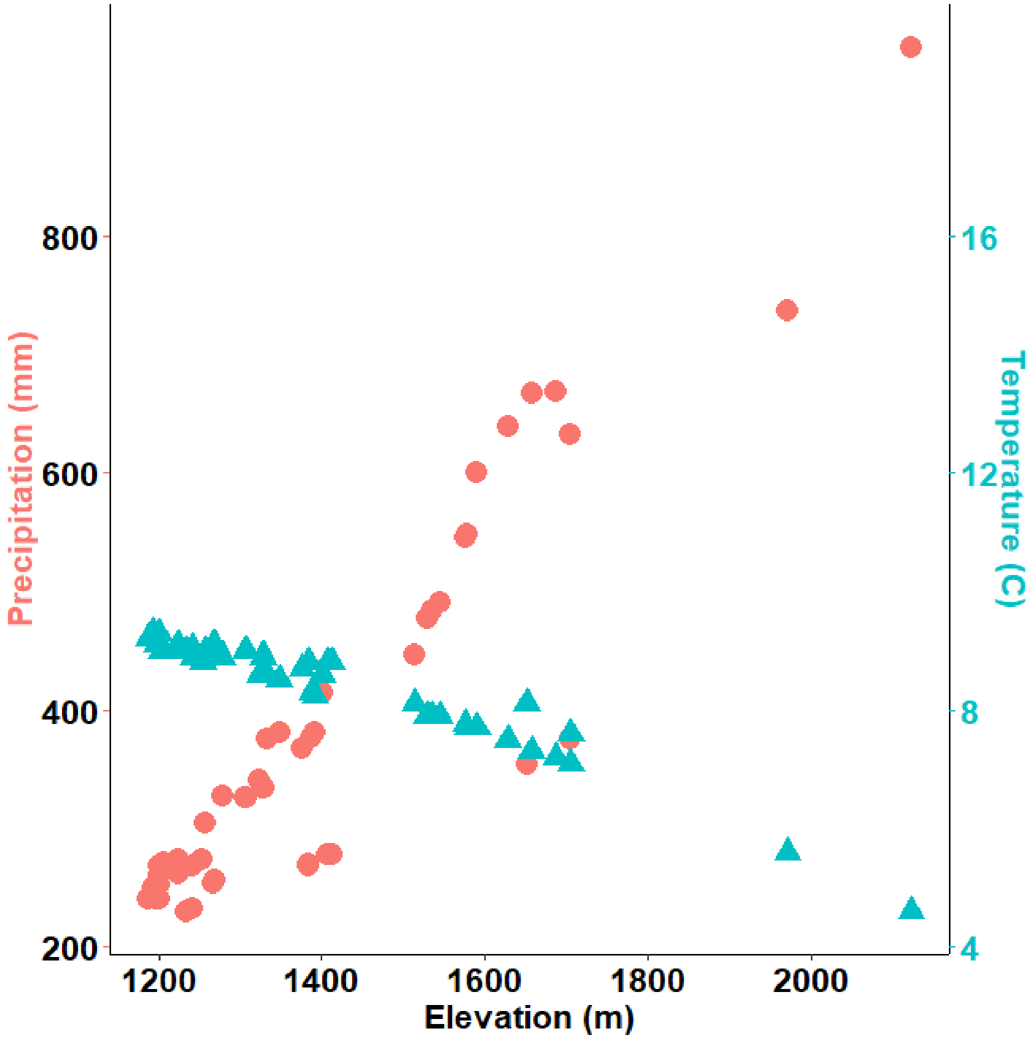


Figure S.4 RCEW precipitation and temperature variation with elevation gradient. Precipitation: average for the period of 1963-2010, temperature: mean normal air temperature for the period of 1984 – 2014 (PRISM Climate Group, 2016, CZO Dataset: Reynolds Creek, 2016)

Supplementary figures for chapter 4

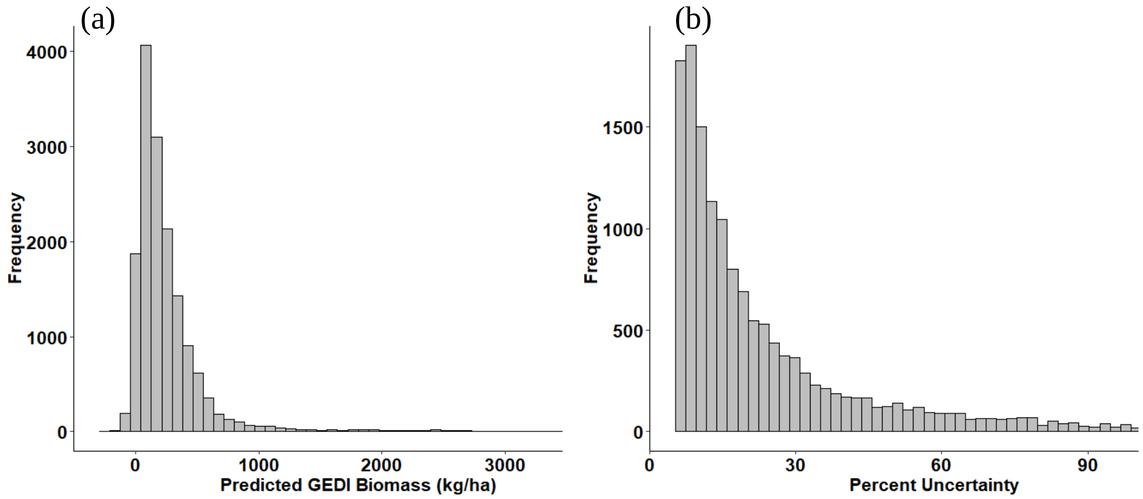


Figure S.5 Frequency distribution of a) predicted GEDI biomass and b) percent uncertainty at GEDI footprint scale.

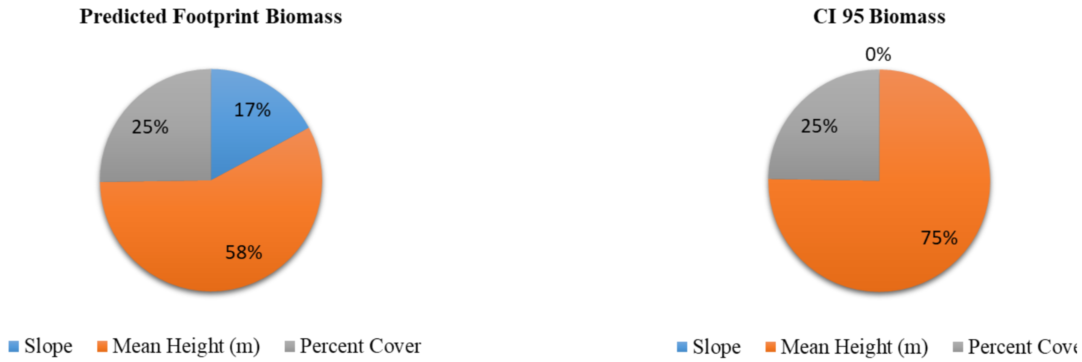


Figure S.6 Variance of predicted footprint biomass and percent uncertainty explained by Slope, Vegetation heights and percent vegetation cover.