Rochester Institute of Technology

RIT Scholar Works

Theses

7-16-2010

Fusing Small-footprint Waveform LiDAR and Hyperspectral Data for Canopy-level Species Classification and Herbaceous Biomass Modeling in Savanna Ecosystems

Diane Sarrazin

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Sarrazin, Diane, "Fusing Small-footprint Waveform LiDAR and Hyperspectral Data for Canopy-level Species Classification and Herbaceous Biomass Modeling in Savanna Ecosystems" (2010). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

Fusing Small-footprint Waveform LiDAR and Hyperspectral Data
for Canopy-level Species Classification and Herbaceous Biomass
Modeling in Savanna Ecosystems

by

Diane Sarrazin

B.S. Royal Military College of Canada, 2005

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

16 July 2010

Signature of the Author _____

Accepted by _____
Coordinator, M.S. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Diane Sarrazin
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science

Dr. Jan van Aardt, Thesis Advisor

Dr. David Messinger

Dr. Emmett Ientilucci

Date

THESIS RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Thesis:

**Fusing Small-footprint Waveform LiDAR and Hyperspectral Data for
Canopy-level Species Classification and Herbaceous Biomass Modeling
in Savanna Ecosystems**

I, Diane Sarrazin, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____ Date _____

Acknowledgements

During my time at RIT, I have been fortunate to work with many talented friends and colleagues. I would like to express my sincere gratitude to Dr. Jan van Aardt, who made his support available in a number of ways from the beginning of the project to the end, and enabled me to understand the matters and issues related to the fusion of wLiDAR and HSI. His motivation, enthusiasm, and knowledge helped me immensely both during the research and the writing of this thesis.

I would also like to thank Dr. David Messinger, who was always available to answer my questions and to make sense of ideas that I had with regards to wLiDAR and HSI, as well as Dr. Emmett Ientilucci, who provided me with insightful knowledge and advice when reading through my work. In addition to my thesis committee, I am grateful to the waveform group, Jiaying Wu, Joe McGlinchy, and David Kelbe with whom many ideas were exchanged and developed. My sincere thanks also go to Cindy Schultz for her encouragement throughout these two past years at RIT. I would also like to express my gratitude to the Canadian Forces for providing the financial support toward the achievement of this thesis.

My deepest gratitude goes to my family for their unwavering love and support throughout this project. They are, without a doubt, the key to my success.

*This thesis is dedicated to my husband, who supported me throughout this endeavor.
Your understanding, patience, and love made all the difference in my life,
even with the distance that kept us apart.*

Fusing Small-footprint Waveform LiDAR and Hyperspectral Data for Canopy-level Species Classification and Herbaceous Biomass Modeling in Savanna Ecosystems

by

Diane Sarrazin

Submitted to the

Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Master of Science Degree
at the Rochester Institute of Technology

Abstract

The study of ecosystem structure, function, and composition has become increasingly important in order to gain a better understanding of how impacts wrought by natural disturbances, climate, and human activity can alter ecosystem services provided to a population. Research groups at Rochester Institute of Technology and Carnegie Institution for Science are focusing on characterization of savanna ecosystems and are using data from the Carnegie Airborne Observatory (CAO), which integrates advanced imaging spectroscopy and waveform light detection and ranging (wLiDAR) data. This component of the larger ecosystem project has as a goal the fusion of imaging spectroscopy and small-footprint wLiDAR data in order to improve per-species structural parameter estimation towards classification and herbaceous biomass modeling. Waveform LiDAR has proven useful for extracting high vertical resolution structural parameters, while imaging spectroscopy is a well-established tool for species classification and biochemistry assessment. We hypothesize that the two modalities provide complementary information that could improve per-species structural assessment, species classification, and herbaceous biomass modeling when compared to single modality sensing systems. We explored a statistical approach to data fusion at the feature level, which hinged on our ability to reduce structural and spectral data dimensionality to those data features best suited to assessing these complex systems. The species classification approach was based on stepwise discrimination analysis (SDA) and used feature metrics from hyperspectral imagery (HSI) combined with wLiDAR data, which could help finding correlated features, and in turn improve classifiers. It was found that fusing data with the SDA did not improve classification significantly, especially compared to the HSI classification results. The overall classification accuracies were 53%

for both original and PCA-based wLiDAR variables, 73% for the original HSI variables, 71% for PCA-based HSI variables, 73% for the original fusion of wLiDAR and HSI data set, and 74% for the PCA-based fusion variables. The *kappa* coefficients achieved with the original and PCA-based wLiDAR variable classifications were 0.41 and 0.44, respectively. For both original and PCA-based HSI classifications, the *kappa* coefficients were 0.63 and 0.60, respectively and 0.62 and 0.64 for original and PCA-based fusion variable classifications, respectively. These results show that HSI was more successful in grouping important information in a smaller number of variables than wLiDAR and thus inclusion of structural information did not significantly improve the classification. As for herbaceous biomass modeling, the statistical approach used for the fusion of wLiDAR and HSI was forward selection modeling (FSM), which selects significant independent metrics and models those to measured biomass. The results were measured in R^2 and RMSE, which indicate the similar findings. Waveform LiDAR performed the poorest with an R^2 of 0.07 for original wLiDAR variables and 0.12 for PCA-based wLiDAR variables. The respective RMSE were 19.99 and 19.41. For both original and PCA-based HSI variables, the results were better with R^2 of 0.32 and 0.27 and RMSE of 17.27 and 17.80, respectively. For the fusion of original and PCA-based data, the results were comparable to HSI, with R^2 values of 0.35 and 0.29 and RMSE of 16.88 and 17.59, respectively. These results indicate that small scale wLiDAR may not be able to provide accurate measurement of herbaceous biomass, although other factors could have contributed to the relatively poor results, such as the senescent state of grass by April 2008, the narrow biomass range that was measured, and the low biomass values, i.e., the limited laser-target interactions. We concluded that although fusion did not result in significant improvements over single modality approaches in those two use cases, there is a need for further investigation during peak growing season.

Contents

List of Figures	xi
List of Tables	xiii
1 Introduction	1
1.1 Literature Review	4
1.1.1 Data Fusion	5
1.1.2 Species Classification	7
1.1.3 LiDAR-based Biomass Modeling	14
2 Methodology	18
2.1 Data Collection	19
2.2 Remote Sensing Instruments	20
2.3 Signal Preprocessing	24
2.4 Species Classification	25
2.4.1 Segmentation	25
2.4.2 Waveform Compositing	25
2.4.3 Feature Metric Extraction	26
2.4.4 Fusion Method - Stepwise Discriminant Analysis	28
2.4.5 Error Analysis	30
2.5 Biomass Modeling	32
2.5.1 Waveform Compositing	32
2.5.2 Gaussian modeling and feature selection	32
2.5.3 Fusion Method - Forward Selection Model	33
2.5.4 Error Analysis	33
3 Results and Discussion	34
3.1 Species Classification	34
3.1.1 Segmentation	34
3.1.2 Feature Metric Extraction	35

3.1.3	Feature-level Fusion - Stepwise Procedure	42
3.1.4	Feature-level Fusion - Discriminant Analysis	49
3.1.5	Z-test	55
3.2	Herbaceous Biomass Modeling	58
3.2.1	Site-level Plot Selection	58
3.2.2	Feature Metric Extraction	58
3.2.3	Feature-level Fusion - Forward Selection Model	67
3.2.4	Herbaceous Biomass Modeling Results	76
4	Conclusion	83
	Appendices	87
A	Species Classification Tables	88
A.1	Stepwise Selection Procedure	88
A.1.1	Waveform LiDAR	88
A.1.2	HSI	90
A.2	Discriminant Analysis	91
A.2.1	Waveform LiDAR	91
A.2.2	HSI	92
A.2.3	Fusion of wLiDAR and HSI	93
A.3	Classification Results	93
A.3.1	Waveform LiDAR	94
A.3.2	HSI	98
A.3.3	Fusion of wLiDAR and HSI	102
B	Herbaceous Biomass Modeling Tables	105
B.1	Forward Selection Model	105
B.1.1	Waveform LiDAR	105
B.1.2	HSI	106
B.1.3	Fusion of wLiDAR and HSI	107
C	GPS Points	109
C.1	GPS Points for Species Classification	109
C.2	GPS Points for Herbaceous Biomass Modeling	114
D	Species Classification Algorithm	121
D.1	Main Algorithm	121
D.2	Height and Crown Thickness	133

E	Herbaceous Biomass Modeling Algorithm	140
E.1	Main Algorithm	140
E.2	PCA HSI Calculations	149
E.3	PCA wLiDAR Calculations	153
E.4	Smoothing of imagery and feature metrics based on a 3x3 window	158
E.5	Building Plot-level Composite Waveforms	159
E.6	Gaussian Modeling of Waveforms	163
E.7	Rise Time to Leading Edge and Fall Time of Trailing Edge of Waveforms	168

List of Figures

1.1	Multispectral versus HSI sensing.	8
2.1	Location of the study area in South Africa.	19
2.2	Site-level sampling design.	20
2.3	LiDAR data acquisition.	22
2.4	Waveform visualization of tree height and crown thickness.	26
3.1	GPS locations and ROIs in the Bushbuckridge communal settlement.	35
3.2	Intensity images for tree height and crown thickness.	35
3.3	Derivatives of HSI and wLiDAR.	36
3.4	Waveform derivatives.	37
3.5	PCA results for HSI.	38
3.6	PCA results for wLiDAR.	39
3.7	Histogram of the variance values of the first 41 PCA bands for wLiDAR.	39
3.8	Eigenvector analysis for HSI.	40
3.9	Eigenvector analysis for wLiDAR.	41
3.10	Stepwise results for original wLiDAR variables.	43
3.11	Stepwise results for PCA-based wLiDAR variables.	44
3.12	Stepwise results for original HSI variables.	46
3.13	Stepwise results for HSI variables, with PCA metrics.	47
3.14	Eigenvectors analysis for wLiDAR, PC 6, and 36.	49
3.15	Eigenvectors analysis for HSI, PCs 5, 14, and 20.	50
3.16	An overview of the sites in KNP and the Bushbuckridge communal settle- ment and a site-level example of GPS locations for herbaceous data collection.	58
3.17	Comparison between dirt road and bare soil composite waveforms.	59
3.18	Composite and road waveforms, and difference between the two.	59
3.19	Derivatives of wLiDAR imagery.	60
3.20	Waveform derivatives.	61
3.21	Grass abundance map.	62
3.22	PCA results for HSI.	63

3.23	PCA results for wLiDAR.	63
3.24	Eigenvector analysis for HSI in the case of image areas used for herbaceous biomass modeling. Note the eigenvectors crossing at the beginning of the red edge region ($700nm$) and the strong negative contribution from PC_3 starting at $1017nm$, indicative of potential vegetation information.	64
3.25	Eigenvector analysis for wLiDAR in the case of herbaceous biomass modeling. All PCs have similar eigenvector contributions in the same time range, i.e., $0.75m$ above and below (assumed multiple scattering) the ground. . . .	65
3.26	Gaussian approximation curve.	66
3.27	Forward selection results for original wLiDAR variables.	68
3.28	Forward selection results for PCA-based wLiDAR variables.	69
3.29	Forward selection results for original HSI variables.	71
3.30	Forward selection results for PCA-based HSI variables.	72
3.31	Forward selection results for original fusion variables.	74
3.32	Forward selection results for PCA-based fusion variables.	75
3.33	Graph of predicted versus field measured data for original wLiDAR data. .	76
3.34	Graph of predicted versus field measured data for PCA-based wLiDAR data.	77
3.35	Graph of predicted versus field measured data for original HSI data. . . .	78
3.36	Graph of predicted versus field measured data for PCA-based HSI data. . .	79
3.37	Graph of predicted versus field measured data for original fusion data. . . .	80
3.38	Graph of predicted versus field measured data for PCA-based fusion data. .	81

List of Tables

2.1	Design specifications of the ALTM 3100 for wLiDAR.	23
2.2	Assessment based on the proportion of correct allocations [Foody, 2004]. . .	31
3.1	Confusion matrix resulting from SDA for original wLiDAR variables.	51
3.2	Confusion matrix resulting from SDA for PCA-based wLiDAR variables. . .	51
3.3	Confusion matrix resulting from SDA for original HSI variables.	52
3.4	Confusion matrix resulting from SDA for PCA-based HSI variables.	52
3.5	Confusion matrix resulting from SDA for original fusion variables.	53
3.6	Confusion matrix resulting from SDA for PCA-based fusion variables. . . .	53
3.7	Assessment based on the proportion of correct allocations between original wLiDAR and PCA-based fusion classifications	55
3.8	Assessment based on the proportion of correct allocations between original HSI and PCA-based fusion classifications	56
3.9	Assessment based on the proportion of correct allocations between original wLiDAR and HSI classifications	56
A.1	Summary of stepwise procedure for original wLiDAR variables.	88
A.2	Summary of stepwise procedure for PCA-based wLiDAR.	89
A.3	Summary of stepwise procedure for original HSI variables.	90
A.4	Summary of stepwise procedure for PCA-based HSI variables.	90
A.5	Linear combination of coefficients for species classification - original wLiDAR variables.	91
A.6	Linear combination of coefficients for species classification - PCA-based wLiDAR variables.	91
A.7	Linear combination of coefficients for species classification - original HSI variables.	92
A.8	Linear combination of coefficients for species classification - PCA-based HSI variables.	92
A.9	Linear combination of coefficients for species classification - original fusion variables.	93

A.10 Linear combination of coefficients for species classification - PCA-based fusion variables.	93
A.11 Classification results for species classification - original wLiDAR variables.	94
A.12 Classification results for species classification - original wLiDAR variables (continued).	95
A.13 Classification results for species classification - PCA-based wLiDAR variables.	96
A.14 Classification results for species classification - PCA-based wLiDAR variables (continued).	97
A.15 Classification results for species classification - original HSI variables.	98
A.16 Classification results for species classification - original HSI variables (continued).	99
A.17 Classification results for species classification - PCA-based HSI variables.	100
A.18 Classification results for species classification - PCA-based HSI variables (continued).	101
A.19 Classification results for species classification - original fusion variables.	102
A.20 Classification results for species classification - original fusion variables (continued).	103
A.21 Classification results for species classification - PCA-based fusion variables.	104
B.1 Parameter estimates for herbaceous biomass modeling - original wLiDAR variables.	105
B.2 Parameter estimates for herbaceous biomass modeling - PCA-based wLiDAR variables.	106
B.3 Parameter estimates for herbaceous biomass modeling - original HSI variables.	106
B.4 Parameter estimates for herbaceous biomass modeling - PCA-based HSI variables.	107
B.5 Parameter estimates for herbaceous biomass modeling - original fusion variables.	107
B.6 Parameter estimates for herbaceous biomass modeling - PCA-based fusion variables.	108
C.1 Tree Species Classification - GPS Points.	109
C.2 Herbaceous biomass field data - GPS Points.	114

Chapter 1

Introduction

Land degradation is a major concern in Africa’s sub-Saharan environment, where multi-spectral remote sensing has been used extensively for monitoring purposes [Asner et al., 2007, Cho et al., 2009]. However, in order to explain degradation patterns in further detail, there is a need to increase the scale and sensitivity of the instruments. For example, coarse resolution satellite data have not been able to capture intricate landscape metrics that would enable the analysis of bare soil fractional cover, while high resolution hyper-spectral data have been proven to do so [Asner et al., 2003]. The variations in three dimensional structure of vegetation could also prove useful for assessing land degradation and how vegetation changes over time. Light detection and ranging (LiDAR) has proven to be a very promising tool in providing large scale spatial information [Lefsky et al., 1999b] and could potentially provide fine scale structural data required to study this environmental issue. The Carnegie Airborne Observatory (CAO) consists of an advanced spectroscopic imaging and waveform laser remote sensing technology and is used to study ecosystems anywhere in the world, given the need [CAO, 2008]. The effort undertaken to build this system was driven by the need to acquire information that could be readily converted to physical and chemical quantities representative of ecosystem processes and properties, achieved by fusing multiple sensor data [Asner et al., 2007]. This could help us understand “how changes in land use, climate, and natural disturbances affect the structure, composition, and functioning of ecosystems, and how these changes alter the services provided by ecosystems to people” [CAO, 2008].

Land degradation can further be assessed by performing vegetation inventory, herbaceous and woody biomass estimations in browsing areas, determining the ratio of bare soil cover versus grass, evaluating grass and tree quality in terms of their chemical constituents, etc. This study will focus on the evaluation of multi-sensor fusion to assess its ability to improve tree species classification within the broader scope of vegetation inventory. It will also investigate the possibility of increasing accuracy of herbaceous biomass modeling with the fine scale sensitivity brought by waveform light detection and ranging (wLiDAR)

along with spectral information detail from hyperspectral imagery (HSI).

Fusion of multi-sensor data has been performed in the past at the pixel, feature, and decision levels. An example of pixel-level fusion, [Mutlu et al., 2008] constructed a new multiband image composed of discrete LiDAR data and multispectral imagery, to which they applied principal component analysis (PCA) and a minimum noise fraction (MNF) transform. They applied the maximum likelihood (ML) and Mahalanobis distance supervised classifications to the original multispectral imagery, the new multiband image, and its PCA and MNF bands in order to classify vegetation (grass, timber, brush, hardwood, etc.). The best results were obtained using the new multiband image and the Mahalanobis distance classification.

Using feature-level fusion, [Sugumaran and Voss, 2007] performed object-oriented classification of multispectral imagery using eCognition Professional 4.0, to which they added LiDAR information that facilitated the segmentation process of tree species. The classification using multispectral and LiDAR information obtained better results than classification of multispectral imagery alone. [Dalponte et al., 2008] employed support vector machines (SVMs) and Gaussian maximum likelihood with leave-one-out-covariance (GML-LOOC) to fuse LiDAR and hyperspectral data in complex forest areas. They obtained better results with the SVMs method. Feature-level fusion will be performed in this research as it seems to be the most promising method in most previous studies, i.e., better results were obtained with feature-level fusion than pixel-level fusion.

The first part of this research consists of performing tree species classification in a savanna environment. Multispectral data were employed to perform species classification for years until the hyperspectral capability was developed, which collects information across a broader coverage of the electromagnetic spectrum in a contiguous fashion. In order to perform species classification, many techniques have been used, which fall into unsupervised or supervised categories of algorithms. However, before handling the HSI, data dimensionality reduction was required in order to manipulate data more efficiently. Frequently used methods are PCA or MNF, which gather most of the variability in a fewer number of bands, but statistical methods such as the stepwise regression procedure can also be used to define the optimal and independent bands in a set. For example, [van Aardt and Wynne, 2007] have used airborne visible/infrared imaging spectrometer (AVIRIS) data (224 bands) to investigate species separability. They applied the regression procedure to reduce dimensionality to 10 bands, which were input into discriminant procedures to evaluate between and within class separability, and obtained an 85% overall accuracy.

[Martin et al., 1998] investigated the application of HSI with AVIRIS data for species identification. They reduced the number of bands to 11 based on known significant band regions that correspond to foliage biochemical composition. The ML classification was applied and compared to field measurements for validation purposes. A 75% overall classification accuracy was achieved. Other species classification studies employed HSI such as [Clark et al., 2005], which used three different classification methods, namely linear

discriminant analysis (LDA), ML, and spectral angle mapper (SAM), while [Cho et al., 2009] evaluated the performance of the SAM classifier in the case of single versus multiple endmembers per species. HSI has the advantage of providing chemical composition of the subjects for identification, however it does not take into account the structural information of materials. This was the reason why LiDAR data seemed like an appropriate choice to fuse with HSI, as it had the potential to provide complimentary structural information to increase species classification accuracy.

LiDAR has been used in the past to extract physical characteristics in forest environments, such as canopy height, above ground biomass, basal area, mean stem diameter, canopy volume, and vertical profiles [Means et al., 1999, Lefsky et al., 1999a, Lefsky et al., 1999b, Dubayah and Drake, 2000]. However, little research has been performed on species-level classification using LiDAR. As one of the few, [Holmgren and Persson, 2004] used discrete LiDAR to discriminate between pine and spruce trees from which they extracted physical features and grouped them in pairs exhibiting the largest correlation coefficients. They implemented Student's t-test to evaluate the most significant metrics, which were input into discriminant functions. This resulted in an overall classification of 95% and they also obtained a high correlation ($r = 0.84$) between calculated crown base height and field data.

Another study, by [Reitberger et al., 2006], used Bayes classification on segmented tree species regions using wLiDAR features such as intensity, crown shape, geometry structure, height, and density of trees. These features were obtained from the decomposition of the waveforms into Gaussian approximation curves. In a leaf-on environment, they successfully discriminated between deciduous and coniferous trees with 80% overall accuracy. [Neuenchwander et al., 2009] have also employed wLiDAR data to classify dead trees, brush, trail, and a few tree species. From the decomposed waveforms, they extracted multiple metrics, such as the canopy and ground energy, height of median energy, rise time to the first peak, etc. They performed a Bayesian pairwise classifier to discriminate between pairs of classes and found that the most valuable metrics were energy ratio and Gaussian amplitude. An 86% overall accuracy was obtained. Very few studies have been performed to evaluate species classification using wLiDAR, especially in a savanna ecosystem, which makes this project unique.

The approach that will be used in this research was based on a number of steps. First comes the preprocessing of the data, followed by segmentation of the tree canopies of interest in the regions where field data are available. Feature extraction from both HSI and wLiDAR will provide the information required for the stepwise regression procedure, which identifies the significant variables to use as input into the linear discriminant procedure. The features that will be extracted for HSI are first and second derivatives, and PCA. As for wLiDAR, structural information will be extracted from the composite waveforms of each segmented tree region such as height, horizontal dimensions of the crown, PCA, first and second derivatives, etc. The evaluation of performance will be in the form of a

confusion matrix along with overall accuracy, *kappa*, and z-scores indicators.

The second part of this research consists of modeling herbaceous biomass in the savanna environment. Spectral imagery has been employed in the past to relate spectral response to biomass by using the ratio of the energy at different wavelengths. As an example, Mirik *et al.* (2005a, 2005b) investigated the application of simple ratio-type vegetation indices (SRTVI) and normalized difference-type vegetation indices (NDTVI) to HSI and used it in a regression-type analysis against field data. Nowadays, LiDAR is the most efficient technology for this purpose, since it captures landscape structural data that is appropriate for volume and biomass estimation. As examples, [Zhao et al., 2009] have investigated scale-invariant estimation methods for forest biomass assessment, while [Drake et al., 2002] studied the relationship between canopy structure and above ground biomass using field data and wLiDAR with a stepwise regression procedure. [Popescu et al., 2003] studied small-footprint discrete LiDAR crown diameter measurements to estimate biomass using stepwise multiple regression to find independent variables and develop equations to relate LiDAR features to field data. Lefsky *et al.* (1999a, 1999b) investigated biomass estimation in a forest environment using wLiDAR metrics in a stepwise multiple regression. However, LiDAR-based efforts for assessment of herbaceous (grass) biomass remain scarce. Wu *et al.* (2009a and 2009b) have also developed a preprocessing approach for wLiDAR and related field data for woody and herbaceous biomass estimation. Their method was to decompose the deconvolved waveform into Gaussian approximation curves. Variables of the Gaussian fit were used in a least squares linear regression to relate to field measurements.

The approach that will be investigated in this part of the research was based on the selection of herbaceous sites where field data were collected. We propose to improve herbaceous biomass modeling by including both wLiDAR and hyperspectral features in a forward selection model (FSM) to find complementary independent variables that can model biomass. The features of interest for wLiDAR will be based on a two-peaks or fewer composite waveform for each site and metrics include the area under the curve, the rise time to the leading edge or the fall time of the trailing edge. The features that will be considered both for HSI and wLiDAR are first and second derivatives of the original bands, and PCA.

1.1 Literature Review

While HSI is a useful tool for identifying elements visible in specific regions of the electromagnetic spectrum, LiDAR can be used to quantify those elements in terms of structural metrics, e.g. height and biomass. The objectives of this research are to achieve fusion of imaging spectroscopy and wLiDAR in order to improve species-level classification and evaluate herbaceous biomass modeling as a use case. The hypothesis was that the two modalities provide complementary information that could improve per-species structural assessment, in context of species classification and biomass modeling, when compared to

single modality sensing systems. The first part of solving this issue was to define the level of fusion and examine the methods that have been used to improve results in both contexts. The second step was to study the methods used for species classification and biomass modeling without fusion, in order to achieve an understanding of the current basic techniques in both fields.

1.1.1 Data Fusion

There are three broad categories of fusion, which derive their names from the level at which the fusion takes place. Decision and feature level fusion employ advanced mathematical and statistical procedures, while pixel level fusion merges the physical parameters from different bands or sensors [Roberts et al., 2008]. In the following paragraphs, we examine pixel and feature level fusion with regards to different data types, notably LiDAR and HSI.

[Roberts et al., 2008] have fused images of multispectral advanced spaceborne thermal emission and reflection radiometer (ASTER) data and a RadarSAT- 1 SAR scene using techniques such as Brovey transformation, intensity, hue, and saturation (IHS), principal component substitution, discrete wavelet transformation (DWT), and a modified DWT approach, at the pixel level. The resulting imagery showed an increase in entropy, which can be due either to increasing noise or information, and only the frequency-based algorithms preserved the spectral information in the post-fusion images. However, classification results did not show an improvement when comparing the original multispectral image to those generated from different pixel-level fusion algorithms [Roberts et al., 2008]. An overall accuracy of 68% was achieved with the ASTER data, as opposed to less than 50% with all other images resulting from the fusion techniques mentioned above.

The complexity of the CAO data set resides in the fact that wLiDAR was composed of multiple time bins, each corresponding to a range that was recorded for an emitted laser pulse, which was tied to the vertical spatial range. However, methods used in the example above usually fuse multispectral and panchromatic imagery, where the later provides high resolution imagery, which differs from wLiDAR data in terms of dimensionality. For example, the IHS approach uses multispectral imagery from which it transforms the red, green, and blue channels to the IHS components and matches the histogram of the intensity component to the panchromatic band specified for fusion. The intensity component is then replaced with the new stretched image and the inverse-transform is performed to return to the RGB color space. Waveform LiDAR does not provide high resolution imagery like that of panchromatic bands, and thus requires a totally different approach to fuse wLiDAR and HSI at the pixel level. A viable approach would merge the spectral information of HSI to structural features provided by wLiDAR, which could be provided by performing data dimensionality reduction on the wLiDAR data.

[Mutlu et al., 2008] employed a unique approach using a combination of discrete LiDAR and multispectral imagery from QuickBird. They constructed a new multiband

image with a total of 10 bands, which was referred to as the LiDAR-Quickbird Stack. The first four bands of the new image were taken from the QuickBird data, the fifth band was the LiDAR-derived canopy cover, bands 6 to 9 are obtained from LiDAR height bins, and the last band consists of the canopy height variance. They applied a PCA and MNF to isolate noise and group most of the variance in a select number of bands. They applied the Maximum Likelihood and the Mahalanobis distance supervised classifications to the Quickbird multispectral image, on the LiDAR-Quickbird Stack, on their principal component (PC), and on their MNF images. The study area was located near Huntsville in eastern Texas and the classes of vegetation were short grass, timber, chaparral, brush, southern rough, closed timber litter, and hardwood. The best results were obtained with the LiDAR-Quickbird stack with an 88% overall accuracy, using the Mahalanobis distance classification. This method might be difficult to apply to wLiDAR because of the abundance of data compared to discrete LiDAR. This may cause the PCA and its variance to spread out across tens of bands, which does not improve the original scenario significantly.

[Sugumaran and Voss, 2007] evaluated the classification of discrete LiDAR fused with multispectral and HSI for tree species identification in the urban environment of the University of Northern Iowa campus. They geometrically and radiometrically corrected the LiDAR images and used them as a reference for processing the spectral imagery. They also performed a MNF transformation on the HSI and kept the bands containing most of the variation. They performed object-oriented classification using eCognition Professional 4.0 on spectral imagery and then added LiDAR information to each classification. The contributions made by LiDAR were elevation data that were used as a classification mask, and intensity data and shadow-free imagery, which helped in the segmentation process. The classes of vegetation used in this study were grass, oak, maple, pine, spruce, ash, honey locust, basswood, and crab apple. 81% overall accuracy was obtained for spectral imagery versus 93% for spectral imagery fused with LiDAR data.

Another technique used by [Dalponte et al., 2008] analyzed the joint effect of hyper-spectral and discrete LiDAR data for the classification of complex forest areas, more specifically the natural reserve of *Bosco della Fontana* in the Po Plain in Italy. The authors used a total of 19 different tree species, but the dominant species were *Carpinus betulus*, *Quercus cerris*, *Quercus robur*, and *Quercus rubra*. They fused data at the feature-level using a distribution-free machine learning classifier based on SVMs and a parametric regularized GML-LOOC applied to the estimation of statistics of the classifier. The result of their research indicated that the distribution-free SVMs are more accurate than the GML-LOOC procedure, resulting in accuracies over 90% for certain species. The SVM method obtained a *kappa* coefficient of 0.69 for HSI, and 0.72 when fused with LiDAR features. For the GML-LOOC method, they obtained a *kappa* of 0.63 and 0.66, respectively. They also report that the fused data set increased separability, especially when tree species have different heights.

[Sarrazin et al., 2010] have performed exploratory species-level classification using wLiDAR and HSI with data collected in the Kruger National Park (KNP) and the Bushbuckridge communal settlement, in South Africa. The authors used three different species: *Acacia nigrescens*, *Combretum apiculatum*, and *Grewia flava*. They deconvolved the wLiDAR data, corrected data for topographic variation, and proceeded to extract structural and spectral features from each modality. These features were fed into a stepwise discriminant analysis (SDA), performed with statistical analysis software (SAS), in order to select independent variables for classification. Three SDAs were performed, the first with wLiDAR associated features, the second with HSI, and the last included metrics both from wLiDAR and HSI modalities. The results indicated that the fused data set did not improve classification accuracy significantly. The SDA-based fusion resulted in a 72% overall accuracy, while the HSI and wLiDAR SDAs exhibited 59% and 68% overall accuracies, respectively. The measured *kappa* coefficients were 0.49 for wLiDAR, 0.36 for HSI, and 0.56 for both modalities.

Most of the mentioned methods were based on HSI and discrete LiDAR and reported an increase in accuracy with fused data sets compared to HSI alone, although the increase is not significant in all cases. While these studies were focused on finding the best possible classification method for fusion of spectral imagery and LiDAR data, this research investigated how fusion of HSI and wLiDAR actually improved results for two use cases when compared to single modality sensing systems. This study focused on species classification and herbaceous biomass modeling as applications.

1.1.2 Species Classification

Spectral-based classification

Multispectral data have been used in the past for forest type classifications [Franklin, 1994, Carleer and Wolff, 2004], while HSI is a more recent modality that might provide a better source of information to this end. Imaging spectrometers measure the interactions between matter and electromagnetic radiation, typically in the spectral range of 400 to 2500nm [Ustin et al., 2004]. Multispectral imaging normally describes the collection of data over one to ten spectral channels. On the other hand, HSI is generally acquired with tens to hundreds of bands [Schott, 2007]. [Ustin et al., 2004] described the difference between multispectral and hyperspectral imagery collected over the same spectral range (Figure 1.1). Figure 1.1(a) and 1.1(b) show more details as more samples were collected over the same range. Figure 1.1(c) and 1.1(d) show how different spectra appear when only six bands of data are collected from the Landsat multispectral imager.

Figure 1.1: (a) Full spectra associated with tree species with green foliage: black cotton-wood (*Populus trichocarpa* - POTR), red alder (*Alnus rubra* - ALRU), Douglas fir (*Pseudotsuga menziesii* - PSME), and western hemlock (*Tsuga heterophylla* - TSHE). (b) Full spectra associated with soil and three types of plants : epiphytic lichen (*Isotheicum stoloniferum* - ISST), bark (PSME), and dry grass. (c) Landsat six-band spectra of the same species shown in (a). (d) Landsat six-band spectra of the same cover types shown in (b) (image courtesy of Ustin *et al.*, 2004).

Different methods for classification of multispectral imagery and HSI are discussed in the literature. As described by [Schott, 2007], the broad categories of classification are supervised and unsupervised classification, where supervised classification allows the user to identify a sample of pixels for each class, after which an algorithm is used to assign similar pixels to corresponding classes. Some of the frequently used algorithms for multispectral images include multivariate minimum distance to the mean, parallelepiped classifiers, and multispectral Gaussian maximum likelihood. As for unsupervised classification, an algorithm determines which pixels exhibit similar characteristics and groups them into classes. The most frequently used algorithms are k-means and iterative self-organizing data, most commonly referred to as ISODATA. Other classification tools revolve around the knowledge of spectral signatures of specific materials in a scene, such as the spectral matched filter (SMF), the SAM, and spectral mixture analysis (SMA).

Dimensionality reduction is required to handle multiple bands efficiently, especially in the case of HSI. Frequently used methods include PCA and MNF transforms, which constrain most of the variability in a few bands, while lower-order bands contain noise and can be discarded [Schott, 2007]. Statistical methods such as the SDA, on the other hand, reduce dimensionality by selecting optimal and independent bands. [Underwood et al., 2003] have used three different methods to reduce dimensionality of AVIRIS data for detection of invasive plant species such as iceplant (*Carpobrotus edulis*) and jubata grass (*Cortaderia jubata*). The study site was in Vandenberg Air Force Base along the central coast of California. They first atmospherically corrected the imagery from radiance to reflectance, then applied georeferencing and denoising tools to the data and defined regions of interest (ROIs) for training purposes, using the ENVI software. They investigated MNF, to which they applied ML supervised classification, and continuum removal (CR), specific to the 932nm water absorption feature. This was followed by a K-means unsupervised classification, and derivation of band ratio indices (BRI), to which they applied the ML classification. Superior overall classification was obtained with the MNF method, with 55% overall accuracy, compared to 39% with CR, and 44% with BRI.

Traditional forest-type classification studies have been performed using multispectral imagery [Franklin, 1994, Schiever and Congalton, 1995]. Frequently used instruments include Landsat Thematic Mapper (TM), Multispectral Scanner (MSS), IKONOS, and the Advanced Very High Resolution Radiometer (AVHRR) [Martin et al., 1998]. As an example, [Carleer and Wolff, 2004] performed a supervised maximum likelihood classification of very high resolution multispectral imagery in order to classify materials in a scene: oak (*Quercus*), scots pine (*Pinus sylvestris*), corsican pine, larch (*Larix decidua*), young beech (*Fagus sylvatica*), old beech, purple beech, Douglas fir (*Pseudotsuga menziesii*), lawn and clearing, road, and water. Their study was performed in the Sonian forest in the southeastern part of Brussels, where two IKONOS images were acquired: one in June and the other in October 2000. They first performed PCA on different groups of bands of the multispectral imagery. Then they applied the normalized difference vegetation index

(NDVI), which is commonly used to highlight vegetation. The last step was the application of a mean filter to all the spectral bands, the PCA, and the NDVI images, which has a smoothing effect and increases class separability by decreasing interclass variability [Foody and Hill, 1996]. The classifier was applied to different combinations of bands, but the best result was obtained with the red, green, blue, near-infrared, PC_1 , and the NDVI bands. The overall classification of the filtered image was 86% with a *kappa* coefficient 0.84, which was considerably better than the classification of the non-filtered image at a 79% accuracy and a *kappa* coefficient of 0.76.

Another example of forest-type classification using hyperspectral data was done by [van Aardt and Wynne, 2001], who used spectroradiometer data (350-2500nm) in order to investigate the spectral separability between six tree species in the Appomattox Buckingham State Forest, Virginia (loblolly pine (*Pinus taeda*), Virginia pine (*Pinus virginiana*), short-leaf pine (*Pinus echinata*), scarlet oak (*Quercus coccinea*), white oak (*Quercus alba*), and yellow poplar (*Liriodendron tulipifera*)). First, data were smoothed using three different types of filters (moving 9-point weighed average, moving 9-point median, and static 10-point average) and then first and second difference operators were employed to describe curve shape. They used SAS to perform a stepwise discriminant procedure to reduce the number of independent variables between the raw reflectance data, a 9-point weighted moving average filter, a 9-point moving median filter, a 10-point static average filter, a Landsat simulation data set, and the first and second derivatives of each of previously mentioned data sets. The independent variables were used as input for a normal discriminant and a canonical discriminant analysis to test the between- and within species separability. The results indicated that cross-validation accuracies varied between 62% and 84% for second difference of raw relative reflectance and smoothed relative reflectance, respectively.

In order to determine if airborne AVIRIS hyperspectral data provide an advantage over canopy-level spectroradiometer data, [van Aardt and Wynne, 2007] then investigated species separability in Appomattox Buckingham State Forest, Virginia using airborne data. The AVIRIS sensor (224 bands) collected data from which pine locations were extracted using 1-pixel and 3x3-pixel windows, for two different analyses. SAS was used to perform the SDA in order to reduce dimensionality to 10 bands. These bands were used as inputs to the linear discriminant and canonical discriminant procedures to evaluate between and within class separability. It was found that the most significant bands were located in the visible and near-infrared regions of the spectrum. The 3x3 neighborhood samples resulted in better accuracies than single pixel data, which was attributed to better sampling and smoothing of the data. The overall cross-validation accuracy was 85% for 3x3-pixel windows, which showed similar results to those in [van Aardt and Wynne, 2001].

Another example of hyperspectral imagery in the classification context is [Gong et al., 1997], which investigated a discriminant analysis and an artificial neural network (NN) using high resolution spectroradiometer data (250-1050nm) to evaluate spectral separability of six coniferous and one hardwood species respectively: sugar pine (*Pinus lambertiana*),

ponderosa pine (*Pinus ponderosa*), white fir (*Abies concolor*), Douglas fir (*Pseudotsuga menziesii*), incense cedar (*Calocedrus decurrens*), giant sequoia (*Sequoiadendron giganteum*) and California black oak (*Quercus kelloggii*). The study site was located at the Blodgett Forest Research Station of the University of California. The feed-forward neural networks algorithm was used along with the error-propagation algorithm as training mechanism [Rumelhart et al., 1986]. They used SAS to implement the linear discriminant analysis to test conifer species identification. It was found that NN achieved a better overall performance than discriminant analysis when used with the same number of training and test samples. Using sunlit samples alone produced 91% overall accuracy. They also found that using average spectra for each tree species may not result in better classification than using individual spectra.

A totally different method was employed to understand the spectral separability between species, by [Clark et al., 2005]. They have used HSI for automated species classification of individual tree crowns in a tropical forest in Costa Rica. Their study involved seven different tree species: *Balizia elegans*, *Ceiba pentandra*, *Dipteryx panamensis*, *Hyeronima alchomeoides*, *Hymenolobium mesoamericanum*, *Lecythis ampla*, and *Terminalia oblonga*. The hyperspectral digital imagery collection experiment (HYDICE) reflectance spectra were used to simulate IKONOS, Landsat ETM+, and ASTER imagery and the spatial resolution was kept constant for all data. The sensor's spectral range varied from 437 to 2434nm. Three different classification methods were used, including LDA, maximum likelihood, and SAM. Combinations of bands were chosen based on a stepwise selection procedure and used as inputs in the three classifiers. The best performing crown scale classification hinged on a selection of 30 bands with the LDA, resulting in 92% overall accuracy, while the SAM classifier performed the worse. They also performed object-based tree crown classification and results indicated that the narrow band hyperspectral HYDICE sensor performed better than multispectral broadband sensors, although this type of classification did not perform as well as crown-scale spectra classification.

[Martin et al., 1998] have investigated the application of high resolution HSI for identification of species composition. The study site was situated in the Harvard Forest, in central Massachusetts, and the 11 species investigated were red maple (*Acer rubrum*), red oak (*Quercus rubra*), white pine (*Pinus strobus*), hemlock (*Tsuga canadensis*), Norway spruce (*Picea abies*), red pine (*Pinus resinosa*), spruce bog, mixed deciduous, deciduous/conifer mix, mixed conifer, and deciduous bog. Data were acquired with 224-bands AVIRIS (400-2400nm) on which ATmosphere REMoval (ATREM) correction was performed. First, they reduced the number of bands to 11 based on known spectral regions corresponding with chemical foliage composition, based on previous multiple linear regression analysis [Martin and Aber, 1997]. They applied the ML classification and used field data such as stem diameter and field measurements of plot level foliar biomass based on stand-level survey, for validation purposes. The overall classification accuracy achieved was 75% for a selection of random validation pixels.

[Cho et al., 2009] have evaluated the performance of the SAM classifier on HSI based on single and multiple endmembers per species, acquired from the CAO in the KNP, South Africa, in May 2008. The data consisted of 72-bands from a high-fidelity imaging spectrometer (HFIS), discrete LiDAR, and GPS inertial measurement unit (IMU). The study site consists of a savanna environment and spectra from the following tree species were collected: *Acacia nigrescens*, *Combretum apiculatum*, *Combretum imberbe*, *Dichrostachys cinerea*, *Euclea natalensis*, *Gymnosporia buxifolia*, *Lonchocarpus capassa*, *Pterocarpus rotundifolius*, *Sclerocarya birrea*, and *Terminalia sericea*. First, they built a tree-mask from the LiDAR canopy height model (CHM), performed with the ROI tool in the ENVI software, which was super-imposed on the HSI. A bootstrapping method was used to obtain training and test data. They used two approaches to generate input into the SAM classifier. First, they used the mean spectrum of the training data for each species and the second approach was to include all the training spectra for each species, as the multiple-endmember approach. The overall classification accuracy for the single endmember approach was 54%, compared to 20% for the multiple endmember approach. They found that band selection improved the classification results and that tree spectra should really be representative of each species in order to perform well, as opposed to results found by [Gong et al., 1997].

Many techniques exist to manipulate multispectral and hyperspectral imagery, from data dimensionality reduction to classification methods, for the many applications within forest-type classification studies. This type of imagery senses the composition of vegetation and thus different spectral metrics can be used to differentiate between tree species. For this research, PCA will play a particularly important role since it groups most of the variance in a fewer number of bands and thus reduces data dimensionality. Such metrics will serve as inputs in the feature fusion approach, to discriminate between species.

LiDAR-based classification

LiDAR has been in use for nearly two decades and many studies have shown that tree horizontal and vertical structural information can be successfully retrieved from it. As early as 1985, the Canadian Forestry Service demonstrated LiDAR's potential for obtaining stand heights, crown cover density and ground elevation below canopies [Aldred and Bonner, 1985, Lim et al., 2003]. [Hyde et al., 2005] were successful in using LiDAR to measure height and biomass at the stand-level. [Lim et al., 2003] affirmed that canopy height, sub-canopy topography, and vertical distributions of canopies can be directly retrieved from LiDAR data. Using empirical methods, many attributes can be predicted from this structural data source, such as above ground biomass, basal area, mean stem diameter, canopy volume, and vertical foliar profiles [Means et al., 1999, Lefsky et al., 1999b, Lefsky et al., 1999a, Dubayah and Drake, 2000]. LiDAR was also used for terrain modeling, which provides the user with three dimensional data point clouds [Flood and Gutelius, 1997]. However, very few studies have investigated the use of these structural metrics to perform

classification, especially species-level classification.

[Holmgren and Persson, 2004] have used discrete LiDAR to discriminate between pine and spruce in the Remningstorp area in Sweden, which is a difficult task even with HSI, since both species have similar spectral signatures. They used field data for training and validation purposes. They extracted features based on height distributions, geometry, and intensity and grouped them in pairs exhibiting the largest correlation coefficients. Student's t-test was implemented to evaluate the most valuable metrics, which were used both in classical linear and quadratic discriminant functions. The overall classification was 95%, which indicates successful discrimination between pine and spruce. They also evaluated crown base height to field data and obtained a high correlation of $r = 0.84$.

[Reitberger et al., 2006] used small-footprint wLiDAR to classify deciduous and coniferous trees from data collected in the Bavarian Forest National Park. They first decomposed the waveforms into Gaussian approximation curves and extracted three dimensional points and attributes from the waveforms. They proposed the extraction of salient features based on waveform intensity, crown shape, internal geometrical structure, height, and density of trees. For classification, they used a two-step procedure. First they clustered tree species and then applied the Bayes classification using those features. This resulted in 80% overall accuracy in a leaf-on environment.

[Neuenschwander et al., 2009] have also used wLiDAR for landcover classification of central Texas vegetation such as dead trees, juvenile juniper, brush, trail, KR Bluestem, mature juniper, and oaks. The motivation behind this study was to find a feature, other than height, useful for species classification. They decomposed the waveforms into Gaussian approximation curves and extracted parameters such as peak values and standard deviations. Other metrics included canopy and ground energy, canopy to ground energy ratio, total waveform energy, height of median energy, rise time to the first peak, and fall time of the last peak. A Bayesian pairwise classifier was chosen for its ability to discriminate between pairs of classes. As such, the most valuable metrics were found to be the energy ratio and Gaussian amplitude. The overall classification accuracy for wLiDAR was 86%, compared to 71% with Quickbird imagery.

There were no studies that evaluated species classification outside of a forestry context. As such, this study contributes by not only evaluating species differentiation in the complex savanna environment, but by doing so based on unique modalities. While HSI reveals many important aspects about chemical composition of each tree or plant species, waveforms are rich in structural information. Previous studies have identified different features that can be useful to extract and fuse with HSI, with the expectation that they would complement each other. The approach of interest for this project relied on the formation of composite waveforms from single canopies, which simulated a large-footprint LiDAR for each tree. Features were extracted from each composite waveform, such as tree height and crown thickness. These metrics, along with hyperspectral features such as PCA values for each canopy, were used as input into a stepwise procedure for data dimensionality reduction.

Then, the output independent variables will be fed into a linear discriminant analysis that will identify the most valuable features for classification.

1.1.3 LiDAR-based Biomass Modeling

The most obvious method to measure forest biomass is to conduct field-based inventory, but this approach is very costly and time-consuming [van Aardt et al., 2006]. In the past, aerial photography volume tables have been used, as described in Avery and Burkhart (1994). Also, spectral imagery has mostly employed empirical methods to estimate biomass that usually relate intensity of electromagnetic radiation to biomass or use the ratio of the energy at different wavelengths [Steininger, 2000, Drake et al., 2002, Mirik et al., 2005a, Mirik et al., 2005b]. [Mirik et al., 2005a, Mirik et al., 2005b] have investigated the utility of a high-resolution hyperspectral sensor to estimate total, live, and individual biomass of matter found in grassland and riparian regions of Yellowstone National Park, Wyoming. The PROBE-1 hyperspectral sensor was used to collect imagery at $1m^2$ resolution. Field data were collected in 56 plots of $1m^2$ (matching the hyperspectral sensor), that represented the range of biomass values across the study area and included litter, grasses, forbs, sedges, sagebrush, and willow. They investigated a large number of SRTVI and NDTVI that were applied to the HSI and used in a regression-type analysis against field data. The results show a strong correlation between SRTVI or NDTVI and the total biomass with R^2 of 0.87 and individual biomass of live, sedge, and willow ranging from R^2 of 0.66 to 0.84. Weak relationships, with R^2 varying from 0.16 to 0.51, were found for the other types of matter, possibly due to spectral signature mixtures. So a variety of biomass components were correctly modeled and showed high correlation with spectral measures, although not all components shared the same success.

Nowadays, LiDAR is viewed as the most efficient method to collect structural data from large areas, which is very appropriate for volume and biomass estimation, especially in forest inventory applications [Lefsky et al., 1999b, Popescu et al., 2004], fuel loads estimations [Seielstad and Queen, 2003], and carbon flux estimations related to natural and anthropogenic processes [Cohen and Spies, 1992, Drake et al., 2002]. For such applications, discrete LiDAR has been used in the past [Popescu et al., 2004, Zhao et al., 2009], however a more recent technology, wLiDAR, has enabled us to capture volume and biomass at much smaller scales [Wu et al., 2009a, Wu et al., 2009b, Means et al., 1999].

[Zhao et al., 2009] have investigated scale-invariant estimation methods for forest biomass assessment with discrete LiDAR in a temperate forest of eastern Texas. The field data that was collected consists of tree species such as loblolly pines (*Pinus taeda*), and deciduous trees like water oak (*Quercus nigra*), red oak (*Quercus falcata*), sweetgum (*Liquidambar styraciflua*), and post oak (*Quercus stellata*). They used the linear functional model and an equivalent nonlinear model with canopy height distributions (CHD) and canopy height quantile functions (CHQ), respectively, as predictors. The CHD was defined as “the probability of observing the top canopy surface at height h ” [Zhao et al.,

2009]. They created a fine resolution tree biomass map by implementing a tree-delineation approach, which served as training and test data to validate the two scale-invariant models. In order to do this, they related biomass to LiDAR by first estimating the diameter at breast height (DBH) of each tree with regression analysis and then they applied allometric equations to derive biomass. They used a sample size of about 50 plots to obtain a good fit of the linear model. To assess the results of the linear and nonlinear models, they built scatterplots relating the above ground biomass (AGB) and obtained an almost perfect match. They were able to demonstrate that biomass can be accurately predicted with an R^2 of 0.8 to 0.95 between models and that mapping biomass is possible at different scales, by comparing the CHD and CHQ results to the training data.

[Drake et al., 2002] have studied the relationship between canopy structure and above ground biomass using field and large-footprint wLiDAR data from the laser vegetation imaging sensor (LVIS) airborne instrument, acquired in La Selva Biological Station, which is a tropical forest research facility in northeast Costa Rica. Data from three land cover types were collected: 18 primary forest plots, 3 secondary forest plots, and 6 agroforestry plots. All plots had annual collection of measurements of stem DBH greater than 10cm. First, they built plot-level “average waveforms”, corrected them for attenuation of the LiDAR energy as the laser pulse travels through canopy, and calculated canopy height profiles (CHPs) based on techniques described by [Harding et al., 2001]. They used a stepwise regression procedure to relate vertical canopy profiles derived from field measurements, to the estimated above ground biomass (EAGB) and found a high correlation R^2 of up to 0.94. In order to find if LiDAR also exhibited this type of correlation with EAGB, the untransformed waveforms were converted to bin profiles and normalized to compensate for difference in units. They also found a high correlation between vertical canopy profile metrics of the LiDAR data and those of field profiles.

[Popescu et al., 2003], among others, have studied the influence of small-footprint discrete LiDAR crown diameter measurements for estimating tree biomass in the Piedmont physiographic area, in Virginia. They collected field data from six types of trees such as pine hardwoods, upland hardwoods, bottomland hardwoods, stands of Loblolly pine, Virginia pine, and short leaf pine. Discrete LiDAR data were collected in addition to a high-resolution leaf-off image, acquired with the multispectral optical airborne terrestrial land application scanner (ATLAS) imagery. First, they identified individual trees by using LiDAR and multispectral imagery using a local filtering method. They calculated the crown diameter as the mean of two measurements along perpendicular directions from each tree top by fitting a polynomial function to each profile. They performed stepwise multiple regression to find independent variables for the model and developed equations to relate LiDAR features to field data. Regression analysis was also performed with and without independent features to estimate plot level biomass. The estimation of crown diameters of deciduous and pine trees was similar with an R^2 of approximately 0.63 for significant trees. However, they found that crown diameter measurements improved the

R^2 values for biomass by up to 0.25 for both pine and deciduous plots.

Lefsky *et al.* (1999a) have studied wLiDAR on 22 plots of Douglas-fir western hemlock stands in the Cascades Range, Oregon. They classified each plot as being young, mature, or old-growth based on age and condition from the Willamette National Forest and a classified TM image. The scanning LiDAR imager of canopies by echo recovery (SLICER) sensor collected wLiDAR and georeferencing was accomplished with the inertial navigation system mounted to the sensor. They used independent waveform metrics such as the CHP and the canopy surface height (CSH) in a stepwise multiple regression, resulting in non-asymptotic predictions of biomass that explained 90% of the variance. The CSH was calculated from the first return, which normally corresponds to the top of canopy and the mean elevation of the ground return. The CHP is a modification of the foliage height profile, which quantifies the distribution of foliage surface area as a function of height. The CHP therefore is a distribution of foliar and woody surface area as a function of height, since wLiDAR cannot distinguish between woody and foliage surfaces. They also included a novel technique in the regression model, called the canopy volume method, which is a volumetric measure of multiple contiguous waveforms in order to describe vegetation structure. They were able to make accurate estimates of the mean and standard deviation of DBH, the number of stems, and basal area estimates.

Lefsky *et al.* (1999b) have investigated methods to predict basal area and above ground biomass using wLiDAR in the coastal plain of Maryland. Two data sets were acquired, the stem map and the chronosequence, from which basal area and above ground biomass were derived from field measurements of DBH. The CHPs for the stem map were obtained with the SLICER sensor, while the chronosequence data set consisted of observation plots in the Smithsonian Environmental Research Center, in Maryland, for which the CHPs were measured in the field. Four different height features were extracted from the CHPs: maximum, median, mean, and quadratic mean canopy height (QMCH). From the chronosequence data set, regression equations were developed and applied to height features calculated from the stem map data set. The regression equations were able to explain most of the variance in the QMCH for the chronosequence data set and most accurately predicted the same stand structure attributes in the stem map data set. For the chronosequence data set, the QMCH explained 70% of the variance for basal area, and 80% of the variance of the above ground biomass. When applied to the stem map data set, the regression equations explained 37% of the variance for basal area and 33% for the biomass. The difference in the coefficients of determination reside in the fact that a greater range of stands was found in the chronosequence data set. They also found that stepwise multiple regression did not improve the accuracy of the two stand structure estimates. While these studies focused on forest environments, examples in savannas and herbaceous biomass remain scarce.

Wu *et al.* (2009a) have investigated the relation between wLiDAR and herbaceous biomass by implementing a structural and statistical feature metric method. The study area was situated in the KNP, in South Africa, and wLiDAR data were acquired by

the CAO in 2008. Field data were collected at 36 sites where herbaceous biomass was weighed within a $0.25m^2$ grid. Their method was based on the decomposition of single deconvolved waveforms into Gaussian approximation curves by using the expectation-maximization (EM) algorithm. Variables of the Gaussian fit were first analyzed to ensure highly correlated independent coefficients were not included in the model. The least squares linear regression was implemented and a definite correlation between wLiDAR and herbaceous biomass was found with an R^2 of 0.59.

Wu *et al.* (2009b) have also developed a signal processing approach for wLiDAR for tree and foliar biomass estimation. The study area was the same as for Wu *et al.* (2009a) and the field data that were used for this research consisted of diameter above basal swelling (D) and height (H) for all trees in specific plot areas, which were used in allometric equations to predict woody biomass. Tree species for this research included *Acacia nigrescens*, *Sclerocarya birrea*, *Terminalia sericea*, *Dichrostachys cinerea*, and *Combretum collinum*. They first deconvolved the waveforms in order to recover the true response distribution of targeted vegetation, as described by [Wagner et al., 2006]. They proposed feature metric extraction from the waveforms such as tree height and crown thickness. Variables were analyzed to ensure highly correlated independent features were not included in the model in order to increase robustness. A least squares linear regression was used to model woody and foliar biomass. The results show a high correlation between the feature metrics and the field data with $R^2 = 0.73$ for foliar biomass estimation and $R^2 = 0.71$ for woody biomass estimation. The height estimation was also compared to field data and results showed an $R^2 = 0.92$, i.e. high correlation with field data.

This research will investigate herbaceous biomass modeling by including wLiDAR and hyperspectral data in a FSM. Field biomass data were gathered at the plot-level, each of which will be represented by a composite waveform. The wLiDAR data will have previously been denoised, deconvolved, and angle corrected so that parts of waveforms that were affected by the off-nadir collection of data are attributed to appropriate voxels. The composite waveforms will be decomposed into Gaussian approximation curves and features attributed to each curve will be input in the model, such as the width, the area under the curve, the rise time to the leading edge, and fall time of the trailing edge. The features that will be considered both for HSI and wLiDAR are first and second derivatives of the original bands, and PCA. Those features will be input into a stepwise selection model to find complementary independent variables, which will be used to model herbaceous biomass. This research evaluates the possibility of increasing biomass accuracy by including hyperspectral data and waveform structural features, which we assume to be two complementary modalities.

Chapter 2

Methodology

This chapter describes the study site and the data collected for this research, the remote sensing instruments used, the signal preprocessing, and the techniques to fuse HSI and wLiDAR at the feature level. One technique relates to tree species classification, which requires segmentation of individual tree canopies, followed by the building of composite waveforms, which effectively simulate a large footprint LiDAR for each tree. Finally, it is possible to extract feature metrics, such as tree height and canopy thickness. The features that will be considered both for HSI and wLiDAR are rst and second derivatives of the original bands, and PCA. These variables will be used in a SDA, implemented in SAS 9.2, which will optimize variable selection with the stepwise selection model for input into a linear discriminant procedure. The original data will be used separately from the PCA data to make sure no correlated data are input in the stepwise regression procedure. The final classification should identify each canopy class and will be evaluated against field data. Other measures of performance will be the producer's and user's accuracies, the *kappa* coefficient, and the z-score, which is based on the proportion of trees correctly classified. The data employed for this part are the compact airborne spectrographic imager (CASI) HSI and LiDAR data from the KNP and the Bushbuckridge communal settlement.

The second part to this research is herbaceous biomass modeling, which requires selection of herbaceous plots, building composite waveforms, decomposition of the deconvolved waveform into Gaussian approximation curves and feature metric extraction based on the Gaussian fit. These variables will be used in a FSM, which will optimize variable selection for biomass modeling. Only waveforms with a maximum of two peaks after deconvolution will be studied since we suppose they are representative of herbaceous and ground information. Variables of interest for biomass modeling are the width, the area under the curve, the rise time to the leading edge, and fall time of the trailing edge. The features that will be considered both for HSI and wLiDAR are rst and second derivatives of the original bands, and PCA. SAS will be used for the forward selection process which will relate herbaceous biomass field data to feature metrics. The coefficient of determination

(R^2) and root mean square error (RMSE) will be calculated in order to measure model performance. This method will use field data from the KNP and the Bushbuckridge communal settlement, along with CASI HSI and LiDAR data, but includes different land uses than the species classification section and those used by [Wu et al., 2009a].

2.1 Data Collection

Data were collected at three different sites in South Africa bounded by ($22^{\circ}8'00''S$; $30^{\circ}34'52''E$) and ($25^{\circ}32'48''S$; $32^{\circ}2'50''E$) that include the KNP, the Sabie Sands game reserve, which is a private conservation area, and the Bushbuckridge communal settlement, which consists of a densely populated area (Figure 2.1) [Wu et al., 2009a]. The area of study is considered a savanna environment, which is mainly characterized by a near-continuous grass layer and various trees and plant species [Archer, 1995, Scholes, 1997]. The airborne data that were collected consist of hyperspectral imagery, waveform, and discrete LiDAR, although wLiDAR was only collected in portions of the KNP and in the communal settlements.

Figure 2.1: a) Location of the KNP in South Africa and b) land uses (west-to-east) from heavily exploited range and farmland, a private game reserve, the KNP, and Skukuza flux-tower site.

Figure 2.2: Site-level sampling design with 36 plots/site [Wu et al., 2009a].

Field data were collected in eight land uses throughout regions the KNP, the Sabie Sands game reserve, and the Bushbuckridge communal settlement. Sites consisted 36 plots in the form of an “E” of 50 x 50m in size, as seen in Figure 2.2, but only 12 sites contained wLiDAR data. HSI was collected in all three land use categories, but wLiDAR was only collected in the two land uses that span the Bushbuckridge communal settlement and the KNP. Herbaceous biomass was weighed within a 0.5 x 0.5m grid and woody biomass and canopy density were estimated for each plot, all at a 10m spacing between-plot [Wu et al., 2009a]. Large and interesting trees were also identified in each land use: spectra were recorded along with stem size, DBH, tree, and canopy height. Each plot and identified tree were associated with a GPS location recorded with a Leica (GS20 Professional Data Mapper with handheld aerial) or Trimble (Recon Handheld with aerial backpack) differential GPS instrument.

The species of interest for the species classification study along with the number of GPS points acquired were: *Combretum hereroense* (Cher-109), *Peltophorum africanum* (Pafr-29), *Sclerocarya birrea* (Sbir-119), and *Terminalia sericea* (Tser-49). The choice of these particular species was based on the amount of specimens available on site for each species along with the feasibility of segmenting each tree using the HSI. The specific data (12 sites) used for tree species classification originated from two area of interest in Bushbuckridge communal settlement and a single area from KNP. As for herbaceous biomass modeling, a total of 319 plots were originally used from both land areas in Bushbuckridge communal settlement and the area in KNP. However, this did not correspond exactly to the same specific regions as for species classification, given to the need for distinguishable trees.

2.2 Remote Sensing Instruments

The CAO-Alpha system consists of a visible/near-infrared (VNIR) imaging spectrometer and a scanning wLiDAR, along with an integrated navigation and data processing system. The result is a geo-rectified product for evaluation of vegetation structure, physiology, and the underlying topography [Asner et al., 2007]. The imaging spectrometer has a

pushbroom design that can achieve a wide swath of 1500 pixels and a spatial resolution of 0.1-1.5m, necessary to resolve individual plant canopies and vegetation clusters. Discrete LiDAR data were also collected with this system. The platform can revert to the CAO-Beta system (AVIRIS sensor), which is intended to fly larger regions at 3-4m spatial resolution [CAO, 2008].

HSI can achieve high spectral resolutions, which is useful to estimate physical parameters such as biomass volume and chlorophyll concentration. It basically provides a detailed description of the spectral signatures of classes [Asner et al., 2007]. The CASI sensor was used for the first part of this project. The imaging spectrometer has 24 bands and a spectral range of 373 to 1052nm: band 1 in the visible violet (394nm), bands 2-4 in the visible blue (423-479nm), bands 5-7 in the visible green (508-564nm), band 8 in the visible orange (593nm), bands 9-13 in the visible red (621-734nm), and bands 14 to 24 in the near infrared (763-1045nm). This range was specifically chosen considering the vegetation absorption features, which will in turn facilitate the classification. The instantaneous field of view is 0.56mrad to match the LiDAR specifications and footprint.

The LiDAR instrument is basically equipped with a transmitter, from which the laser is emitted, and a receiver to capture the reflected signal from the ground. LiDAR systems are typically categorized as “discrete” or “full waveform” systems, which differ in how they vertically and horizontally sample a target’s three-dimensional structure. The vertical sampling for wLiDAR is much higher (1ns) than discrete LiDAR, and generates more returns that typically describe the vertical distribution of objects in space. This can be viewed as a function, where the intensity of the signal is a function of time, as can be seen in Figure 2.3. Waveform LiDAR systems digitize the entire target backscattered signal at system-dependent intervals, while discrete LiDAR only registers returns above a specific intensity threshold, and then only after approximately 6.7ns (return trip time) have elapsed since a previous signal detection, up to a maximum of four returns per pulse. Pulses are typically classified as first, second, and up to four returns. Those returns that hit the lower structure (ground) of a scene are used to derive the digital elevation model (DEM), while the hits from the top structures of a forest scene form the CHM.

Waveform LiDAR is a potentially useful tool for finding tree height, forest structure, mean stem diameter, volume and biomass estimates, and other parameters related to the vertical distribution of a particular scene under analysis, given that the vertical sampling is high enough. For example, it can be used to distinguish between coniferous and deciduous trees [Dalponte et al., 2008]. In this specific research, the CAO-Alpha system employed the Airborne Laser Terrain Mapper (ALTM) 3100 wLiDAR in order to describe the three dimensional vegetation structure of South Africa savanna systems. Specifications of the LiDAR system are included in Table 2.1 [Mallet and Bretar, 2008]. Careful choices were made in order to retrieve the most useful information from the ecosystem, for example the infrared wavelength (1064nm) was chosen for its high reflectivity from vegetation targets.

The standard LiDAR equation is shown in equation 2.1. The equation takes into ac-

Figure adapted from Blair, Harding, NASA/GSFC

Figure 2.3: Waveform LiDAR data acquisition (image courtesy of Blair Harding, NASA GSFC).

count that targets are distributed in space and that the reflected signal is the superposition of echoes at different distances. The received power, P_r , is given by

$$P_r(t) = \frac{D^2}{4\pi\lambda^2} \int_0^H \frac{\eta_{sys} \eta_{atm}}{R^4} P_t \left(t - \frac{2R}{v} \right) \sigma(R) dR \quad (2.1)$$

where t is the time, D the aperture diameter of the receiver optics, λ the wavelength, H the flying height, η_{atm} and η_{sys} are the atmospheric and system transmission factors, respectively, R the distance from the system to the target, P_t the emitted power, v the velocity of the laser pulse, and $\sigma(R) dR$ the apparent effective differential cross-section. We say *apparent*, since an object reflecting the signal at a certain distance can occlude another object further in the laser path [Mallet and Bretar, 2008].

The power of the received signal can also be considered as the sum of the contribution of N targets along with each of their characteristics,

$$P_r(t) = \sum_{i=1}^N P_{r,i}(t) * \eta_{sys}(t) * \eta_{atm}(t) \quad (2.2)$$

where $P_{r,i}(t)$ is the echo of the i^{th} object and is given by

Table 2.1: Design specifications of the ALTM 3100 for wLiDAR.

Description	Specification
Company manufacturer	Optech
Beam deflection	Oscillating mirror
Wavelength	1064nm
Flying height	$\leq 1km$
Number of bands	256ns
Pulse rate	$\leq 70kHz$
Pulse energy	$< 0.2mJ$
Pulse width	15ns
Scan rate	$< 70Hz$
Scan angle	50 deg
Beam divergence	0.56mrad
Footprint size	0.56m (at 1000m above ground level)
Range accuracy	1cm
Digitizer	1ns

$$P_{r,i}(t) = \frac{D^2}{4\pi\lambda^2} \int_{R_i-\Delta R}^{R_i+\Delta R} \frac{1}{R^4} P(t) \left(t - \frac{2R}{v} \right) \sigma_i(R) dR. \quad (2.3)$$

Here, R_i is the mean distance, the integral interval $[R_i - \Delta R, R_i + \Delta R]$ represents the spatial spread, and $\sigma_i(R)$ is the effective differential backscattering cross-section. So when $\Delta R \ll R$, we get

$$P_{r,i}(t) = \frac{D^2}{4\pi\lambda^2 R_i^4} P_t(t) * \sigma'_i(t) \quad (2.4)$$

where $\sigma'_i(t)$ is the apparent cross-section of illuminated areas within each range interval. The power of the received signal [Mallet and Bretar, 2008] can thus be expressed as

$$P_r(t) = \sum_{i=1}^N \frac{D^2}{4\pi\lambda^2 R_i^4} P_t(t) * \eta_{sys}(t) * \eta_{atm}(t) * \sigma_i(t). \quad (2.5)$$

2.3 Signal Preprocessing

Features of the raw incoming waveform depend on the sensor's variable outgoing pulse signal, the receiver impulse response, and the system noise. In order to obtain the true response distribution of specific interactions along the LiDAR footprint, signal preprocessing is necessary. We applied the steps described in Wu *et al.* (2009a and 2009b).

Noise is usually present in the form of high frequency components of the raw waveform domain and is thus removed by setting a cut-off frequency (0.5 *cycles/mm*) in the frequency domain, followed by a conversion back to the time domain. The noise-filtered waveform, represented in a simplified version of equation 2.5, is

$$P_r(t) = P_t(t) * \sigma(t) * \Gamma(t) \quad (2.6)$$

where $P_t(t)$ is the outgoing waveform, $\sigma(t)$ represents the cross-section or the true response distribution of the target, and $\Gamma(t)$ is the receiver impulse response. We deconvolved the incoming waveform from the outgoing waveform and receiver impulse response in order to find the true response. The Richardson-Lucy algorithm was applied to deconvolve the system impulse response from the incoming waveform to obtain a better estimate of the target [Easton, 2008]. The mathematical solution for $\sigma(t)$ is

$$\hat{\sigma}_{i+1}(t) = \hat{\sigma}_i(t) \left[\frac{P_r(t)}{h(t) * \hat{\sigma}_i(t)} * h(t) \right] \quad (2.7)$$

where $h(t) = p_t(t) * \Gamma(t)$. The residuals for each iteration i is defined by

$$r_i(t) = P_r(t) - h(t) * \hat{\sigma}_i(t) \quad (2.8)$$

and will converge as the iteration progresses.

The Richardson-Lucy algorithm [Lucy, 1974] was applied to find the true target response distribution, while also enhancing the vertical signal resolution [Wu et al., 2009a]. A correction was required to ensure that each waveform time bin is associated with its respective vertical height or voxel, since the waveforms were collected by a line scanner, effectively allowing off-nadir scan angles and waveforms crossing multiple pixels. The waveforms were also aligned at the ground peak location, meaning that a certain waveform feature was assigned to be ground for all waveforms, after which waveforms were aligned with respect to that ground feature. This effectively normalizes for ground elevation differences. The timescale for each time bin can easily be attributed to height, since we know each pulse travels at the speed of light ($3 \times 10^8 m/s$) and each detection time bin lasts $1ns$. Thus each time stamp of $1ns$ represents $0.15m$ in space. The time bins above the ground bin are assigned to targets above ground, while it is believed the time bin after the ground bin include multiple scattering information and potential system noise. The ground peak of each waveform was determined to be the last peak, while setting a condition that the ground peak has to be at least 20% of the maximum peak in that waveform.

This prevents the ground peak to be assigned to multiple scattering peaks that may have occurred and that were delayed relative to the ground return. It is important to note that if the waveform has to go through high density matter, there will be energy attenuation and the ground signal may not be the strongest peak. As no absolute normalization of the waveform was performed to compensate for this energy attenuation through vegetation, intensity independent variables/metrics were required. However, a normalization method was applied to normalize each time bin by the total energy or area under the curve of each waveform [Peterson et al., 2005]. This accounts for the variation in energy from one footprint to another and allows for improved comparison between the shapes of individual waveforms.

2.4 Species Classification

2.4.1 Segmentation

Segmentation of the scene was required to identify individual trees for classification. ROIs were used to identify each tree, since segmentation was not the focus of this study. It was important that the right trees were identified and the similarity and proximity of certain canopies made this task challenging.

ROIs were determined manually by selecting regions corresponding to certain predetermined classes or species. Each tree was identified by a GPS coordinate from field data and these points were overlaid as vector files on the images. There is an error associated with each GPS location and this may cause corresponding points of the vector file to not always fall directly on canopies in the CHM. We thus assumed that these points are associated with the closest tree on the map. ROIs were also made smaller than whole tree regions in order to diminish the probability of selecting wrong classes in one ROI. We thus assumed that all pixels in each ROI were associated with one tree species.

2.4.2 Waveform Compositing

The ground bin for tree classification was assigned to bin 100 in the ground-aligned waveforms, permitting 15m of information above ground (100 bins x 0.15m), while 30 bins after the ground bin allowed for the inclusion of multiple scattering information. Thus, the total number of bands for all waveforms was set to 130 bands. Composite waveforms were constructed from the small-footprint waveform data to represent each tree. The method was used to simulate a larger footprint [Nayegandhi et al., 2006] and is represented by

$$\beta^{comp} = \frac{\sum_{i=1}^n \beta_i^{ind}}{n} \quad (2.9)$$

where β^{comp} is the resulting backscatter of the composite waveform, β^{ind} is the backscatter count for each waveform i , and n is the total number of waveforms used to construct the composite waveform.

2.4.3 Feature Metric Extraction

Features were extracted from the composite waveforms, such as the height and crown thickness of each tree, crown thickness to height ratio, first and second derivatives, and PCA. Features that were extracted from HSI included the first and second derivatives and PCA.

Tree Height and Crown Thickness

Typical tree waveforms usually exhibit more than one peak, due to multiple returns of the signal before hitting the ground. Tree height is advanced by [Wu et al., 2009b] to be the distance, d_1 , from the full width half maximum (FWHM) of the first peak to the last peak, where the last peak is generally the ground response of the deconvolved waveform. The second metric, d_2 , is advanced to be the distance between the FWHM of the first peak to the first valley. These metrics are shown in Figure 2.4 [Wu et al., 2009b].

Figure 2.4: Waveform visualization of tree height and crown thickness [Wu et al., 2009b].

Each pixel will have a d_1 and d_2 value. Two variables were associated with height for each tree, namely the mean and the maximum height recorded for all waveforms

inside each ROI. As for the crown thickness, this was calculated as the mean of all values in each ROI,

$$\bar{d}_2 = \frac{\sum_{i=1}^N d_{2,i}}{N} \quad (2.10)$$

where $\bar{d}_2$ is the mean crown thickness, N is the total number of pixels for each ROI, and i is the pixel count within each ROI. The crown thickness to height ratio (d_3) was also calculated as a feature in the model

$$d_3 = \frac{d_2}{d_1}. \quad (2.11)$$

First and Second Derivatives

The first and second derivatives can be found for each waveform, as

$$f'(x) = \frac{\delta f}{\delta x} \quad (2.12)$$

$$f''(x) = \frac{\delta^2 f}{\delta x^2}. \quad (2.13)$$

Their values were used as inputs to the SDA, as it was thought to provide valuable information related to the shape of the signal.

Principal Components Analysis

PCA is a method that has been used extensively in hyperspectral studies to decorrelate the data, maximize the variability in all bands, and reduce data dimensionality, but has not been applied to wLiDAR data. Thus PCA was applied to wLiDAR and an analysis was done to evaluate the possibility of including the results in our feature classification. PCA was also applied to HSI. An explanation of PCA is found in [Schott, 2007].

First, the covariance of the specific image is found. Each feature value in the transformed data set is a linear combination of the bands in the input data set [Schott, 2007],

$$PC_1 = \mathbf{e}_1^T \mathbf{x} = e_{11}DC_1 + e_{12}DC_2 + \cdots + e_{1N}DC_N \quad (2.14)$$

where PC_1 is the brightness value of the first principal component feature, $\mathbf{e}_1$ is the first eigenvector, and $\mathbf{x}$ is a particular digital count (DC) vector in all the bands N . The eigenvalues are first placed in decreasing order and the eigenvectors are then ordered according to their associated eigenvalues. In order to capture the same variance in all resulting bands of the PCA, all the sites across land uses under study were entered as

one big image when finding the covariance. Each band of each subset was reshaped into column vectors, which were then concatenated to form a big image with the same number of columns. Since each subset had a different number of bands, some padding had to be performed to bring each subset to the same size, i.e., the image with the largest number of bands.

This implies that there were N principal component features in the vector $\mathbf{PC}$, one for each pixel or vector $\mathbf{x}$. This transferred to the new PC transformed data set, written as

$$\mathbf{PC} = \begin{bmatrix} PC_1 \\ PC_2 \\ \vdots \\ PC_N \end{bmatrix} \quad (2.15)$$

In order to understand the effect of PCA on wLiDAR, we performed the following analysis to help find the contribution of the original bands for each PC image and more precisely, for the first PC. We also repeated this exercise on the HSI data for a comparative assessment.

Equation 2.16 shows the matrix $\mathbf{E}$ of eigenvectors ordered as per the eigenvalues (equation 2.17) and the vector $\mathbf{x}$:

$$\mathbf{PC} = \mathbf{E}^T \mathbf{x} \quad (2.16)$$

$$\mathbf{E} = \left[\begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1N} \end{pmatrix} \begin{pmatrix} e_{21} \\ e_{22} \\ \vdots \\ e_{2N} \end{pmatrix} \cdots \begin{pmatrix} e_{N1} \\ e_{N2} \\ \vdots \\ e_{NN} \end{pmatrix} \right] \quad (2.17)$$

$$\mathbf{PC} = \begin{bmatrix} (e_{11} \ e_{12} \ \dots \ e_{1N}) \\ (e_{21} \ e_{22} \ \dots \ e_{2N}) \\ \vdots \\ (e_{N1} \ e_{N2} \ \dots \ e_{NN}) \end{bmatrix} \begin{bmatrix} DC_1 \\ DC_2 \\ \vdots \\ DC_N \end{bmatrix} = \begin{bmatrix} PC_1 \\ PC_2 \\ \vdots \\ PC_N \end{bmatrix}. \quad (2.18)$$

We can see that the the first row of the eigenvector matrix is associated with the PC_1 , the second row with PC_2 , etc. In order to find which original band contributes most to each PC, we needed to find which eigenvectors had the highest values and associate them with their respective bands from equation 2.18.

2.4.4 Fusion Method - Stepwise Discriminant Analysis

We applied [Abdullah et al., 2001] interpretation of SDA in the following section. SDA is used to identify a subset G_i of $i = 1, 2, \dots, g$ feature metrics that maximizes discrimination between tree classes. This can be interpreted in terms of the sample Mahalanobis distance as

$$D_i^2 = (\mathbf{x} - \bar{\mathbf{x}}_i)^T S^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) \quad (2.19)$$

where $\mathbf{x}$ is the vector containing feature metrics, $\bar{\mathbf{x}}$ is the vector containing average feature metrics for a group i , and S is the pooled covariance matrix. This can also be written as a linear discriminant function with constants and a linear combination of $\mathbf{x}$, that

$$f_i(x) = (\bar{\mathbf{x}}_i)^T S^{-1} \mathbf{x} - \frac{1}{2} (\bar{\mathbf{x}}_i)^T S^{-1} (\bar{\mathbf{x}}_i) \quad (2.20)$$

where

$$S = \frac{\sum_{i=1}^g (n_i - 1) Cov_i}{\sum_{i=1}^g (n_i - g)} \quad (2.21)$$

$$Cov_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T. \quad (2.22)$$

We used a training test sample to compute S and Cov_i , and then solved equation 2.20 for an unknown sample. This produced discriminant scores for each feature metric which can be arranged as

$$\hat{f}_k(x) = \{f_1(x), f_2(x), \dots, f_g(x)\}. \quad (2.23)$$

We allocated group k to the feature x with the largest discriminant score. In order to optimize the subset which maximizes the classification rate, we used Wilks Λ selection criteria, which is known as the SDA itself. This reduction in the number of variables speeds the processing time and results in a more robust classification. The vector $\mathbf{x}_p = (x_1, x_2, \dots, x_p)$ of original feature metrics is generated from the training population comprising of g groups and p feature variables, while $\mathbf{x}_{p+1} = (x_1, x_2, \dots, x_p, x_{p+1})$ denotes the new vector that results when adding a new feature metric to x_p . To describe Wilk's Λ statistics, we used B and W , the between group and within-group sums of squares and products matrices, respectively, calculated from B_1 and W_1 as

$$\Lambda(x_p) = \frac{|W|}{B + W} \quad (2.24)$$

$$\Lambda(x_{p+1}) = \frac{|W_1|}{B_1 + W_1}. \quad (2.25)$$

The multiplicative increment resulting from partially adding x_p to x_{p+1} is

$$\tilde{\Lambda} = \frac{\Lambda(x_{p+1})}{\Lambda(x_p)}. \quad (2.26)$$

The F statistic,

$$F = \frac{n - g - p}{g - 1} \frac{1 - \tilde{\Lambda}}{\tilde{\Lambda}} \quad (2.27)$$

is applied to test the significance of the change between $\Lambda(x_p)$ and Λ_{p+1} . At each step of the SDA, the x_p with the largest F value is added to the set if larger than a specific threshold, F_{in} . Subsequently, the other variables in the set are re-examined and the one with the smallest F value is deleted if the F value is less than a second threshold, F_{out} . A moderate significance level of approximately 0.10 to 0.25 typically performs better than the use of a smaller or larger significance level in the model selection [Fernandez, 2003]. This method selects the best variables from the whole subset and was implemented with SAS.

2.4.5 Error Analysis

The final classification should identify each species and was evaluated against field data, thereby describing the overall accuracy. Additional measures of performance, developed by [Congalton and Green, 1999], were the producer's accuracies, which describes the number of correctly classified trees per species according to the training data (the row total), and the user's accuracies, which describes the total number of trees that were classified correctly for a species over the total number of trees that were classified under that class (the column total) [Lillesand et al., 2008]. The latter is a measure of commission error and indicates the probability of a tree being classified into a class and actually belonging to that class on the ground.

Finally, the *kappa* coefficient (k) will be calculated as

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (2.28)$$

where N is the total number of observations in the confusion matrix, r is the number of rows, x_{ii} is the number of observations in row i and column i of the major diagonal, x_{i+} is the total number of observations in row i , and x_{+i} is the total number of observations in column i . It indicates how well a classification performs when compared to a chance assignment. k varies between 0 and 1, where 0 indicates that the classification is no better than a random assignment of pixels and 1 indicates that the classification is 100% better than a chance assignment [Congalton and Green, 1999, Lillesand et al., 2008].

An approach described in [Foody, 2004] was employed to assess significant differences between the three classifications by SDA on wLiDAR, HSI, and the fused data set, and consisted of a comparison of accuracy based on the "proportion of correct allocations". This method is one of the most commonly used accuracy metrics in remote sensing [Trodd, 1995]. The same study site was used for all three classifications, which implies a dependent sampling approach. A standardized normal statistical test based on binary distinction between correctly and incorrectly classified allocations can be written as

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (2.29)$$

where f_{ij} represents the frequency of sites lying in confusion matrix element i and j . The critical z -value is selected as $|z| = 1.96$ ($\alpha = 0.05$), where H_0 , signifying no difference between classification accuracies, is rejected at absolute z -values > 1.96 .

$$z = \frac{f_{ci} - f_{ic}}{\sqrt{f_{ci} + f_{ic}}} \quad (2.30)$$

Table 2.2: Assessment based on the proportion of correct allocations [Foody, 2004].

Allocation	Classification 1	
	Correct	Incorrect
Classification 2		
Correct	f_{11}	f_{12}
Incorrect	f_{21}	f_{22}

2.5 Biomass Modeling

The first step in biomass modeling were selection of herbaceous plots, from which the preprocessed waveforms were grouped into composite waveforms. This was followed by the decomposition of the composite waveforms into Gaussian approximation curves and feature metric extraction based on the Gaussian fit and other useful parameters, such as rise and fall time to the ground peak. Feature extraction was also performed with HSI. These features were used as input in a FSM, which optimized variable selection for a specific system modeling, in this case biomass modeling.

2.5.1 Waveform Compositing

Differential GPS points were collected at locations where herbaceous biomass was measured, but in order to account for the GPS location error of 1-2m on all imagery, data inside a 3x3 pixel region centered about each GPS point were averaged, meaning that composite waveforms were formed for each plot. Since herbaceous matter is only present below about 2m in South Africa, all waveforms were cropped to 13 bins ($0.15m \times 13 = 1.95m$) above the ground peak. Also, we hypothesized as mentioned by Wu *et al.* (2009a), that herbaceous biomass could be linked to the waveform component just below the ground peak, which is normally hidden in the raw waveform before deconvolution. This is believed to be caused by multiple scattering, and could be indicative of herbaceous biomass, therefore the waveforms were cropped to 16 bins after the ground peak in order to account for this phenomenon. Waveform data from trees were excluded from the sites as this could have induced error. This was done by excluding waveforms that exhibited more than two peaks after deconvolution. The final cropped waveforms had 30 bands and all the peaks were aligned at bin 14. Equation 2.9 was used to build each composite waveform.

2.5.2 Gaussian modeling and feature selection

The decomposition of composite waveforms into Gaussian approximation curves is modeled such that

$$f(t) = Ne^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad (2.31)$$

which leads to extraction of waveform metrics such as the intensity of each peak (N) and the standard deviation or width of each pulse (σ). We already knew that the location of the estimated peaks measured for each waveforms (μ) would be the same. The rise time to the leading edge and fall time of the trailing edge of the ground peak, in time bins, were identified as useful features and were thus calculated. Also, we hypothesized that it would be useful to find the typical ground waveform without any matter above ground. Each composite waveform could then be compared to this pure ground waveform and we hypothesized that the area difference between the two could be an indication of herbaceous biomass. The most uniform area in a scene of the KNP was on a dirt road, so a composite waveform was built from as many individual road waveforms as possible, and was subsequently area-normalized. In order to make sure that the road composite waveform really is representative of the undercover of each land use, the same method was employed with bare ground pixels for comparison purposes.

The metrics that were considered both for HSI and wLiDAR were first and second derivatives of the original bands. Also, the same method as described in section 2.4.3 was applied to find

the PCA of both HSI and wLiDAR data. Another HSI feature of interest was determined to be the unmixed "grass signal", which provided an abundance map of grass. The grass spectra used in the unmixing was obtained from an area on the map through visual inspection. The features were then entered in the FSM, as described in the next section, in order to optimize the variable selection and relate herbaceous field data to feature metrics. This method used field data from the KNP and the Bushbuckridge communal settlement, along with CASI HSI and LiDAR data.

2.5.3 Fusion Method - Forward Selection Model

The forward selection method begins with no variables, and a predictor feature or metric is added to the model at each step of the procedure. The selection process is based on Wilks' lambda, as described in Section 2.4.4, which statistically reflects how each independent variable would contribute to the model if they were added. The most contributing variable is added to the model, and cannot be removed in future steps. When no variable produces a significant difference in the model, the procedure stops [Fernandez, 2003, SAS Institute Inc., 2008].

We assumed that the original and the PCA data sets were highly correlated, since they represent the same data; the PCA is a linear transformation from the original data set. This is the reason why the original and PCA-based variables were treated separately as inputs into this model.

2.5.4 Error Analysis

The coefficient of determination (R^2)

The coefficient of determination (R^2) is used to explain the proportion of variance in dependent variables explained by the independent variables in the regression model [Nagelkerke, 1991]. It usually varies between 0 and 1, 0 not explaining any variance and 1 predicting each output perfectly. In this case, the coefficient of determination is simply the square of the correlation coefficient between the field data and the predicted biomass values, given as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (2.32)$$

where X and Y are field and predicted variables correspondingly, $cov(X,Y)$ is the covariance between both variables, and σ_X and σ_Y are standard deviations [Ozer, 1985].

RMSE

Another measure of performance between estimated and field measured biomass is the root mean square error (RMSE), given as

$$RMSE_{X,Y} = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}} \quad (2.33)$$

where lower values are indicative of better predictions. [Willmott and Matsuura, 2005]

Chapter 3

Results and Discussion

3.1 Species Classification

3.1.1 Segmentation

Different environmental factors could affect the results of the proposed classification techniques, such as the slope of the terrain, the density of trees and their vertical structure, the time of year at which data were gathered, vegetation phenology, etc. Tree segmentation is directly affected by the density of the forested area, even in such a savanna environment. Such considerations are of prime importance when selecting trees corresponding to the appropriate classes. Of the species of interest, there was a total of 306 trees for which GPS locations were recorded. Some points were completely excluded from the analysis based on the fact that it was too difficult to associate them with canopies. This had the effect of diminishing the samples size of each species, from 109 to 65 for *Combretum hereroense*, from 119 to 74 for *Sclerocarya birrea*, from 49 to 33 for *Terminalia sericea*, from 29 to 20 for *Peltophorum africanum*. This effectively increased our confidence in tree positions, GPS data records, and associated per-tree spectra. Figure 3.1(a) shows a subset of the CHM for a land use area from the Bushbuckridge communal settlement along with the GPS locations that were recorded. Figure 3.1(b) shows the same subset along with the ROIs that were constructed to represent each tree. The exact GPS points identifying each tree used in this section can be found in Appendix C.

Figure 3.1: GPS locations super-imposed on a subset of the Bushbuckridge communal settlement and their ROIs.

3.1.2 Feature Metric Extraction

The main Matlab (version 7.8) algorithm that was coded to calculate features from each ROI is shown in Appendix D.

Tree Height and Crown Thickness

Features were extracted from the composite waveforms, such as the crown width and the height, as described in Section 2.4.3, which resulted in d_1 and d_2 values for each pixel. For each ROI, the mean of the crown width and height values were calculated, along with the maximum height, in order to represent each tree. The Matlab (version 7.8) code that was developed to calculate these metrics is shown in Appendix D.2. Figure 3.2 shows the height and crown thickness images for a subset of a land use situated in the Bushbuckridge communal settlement.

Figure 3.2: Intensity image representing a) tree height, d_1 , and b) crown thickness, d_2 [Wu et al., 2009b].

The values obtained by using this method were accurate to $0.15m$, since each value for the FWHM, peak, or valley was calculated to the closest $1ns$ bin values, which each represented $0.15m$.

From these two images, not all pixels seem to have the appropriate height and crown thickness values. This was attributed to the fact that not all waveforms hit the top canopy of each tree, some penetrate and return a signal further down the tree structure. Such interactions are arguably species-, phenology-, canopy-gap-, and system dependent.

First and Second Derivatives

Images of the first and second derivatives of HSI (at $1045nm$) and wLiDAR (at $3.75m$ above ground) are shown in Figure 3.3. Both derivatives for HSI show interesting details that seem to represent ground detail. As for wLiDAR, both the original and the derivative images seem to represent trees, while the intensities vary between the three images.

Figure 3.3: a) Original, b) first and c) second derivatives of HSI at $1045nm$. d) Original, e) first and f) second derivatives at $3.75m$ above the ground of wLiDAR for a subset of the Bushbuckridge communal settlement.

Figure 3.20(a) shows an example of a selected waveform in its original deconvolved form, with the ground peak aligned at time bin 100, while Figures 3.20(b) and 3.20(c) show the signal after calculating first and second derivatives, respectively.

(a) Original waveform.

(b) First derivative.

(c) Second derivative.

Figure 3.4: The same waveform, chosen at random, is represented for the a) deconvolved subset, b) the first derivative and c) the second derivatives in the Bushbuckridge communal settlement.

PCA

The PCA was applied to HSI and wLiDAR. The first three PCs of HSI are shown in Figure 3.5. We can see the outline of vegetation in PC_1 , compared to PC_2 and PC_3 , which represent more subtle vegetation features. This is underscored by the fact that Figure 3.5(a) contains about 94% of the variation compared to approximately 5% in PC_2 and 1% for PC_3 . As a matter of fact, 99.5% of the variation was contained in the first three PCs.

Figure 3.6 shows the first three PCs for wLiDAR. The variance majority was contained in more bands than for HSI, which is what we would expect, as the data were from two totally different sensors modalities that have different numbers of bands to start with and represent different phenomenologies, i.e., spectral vs. structural domains. As a matter of fact, 99.5% of the variance was contained in the first 41 bands of the PCA, so the information was more spread out. As a comparison, the first three PCs of wLiDAR contained approximately 78% of the variance, which was significantly less of that of HSI. This may be due to the fact that the ground was standardized to a specific time bin for all waveforms, which may not be the optimum method, as the variation of vegetation relative to sea level is not captured. This could also be explained by the fact that waveforms were not normalized for energy attenuation through vegetation, so a first return may seem very high comparative to the ground return. These are just some of the reasons that might explain why only 78% of the variance was contained in the first few bands. Figure 3.7 shows the histogram of the variance values of the first 41 PCA bands, which varied between 0 and 44%. This shows that a majority of those 41 bands explained between 0 and 5% variance.

Figure 3.5: Images of the first three bands of the PCA on HSI, a) band 1, b) band 2 and c) band 3.

Figure 3.6: Images of the PCA for wLiDAR for a) band 1, b) band 2 and c) band 3.

Figure 3.7: Histogram of the variance values of the first 41 PCA bands for wLiDAR.

In order to find which original band contributed most to each PC, we constructed graphics of the eigenvectors versus the spectral range for HSI and time for wLiDAR. Figure 3.8 shows the results for the first three eigenvectors of HSI, which represent the first three PCs of HSI. PC_1 shows an ascending contribution from the eigenvector elements associated with the spectral range between 678 and 1045nm, which is represents the red edge and near-infrared regions, both of which are important vegetation indicators [van Aardt and Wynne, 2001]. PC_2 exhibits a negative contribution of all eigenvector elements, except for those associated with the range between 743 to 932nm. It is also worth noting there is a minimum at 678nm. PC_3 has a small peak at 678nm and a minimum at 763nm. There is a strong increasing contribution of the elements of eigenvectors between 791 and 1045nm.

Figure 3.8: Eigenvector analysis for HSI.

In Figure 3.9, we know each time stamp of 1ns represents 0.15m in space as explained in section 2.3, with the ground being at time bin 100. The first time bins correspond to higher elevation data and time bins after 100 correspond to potential multiple scattering elements. Such scattering effects can be thought of as delayed returns and are believed to occur in more complex structures. All PCs have contributions close to the ground and in the multiple scattering region (after time bin 100), but let's examine the eigenvector elements for PCs one to three that contribute most to above ground variation (time bin 1 to 100). The first PC has a positive contribution from the ground to about 4m. The second PC is neutral above ground, with the main contribution coming from multiple scattering. This is explained by the image of PC_2 in Figure 3.6 where the brightest pixels seem to designate non-tree areas. PC_3 is defined by a negative contribution from 1m to approximately 4m above ground. We can deduce that the ground itself and multiple scattering contributed to most of the PCs, but eigenvectors one and three had elements that favor a contribution of similar timescale

Figure 3.9: Eigenvectors vs. time for wLiDAR.

ranges above ground, thus we mostly see information that corresponded to structure between the ground and $4m$. This corroborated the information seen from the eigenvector analysis for HSI, which indicated that the first three PCs contain near-infrared interactions.

3.1.3 Feature-level Fusion - Stepwise Procedure

For all feature-bands obtained previously (first and second derivative, PCA, crown thickness, etc.), we retained the mean per-band value in each ROI, as inputs to the SDA model. In addition, the maximum height for each ROI was also chosen as a variable for each tree. Six SDAs were performed, the first two included wLiDAR variables only, with and without PCA variables, the third and fourth used HSI variables only, with and without PCA variables, and the fifth and sixth were a fusion of independent variables that were selected from both previous SDAs from different sensors, with and without PCA variables. A summary of all four stepwise procedures can be found in Appendix A. The a priori probabilities for all the classes were set to $1/4$, since there are four classes of species. The following 620 features from both sensors were used:

- HSI (24 bands) and composite wLiDAR (130 bands), for which band 100 corresponded to the ground, with each previous band representing matter above ground, and each subsequent band representing potential multiple scattering;
- PCA of HSI (24 bands) and wLiDAR (130 bands);
- first derivative of HSI (24 bands) and wLiDAR (130 bands) in the x direction;
- second derivative of HSI (24 bands) and wLiDAR (130 bands) in the x direction;
- mean and maximum height from wLiDAR data (2 band);
- mean crown thickness from wLiDAR data (1 band); and
- crown thickness to height ratio from wLiDAR data (1 band).

Waveform LiDAR

The stepwise procedure in SAS revealed that the most significant variables for effective clustering of wLiDAR data in terms of species classification were:

1. intensity at $0.45m$ above ground;
2. intensity at time bin 102, assumed $0.3m$ underground (multiple scattering);
3. intensity at $7.8m$ above ground; and
4. intensity at time bin 124, assumed $2.25m$ underground (multiple scattering).

The significance level for variable entry into the model was fixed at 0.0008. This number was obtained by trial and error in order to obtain four different feature outputs from the stepwise regression for all attempts. Figure 3.10 shows the resulting images that were selected in the stepwise procedure of wLiDAR, without PCA variables, for a specific subset in the area of interest, situated in the Bushbuckridge communal settlement. Figures 3.10(a) and (b) represent data very close to the ground, while Figures 3.10(c) and (d) represent information that is further from the ground, which explains why there are so many zero values.

Figure 3.10: Selected features for species classification from the original wLiDAR bands: a) intensity at $0.45m$ above ground, b) intensity at time bin 102, theoretically $0.3m$ underground (multiple scattering), c) intensity at $7.8m$ above ground, and d) intensity at time bin 124, theoretically $2.25m$ underground (multiple scattering).

The stepwise procedure in SAS revealed that the most significant variables for effective clustering of wLiDAR data with PCA variables were:

1. component 6 of the PCA;
2. maximum height;
3. component 36 of the PCA; and
4. mean height.

The significance level was fixed at 0.0001. This number was obtained by trial and error in order to obtain four different feature outputs from the stepwise regression. Figure 3.11 shows the resulting images that were selected in the stepwise procedure of wLiDAR, with PCA variables. Figures 3.11(a) and (c) do not highlight any particular feature, however, while Figure 3.11(b) clearly identifies trees based on their height.

Figure 3.11: Selected PCA-based features for wLiDAR: a) component 6 of the PCA, b) height, and c) component 36 of the PCA.

HSI

The stepwise procedure in SAS revealed that the most significant variables for effective species-based clustering of the HSI data without PCA metrics were:

1. reflectance at $423nm$ (blue region);
2. reflectance at $394nm$ (blue region);
3. reflectance at $649nm$ (red region); and
4. second derivative at $763nm$ (red-edge/NIR region).

The significance level was fixed at 0.00001, which is lower than for original wLiDAR features. Figure 3.12 shows the resulting images that were selected in the stepwise procedure of original HSI bands, which is a mix of spectral and second derivative features. Figures 3.12(d) does not seem to capture any particular features of interest when compared to the spectral images, which capture vegetation at different wavelengths.

The stepwise procedure in SAS revealed that the most significant variables for effective clustering of the HSI data with PCA metrics were:

1. component 5 of the PCA;
2. component 14 of the PCA;
3. component 2 of the PCA; and
4. component 20 of the PCA.

The significance level was fixed at 0.00001, which is in the same as the previous section. Figure 3.13 shows the resulting images that were selected in the stepwise procedure of HSI. These images all seem to exhibit some noise components and no features are apparent in terms of explaining their selection, apart from Figures 3.13(a) and (c), which are components 5 and 2 of the PCA that exhibit close to zero and 5% variance, respectively. However, although difficult to interpret, we have to conclude that these PCs were useful in explaining between-species variability.

Figure 3.12: Selected features for HSI: a) reflectance at $423nm$, b) reflectance at $394nm$, c) reflectance at $649nm$, and d) second derivative at $763nm$.

Figure 3.13: Selected PCA features for HSI: a) component 5 of the PCA, b) component 14 of the PCA, c) component 2 of the PCA, and d) component 20 of the PCA.

Waveform LiDAR and HSI

The independent variables that were chosen for non-PCA variables for both sensors and used as input into the discriminant analysis of original fusion variables were:

1. waveform intensity at $0.45m$ above ground;
2. waveform intensity at time bin 102, theoretically $0.3m$ underground (multiple scattering);
3. waveform intensity at $7.8m$ above ground;
4. waveform intensity at time bin 124, theoretically $2.25m$ underground (multiple scattering);
5. reflectance at $423nm$;
6. reflectance at $394nm$;
7. reflectance at $649nm$; and
8. second derivative of reflectance at $763nm$.

The independent variables that were chosen for PCA variables for both sensors and used as input into the discriminant analysis of PCA-based fusion variables were:

1. component 6 of the PCA of wLiDAR;
2. maximum height;
3. component 36 of the PCA of wLiDAR;
4. mean height;
5. component 5 of the PCA of HSI;
6. component 14 of the PCA of HSI;
7. component 2 of the PCA of HSI; and
8. component 20 of the PCA of HSI.

Figure 3.14 shows the eigenvectors that contributed most to PC_6 and PC_{36} of wLiDAR, which were selected as wLiDAR independent variables in the stepwise procedure of wLiDAR. PC_6 exhibited a varying signal from time bins 60 to 100, from the ground to approximately $6m$. This PC also included a contribution from potential multiple scattering occurring from time bins below the ground. PC_{36} only contained a noticeable contribution from multiple scattering bands, after time bin 100. Figure 3.15 shows the eigenvectors that contribute the most to PC_5 , PC_{14} , and PC_{20} of HSI, all of which were selected as HSI independent variables in the stepwise procedure of HSI. PC_5 exhibited a strong contribution from $394nm$ to $536nm$ and two noticeable peaks at $706nm$ and $989nm$. PC_{14} had three strong positive contributions at $649nm$, $734nm$, $876nm$ and a strong negative peak at $819nm$. PC_{20} contained three major positive spectral contributions at $423nm$, $536nm$, and $678nm$. These wavelength regions (blue-green and near-infrared) have been shown to be useful in species classification in other studies [van Aardt and Wynne, 2001, Martin et al., 1998].

As mentioned by [Fernandez, 2003], the SDA method is not always guaranteed to give the best results, since only one variable can enter the model at each step and this means that not all

Figure 3.14: Eigenvector analysis for wLiDAR, PC 6 and 36. It is clear that most of the variation originates close to the assumed ground ($100ns$), although PC 36 exhibits a strong contribution from potential multiple scattering (greater than $100ns$).

possible combinations of predictors and interactions among them are evaluated. Certain important variables could then be excluded in the selection process. It has also been shown that highly correlated predictor variables should be excluded from the model if they do not have discriminatory power [Fernandez, 2003].

3.1.4 Feature-level Fusion - Discriminant Analysis

Tables 3.1, 3.3, and 3.5 show the confusion matrices for wLiDAR, HSI, and both data sets, respectively, without PC variables. We assumed that the original data set and the PCA are highly correlated, since they represent the same data; the PCA is a linear transformation of the original data set. Tables 3.2, 3.4, and 3.6 show the confusion matrices for wLiDAR, HSI, and both data sets, respectively, with PCA variables. This provides the number of classified and mis-classified observations on a per-species basis, while the producer's and user's accuracies are also included in each table. The linear discriminant function for classification of each species are shown in Appendix A, for all six classifications.

Figure 3.15: Eigenvector analysis for HSI, PCs 5, 14, and 20. Note the strong negative peak in PC 14 from the red edge region, which is most likely a strong indication of vegetation.

For HSI variables, we can see that 41 *Combretum hereroense* were correctly classified, which represent 63% with an associated error rate of 37%. 81% of the *Sclerocarya birrea*, 70% of the *Terminalia sericea*, and 85% of the *Peltophorum africanum* were correctly classified. The overall classification for the original HSI variables was 73% and the overall error rate was 27%, which is better than both previous wLiDAR classifications. The $kappa$ coefficient was $k = 0.63$.

Reference Data	Species - Classification					Producer's Accuracy
	Cher	Sbir	Tser	Pafr	Row Total	
Cher	41	0	5	19	65	63%
Sbir	4	62	2	9	77	81%
Tser	2	3	23	5	33	70%
Pafr	2	1	0	17	20	85%
Column Total	49	66	30	50	195	
User's Accuracy	84%	94%	77%	34%		
Overall accuracy = 73%						
Kappa statistic = 0.63						

Table 3.4: Confusion matrix resulting from SDA for PCA-based HSI variables.

	Species - Classification					
Reference Data	Cher	Sbir	Tser	Pafr	Row Total	Producer's Accuracy
Cher	40	3	4	18	65	62%
Sbir	5	61	7	4	77	79%
Tser	0	3	25	5	33	76%
Pafr	5	1	2	12	20	60%
Column Total	50	68	38	39	195	
User's Accuracy	80%	90%	66%	31%		
Overall accuracy = 71%						
<i>Kappa</i> statistic = 0.60						

For the original fused data variables, the overall classification remained unchanged compared to the original HSI variable classification with 73% for both classifications. The original wLiDAR overall classification was much lower, achieving only 53%. For the PCA-based fused data variables, the classification increased marginally when compared to the PCA-based HSI classification, with 74% versus 71%, respectively. The PCA-based wLiDAR classification was the same as the original results, obtaining only a 53% overall classification. It is surprising that the fusion of PCA-based variables obtained a better result, considering that both wLiDAR classifications were the same and that the PCA-based HSI classification was lower than that of the original HSI classification. This was attributed to the PCA resulting in a better alignment of between-species variation when inputs from both modalities were used.

For all classes, the wLiDAR producer's and user's accuracies were equal or smaller than for HSI classifications. The only two classes where the fused data set improved the producer's accuracy with the original variables were with *Combretum hereroense* and *Sclerocarya birrea*. This indicator is a good measure of how well the species were classified, as it refers to the validation data. In the first case, the producer's accuracy was 43% for wLiDAR, 63% for HSI, and 68% for the fused data set. In the second case, the producer's accuracy was 81% for wLiDAR, 81% for HSI, and 86% for the fused data set. For the PCA-based classification, there were three classes where the producer's accuracy improved in the fused data set: *Combretum hereroense*, *Sclerocarya birrea*, and *Peltophorum africanum*. In the first case, the producer's accuracy for wLiDAR was 52%, 62% for HSI, and 68% for the fused data set. In the second case, the producer's accuracy was 78% for wLiDAR, 79% for HSI, and 84% for the fused data set. In the third case, the producer's accuracy was 15% for wLiDAR, 60% for HSI, and 80% for the fused data set. For these particular species, the fused results were an improvement, with an increase in the producer's accuracy of at least 5% compared to HSI and/or wLiDAR alone. This could mean that a couple of poorly classified and potentially variable responses and structures might have skewed our final classification results. *Terminalia sericea* is the only one for which the fused data-set did not improve the producer's accuracy and *Peltophorum africanum* was only improved for the PCA-based fused classification. This could also be due to the fact that these two species are not separable from the others, or again it could be due to their inherently high spectral and structural variability. The fact remains that good producer's accuracies were achieved for a little more than half of the classifications, which is a sign that the fusion itself might be a reliable tool to use in the future. However, the overall classification accuracy would potentially only improve if we identify species that do not benefit from set fusion.

There was no improvement in user's accuracy with the original fused variables compared to HSI alone, however it was higher than for wLiDAR. As for the PCA-based data set, the three cases where the user's accuracies improved were *Combretum hereroense*, *Terminalia sericea*, and *Peltophorum africanum*. In the first case, the user's accuracy was 67% for wLiDAR, 80% for HSI, and 83% for the fused data set. In the second case, the user's accuracy was 47% for wLiDAR, 66% for HSI, and 80% for the fused data set. In the third case, the user's accuracy was 16% for wLiDAR, 31% for HSI, and 41% for the fused data set.

We therefore concluded that, although fusion did not result in overall classification improvements, it did contribute to species-specific improvements. This was attributed to potentially large variation in the case of *Terminalia sericea* and *Peltophorum africanum*. We noticed that *Sclerocarya birrea* obtained the best producer's accuracy in most classifications. *Sclerocarya birrea* also always obtained the best user's accuracies. *Sclerocarya birrea* is a tree that generally has a large canopy that is easily distinguished while defining ROIs, given its distinct blue hues in the

visible region of HSI. The trees in the scene might also have had a wide range of maturity (phenological stages) since it's the class that performed best. We concluded that there was less error associated with these ROIs. The GPS locations also have an inherent inaccuracy, which in turn could have led to mixed pixels in the three other species ROIs. Finally, the results were also ascribed to structural variations due to the phenology of the different species during the time of data collection in April 2008.

The *kappa* statistics were generally representative of the classification results as wLiDAR obtained between 0.41 and 0.44 accuracy for both classifications, HSI obtained between 0.60 and 0.63 for both classifications, and the fusion of variables obtained between 0.62 and 0.64. The HSI and the fused classification obtained a relatively good *kappa* accuracy. The PCA-based approach proved to be more useful in the fusion classification (74%) than for HSI (71%), the latter exhibiting more successful classification with original variables (73%).

3.1.5 Z-test

To compare the proportion of correct allocation between the classification of different data sets, the z-test was performed, as detailed in Section 2.4.5. It is based on 2 x 2 confusion matrices indicating correct and incorrect allocations of tree ROIs in pairs of classifications under evaluation. Three comparisons were performed relating the best classifications from wLiDAR, HSI, and the fusion classification, as shown in Tables 3.7, 3.8, and 3.9. This was performed in order to evaluate which classification was significantly better than others. The classification results used to calculate this metric are shown in Appendix A.

Table 3.7: Assessment based on the proportion of correct allocations between original wLiDAR and PCA-based fusion classifications

Allocation	Classification of PCA-based fusion classification	
	Correct	Incorrect
Classification of original wLiDAR variables		
Correct	96	17
Incorrect	49	33
z-value	-3.9389	

Table 3.8: Assessment based on the proportion of correct allocations between original HSI and PCA-based fusion classifications

Allocation	Classification of PCA-based fusion classification	
	Correct	Incorrect
Classification of original HSI variables		
Correct	129	14
Incorrect	16	36
z-value	-0.3651	

Table 3.9: Assessment based on the proportion of correct allocations between original wLiDAR and HSI classifications

Allocation	Original HSI classification	
	Correct	Incorrect
Original wLiDAR classification		
Correct	92	21
Incorrect	51	31
z-value	-3.5355	

To determine if there was a significant difference in the proportion of correct allocations for a two-tailed normal distribution, the null hypothesis H_0 , or the point of no significant difference, would be rejected at $\alpha = 0.05$, which corresponds to a tabled z-value of 1.96 [Foody and Hill, 1996, Congalton and Green, 1999]. The first and last classification comparisons were inside the standard test performance, which means that there were significant differences between the evaluated classifications. But comparison between original HSI and PCA-based fusion classifications are outside the standard test performance, meaning there was no significant difference between the two classifications [Congalton and Green, 1999].

It is evident that tree features need to be located more accurately in the future to correspond to each composite waveform. The waveforms themselves are more accurate when compared to the positional accuracy of the GPS, since we had to include 3x3 composite waveforms to compensate for that fact. The choice of this particular number was very conservative, so if the GPS error was greater than 1m, the composite waveform would need to be enlarged to include more voxels.

These outcomes were similar than the ones obtained with a previous data set from the Skukuza flux site in the KNP [Sarrazin et al., 2010], indicating that the fusion of wLiDAR structural features to HSI reflectance metrics did not increase classification accuracy significantly. Although species were different between the studies, the amount of preprocessing required to extract viable metrics

from wLiDAR would not, at the moment, be worth the effort if resources were a restraining factor. The overall classification results obtained in [Sarrazin et al., 2010]’s study were very similar to our 74% achieved with the PCA-based fusion data set. They obtained a 72% overall accuracy with their fusion data set, which included both original and PCA-based features in one SDA. Their best *kappa* coefficient was 0.56, while our best was 0.64. It seems that treating original and PCA features separately to prevent the inclusion of dependent variables in the SDA model is an essential consideration prior to classification.

3.2 Herbaceous Biomass Modeling

3.2.1 Site-level Plot Selection

Differential GPS points were collected at locations where herbaceous biomass was measured. An overview of the sites can be viewed in Figure 3.16(a), as part of land uses 2, 7, and 8. A subset image of such a site from land use 8 is shown in Figure 3.16(b), with plot extractions centered about a 3x3 pixel region to compensate for GPS error. The exact GPS locations can be found in Appendix C. Using the method described in Section 2.5.1, composite waveforms were generated at each plot location. The Matlab (version 7.8) algorithm that was coded to calculate the composite waveform is listed in Appendix E.5.

Figure 3.16: (a) An overview of the sites as part of the landuses where wLiDAR data was collected in the KNP and the Bushbuckridge communal settlement and (b) a site-level example of landuse 8, showing 3x3 pixel areas of imagery data for each GPS location.

3.2.2 Feature Metric Extraction

The main Matlab (version 7.8) algorithm that was developed in order to calculate feature variables for each herbaceous plot is found in Appendix E. The algorithm that was used to calculate the rise time to the leading edge and fall time of the trailing edge can be found in Appendix E.7.

Area Difference Feature

The typical waveform, without any above ground vegetation, was calculated from the most uniform area on the map (a dirt road in the KNP), so that each composite waveform in the image could be compared. In order to make sure that the road composite waveform really is representative of a bare ground area, a composite waveform was constructed with bare ground pixels within a site.

Figure 3.17 shows the resulting composite waveforms, area normalized for comparison purposes. The shape is very similar, which confirms that we could use the road composite waveform to represent a typical waveform without any above ground vegetation. Figure 3.18 shows an example of a road waveform superimposed on a plot composite waveform and their area difference, which was hypothesized to be correlated with herbaceous biomass. This area difference was used as a feature in the FSM.

Figure 3.17: Comparison between dirt road and bare soil composite waveforms.

Figure 3.18: An example of herbaceous biomass plot composite waveform compared to a dirt road waveform, and the difference between the two.

First and Second Derivatives

Principal metrics that were considered both for HSI and wLiDAR were the first and second derivatives of the original data. Section 3.1.2 of species classification shows the resulting images of those metrics on HSI and wLiDAR. The only difference is that wLiDAR in this section had a smaller number of bands, thus Figure 3.19 shows the resulting images corresponding to the first and second derivatives of wLiDAR at $0.30m$ above ground, which corresponds to herbaceous vegetation information. Figure 3.20 shows a random waveform and its resulting first and second derivatives.

Figure 3.19: a) Original, b) first and c) second derivatives of wLiDAR at $0.3m$ above ground for a subset of the Bushbuckridge communal settlement. The images seem to show less vegetation detail from one image to another with intensity values also diminishing between each derivative images.

(a) Original waveform.

(b) First derivative.

(c) Second derivative.

Figure 3.20: The same randomly chosen waveform is represented from the a) deconvolved subset, b) first derivative, and c) second derivatives in the Bushbuckridge communal settlement.

HSI Unmixing

Another important HSI feature was related to the per-pixel grass abundance, a product of a linear pixel unmixing approach, based on defined endmembers. The selected endmembers were based on common general areas visible from the HSI for all sites, such as tree, shadow, bare soil, and grass. A typical grass spectrum was chosen in order to perform the unmixing with ENVI software (version 4.6), since results were similar when different representative grass curves were selected. Figure 3.21 shows the resulting image. It should be noted that unmixing results were not verified, since the goal was to evaluate if stand-alone HSI abundance data could constitute a valuable modeling metric.

Figure 3.21: Grass abundance map as calculated with ENVI software (version 4.6) using a linear unmixing approach [ENVI, 2008].

PCA

The PCA was applied to HSI and wLiDAR. The first three PCs of HSI are shown in Figure 3.22 and are a little different from section 3.1.2, since it was derived from the co-variance metrics of different images or image areas. We can see the outline of vegetation with different clarities through this figure. This is explained by the fact that Figure 3.22(a) contains about 87% of the variation, compared to approximately 12% in PC_2 and 1% for PC_3 . As a matter of fact, 99.5% of the variation was contained in the first four PCs.

Figure 3.23 shows the first three PCs for wLiDAR. The variance majority was contained in more bands than for HSI, which was expected, as the data was from two totally different sensors that in turn measures vastly different phenomena. As a matter of fact, 99.5% of the variance was contained in the first 11 bands of the PCA. As a comparison, the first four PCs of wLiDAR contained approximately 96% of the variance, which was approximately the same as for HSI. This may be due to the smaller number of bands, i.e. constrained to close-to-ground interactions, than was the case in the previous section with species classification.

Figure 3.22: Images of the first three bands of the PCA on HSI for herbaceous biomass modeling, a) band 1, b) band 2 and c) band 3. Note the clear vegetation structure present in all tree PCs.

Figure 3.23: Images of the first three bands of the PCA on wLiDAR for herbaceous biomass modeling, a) band 1, b) band 2 and c) band 3.

The code that was used to calculate the PCA is found in Appendix E.2. In order to find which original band contributed most to each PC, graphics of the eigenvectors versus the spectral range (HSI) and time (wLiDAR) were constructed. Figure 3.24 shows the results for the first three eigenvectors of HSI, which represent the first three PCs of HSI. PC_1 shows a noticeable ascending contribution from the eigenvector elements associated with the spectral range of 706 to 1045nm. PC_2 exhibits a positive contribution before 678nm and a mostly negative contribution after that same point, except at 1045nm. PC_3 has peaks at 564nm, and 763nm, and then strongly descends to 1045nm. This is a little different than the eigenvector contribution from the previous section, which is attributed to different image areas being used for the two use cases, namely species classification and herbaceous biomass modeling. However, the red-edge and near-infrared regions were again prominent.

Figure 3.24: Eigenvector analysis for HSI in the case of image areas used for herbaceous biomass modeling. Note the eigenvectors crossing at the beginning of the red edge region ($700nm$) and the strong negative contribution from PC_3 starting at $1017nm$, indicative of potential vegetation information.

In Figure 3.25, each time stamp of $1ns$ represents $0.15m$ in vertical space, as mentioned in section 2.3. Everything below $14ns$ represents above ground information and the information higher than $14ns$ represents potential multiple scattering elements. All three wLiDAR PCs exhibit contributions in the multiple scattering region, i.e. higher than $14ns$. The first PC has a strong negative contribution from the ground. All three PCs have similar varying contributions from the ground to $0.75m$ above ground. This is the vertical range where grass biomass is normally found. We hypothesize that the inclusion of multiple scattering effects are indicative of close-to-ground herbaceous scattering.

Figure 3.25: Eigenvector analysis for wLiDAR in the case of herbaceous biomass modeling. All PCs have similar eigenvector contributions in the same time range, i.e., $0.75m$ above and below (assumed multiple scattering) the ground.

Gaussian Modeling

An example of Gaussian approximation curves was obtained and is shown in Figure 3.26, which is a close representation of the deconvolved waveform. The Gaussian modeling algorithm was only performed on waveforms with two peaks or fewer, such as shown here. The Gaussian modeling algorithm did not perform well for all composite waveforms that did not exhibit an almost perfect Gaussian shape due to averaging or angle correction, which resulted in 16 being dropped from the 319 original plots, for a remaining 303 accurate Gaussian approximations. This corresponded to a 5% error rate, which was deemed acceptable. The function that was used to calculate the metrics for the Gaussian approximations is shown in Appendix E.6.

Figure 3.26: Gaussian approximation curve.

3.2.3 Feature-level Fusion - Forward Selection Model

For each feature set obtained previously (first and second derivatives, PCA, rise and fall time, etc.), we only retained the mean value in each plot, per band, for inputs to the FSM model (see Appendix E.4 for smoothing algorithm). Six FSMs were performed, the first two included wLiDAR variables only, with and without PCA variables, the third and fourth used HSI variables only, with and without PCA variables, and the fifth and sixth were a fusion of wLiDAR and HSI independent variables, with and without PCA variables. The parameter estimates for all six cases are shown in Appendix B. The following 227 features from both sensors were used:

- HSI (24 bands) and composite wLiDAR (30 bands), for which band 14 correspond to the ground, with each previous band representing matter above ground, and each subsequent band representing potential multiple or delayed scattering;
- PCA of HSI (24 bands) and wLiDAR (30 bands);
- first derivative of HSI (24 bands) and wLiDAR (30 bands) in the x direction;
- second derivative of HSI (24 bands) and wLiDAR (30 bands) in the x direction;
- mean and sum of grass unmixing (abundance) value for each plot (2 bands);
- difference between composite waveform and road waveform (1 band);
- rise time to leading edge and fall time to trailing edge of ground peak from wLiDAR data (2 bands); and
- amplitude, sigma, and area under the curve of a maximum of two modeled Gaussian curves (6 bands).

Waveform LiDAR

The independent variables that were selected from the original wLiDAR features were the following:

1. wLiDAR intensity 0.30m above ground;
2. wLiDAR intensity at ground level;
3. wLiDAR intensity 0.15m below ground (theoretical multiple scattering);
4. wLiDAR intensity 1.05m below ground (theoretical multiple scattering); and
5. first derivative of wLiDAR intensity at 0.15m below assumed ground.

The linear combination of those variables that best modeled the herbaceous biomass are shown in Appendix B, for a fixed significance level of 0.12. This number was obtained by trial and error in order to obtain from four to eight different feature outputs from the FSM. Figure 3.27 shows the resulting images for features that were selected in the forward selection procedure for wLiDAR, without PCA variables, for a specific subset in the area of interest, situated in the Bushbuckridge communal settlement. Figures 3.27(a) and (c) are the most noisy images, and represent data very close to the ground and assumed multiple scattering. Figures 3.10(d) represent information that is further from the ground, which explains why there are so many zero values.

Figure 3.27: Selected features for original wLiDAR bands in the case of herbaceous biomass modeling: a) wLiDAR $0.30m$ above ground, b) wLiDAR at ground level, c) wLiDAR $0.15m$ below assumed ground, d) wLiDAR $1.05m$ below assumed ground, and e) first derivative of wLiDAR at $0.15m$ below assumed ground.

The independent variables that were selected from the wLiDAR-related PCA features were the following:

1. component 1 of the PCA;
2. component 19 of the PCA;
3. component 28 of the PCA; and
4. amplitude of the first modeled gaussian for each composite waveform.

The linear combination of those variables that best modeled the herbaceous biomass are shown in Appendix B, for a fixed significance level of 0.05. Figure 3.28 shows the resulting images that were selected in the forward selection procedure from PCA wLiDAR variables, for the same area, except for the amplitude of the first modeled gaussian, as this was only computed for each composite waveform. Figures 3.28(b) and (c) contain close to zero percent variance contributions compared to the first selected feature, which explained 72% of the variance.

Figure 3.28: Selected PCA features for original wLiDAR bands in the case of herbaceous biomass modeling: a) component 1 of the PCA, b) component 19 of the PCA, and c) component 28 of the PCA. Note the lack of vegetation features in (b) and (c).

HSI

The independent variables that were selected from the original HSI features were the following:

1. reflectance at $394nm$;
2. reflectance at $422nm$;
3. reflectance at $508nm$;
4. reflectance at $564nm$;
5. reflectance at $763nm$;
6. reflectance at $791nm$;
7. reflectance at $876nm$;
8. reflectance at $989nm$; and
9. first derivative of reflectance at $791nm$.

The best linear combination of those variables for herbaceous biomass modeling are shown in Appendix B; the significance level was fixed at 0.02, which is lower than both wLiDAR modeling attempt. The reason why more than eight features were selected is that the procedure would output either a really small number of variables, with poor modeling results, while the next α -level output resulted in nine variables. Figure 3.29 shows the resulting images of features that were selected in the forward procedure of original HSI bands, which is a mix of spectral and first derivative features.

The independent variables that were selected from the PCA-based HSI features were the following:

1. component 4 of the PCA;
2. component 6 of the PCA;
3. component 12 of the PCA;
4. component 13 of the PCA; and
5. component 19 of the PCA.

The linear combination of those variables that were used for herbaceous biomass modeling are shown in Appendix B; the significance level was fixed at 0.01, which is comparable to the previous modeling effort of HSI variables. Figure 3.30 shows the resulting images of features that were selected in the forward procedure of PCA-based HSI features. The selected variables all exhibited close to zero percent variance contributions.

Figure 3.29: Selected reflectance features for HSI in the case of herbaceous biomass modeling: a) reflectance at $394nm$, b) reflectance at $422nm$, d) reflectance at $508nm$, e) reflectance at $564nm$, f) reflectance at $763nm$, g) reflectance at $791nm$, h) reflectance at $876nm$, i) reflectance at $989nm$, and j) first derivative of reflectance at $791nm$.

Figure 3.30: Selected PCA-based reflectance features for HSI for herbaceous biomass modeling: a) component 4 of the PCA, b) component 6 of the PCA, c) component 12 of the PCA, d) component 13 of the PCA, and e) component 19 of the PCA. Note the "uniform" intensity in certain vegetation features, such as the bottom left tree versus the noisy grass response in most of the images.

Fusion of wLiDAR and HSI

The independent variables that were selected from the original fusion features were the following:

1. wLiDAR 0.3m below the assumed ground (multiple scattering);
2. reflectance at 394nm;
3. reflectance at 422nm;
4. reflectance at 508nm;
5. reflectance at 564nm;
6. reflectance at 763nm;
7. reflectance at 791nm;
8. reflectance at 876nm;
9. reflectance at 989nm;
10. second derivative of wLiDAR at 0.90m below assumed ground; and
11. first derivative of reflectance at 791nm.

The best linear combination of those variables for herbaceous biomass modeling are shown in Appendix B; the significance level was fixed at 0.015, which is lower than that for both wLiDAR and HSI modeling. The reason why more than eight features were selected is that the procedure would output either a small number of variables, with poor modeling results, while the next α -level resulted in 11 variables. Figure 3.31 shows the resulting feature images that were selected in the forward procedure of original fusion features. 18% of those variables were from wLiDAR and 82% are from HSI.

The independent variables that were selected from the PCA-based fusion features were the following:

1. component 1 of the PCA of wLiDAR;
2. component 19 of the PCA of wLiDAR;
3. component 4 of the PCA of HSI;
4. component 13 of the PCA of HSI;
5. component 18 of the PCA of HSI; and
6. component 19 of the PCA of HSI.

The linear combination of those variables that was used for herbaceous biomass modeling are shown in Appendix B; the significance level was fixed at 0.015, which was the same as the previous fusion model. Figure 3.32 shows the resulting feature images that were selected in the forward procedure of PCA-based fusion features.

Figure 3.31: Selected features for wLiDAR and HSI fusion for herbaceous biomass modeling: a) wLiDAR 0.3m below the assumed ground (multiple scattering), b) HSI at 394nm, c) HSI at 422nm, d) HSI at 508nm, e) HSI at 564nm, f) HSI at 763nm, g) HSI at 791nm, h) HSI at 876nm, i) HSI at 989nm, j) second derivative of wLiDAR at 0.90m below assumed ground, and k) first derivative of HSI at 791nm.

Figure 3.32: Selected features for PCA-based fusion of wLiDAR and HSI: a) component 1 of the PCA of wLiDAR, b) component 19 of the PCA of wLiDAR, c) component 4 of the PCA of HSI, d) component 13 of the PCA of HSI, e) component 18 of the PCA of HSI, and f) component 19 of the PCA of HSI.

3.2.4 Herbaceous Biomass Modeling Results

Figures 3.33, 3.35, and 3.37 show the predicted versus the field measured data for original wLiDAR, HSI, and the fusion of both data sets. Figures 3.34, 3.36, and 3.38 show the predicted versus the field measured data for wLiDAR, HSI, and both data sets, with PCA variables. The field measured biomass was sorted into $0.005kg$ group intervals in order to obtain a better fit, i.e. this would act as a smoothing action, but results did not significantly differ. Note that most of these figures seem to exhibit an exponential curve fit, however no transformation was applied to this modeled herbaceous biomass, so a linear trend line is shown. Variable transformations, e.g. squared, square-root, and log were attempted, but none resulted in improved results.

Waveform LiDAR

The model for obtaining predicted HB (y_p) from variables selected in the original wLiDAR FSM is $y_p = 166.30 \cdot H12 - 85.87 \cdot H14 + 119.24 \cdot H15 - 1054.31 \cdot H21 + 133.75 \cdot J15 + 22.65$, where H12 is wLiDAR $0.30m$ above ground, H14 is wLiDAR at ground level, H15 is wLiDAR $0.15m$ below assumed ground, H21 is wLiDAR $1.05m$ below assumed ground, and J15 is the first derivative of wLiDAR at $0.15m$ below assumed ground. The RMSE was $0.08kg/m^2$ and the coefficient of determination was $R^2 = 0.07$.

Figure 3.33: Graph of predicted versus field measured data for original wLiDAR data.

The model for obtaining predicted HB (y_p) from variables selected in the PCA-based wLiDAR FSM is $y_p = -0.03 \cdot I1 + 4.54 \cdot I19 - 19.07 \cdot I28 - 0.45 \cdot M1 + 28.51$, where I1 is component 1 of the PCA of wLiDAR, I19 is component 19 of the PCA of wLiDAR, I28 is component 28 of the PCA of wLiDAR, and M1 is the amplitude of the first modeled Gaussian for each composite waveform. The RMSE was 0.08 kg/m^2 and the coefficient of determination was $R^2 = 0.12$.

Figure 3.34: Graph of predicted versus field measured data for PCA-based wLiDAR data. Note how there seems to be an accentuated exponential fitting here compared to the previous figure, although the R^2 value increased.

HSI

The model for obtaining predicted HB (y_p) from variables selected in the HSI FSM is $y_p = 0.01 \cdot C1 - 0.72 \cdot C2 + 0.77 \cdot C5 - 0.29 \cdot C7 + 0.26 \cdot C14 - 0.24 \cdot C15 + 0.03 \cdot C18 - 0.03 \cdot C22 + 0.02 \cdot D15 + 79.54$, where C1 is reflectance at $394nm$, C2 is reflectance at $422nm$, C5 is reflectance at $508nm$, C7 is reflectance at $564nm$, C14 is reflectance at $763nm$, C15 is reflectance at $791nm$, C18 is reflectance at $876nm$, C22 is reflectance at $989nm$, and D15 the first derivative of reflectance at $791nm$. The RMSE was $0.07kg/m^2$ and the coefficient of determination was $R^2 = 0.32$.

Figure 3.35: Graph of predicted versus field measured data for original HSI data.

The model for obtaining predicted HB (y_p) from variables selected in the PCA-based HSI FSM is $y_p = 0.04 \cdot F4 + 0.06 \cdot F6 + 0.24 \cdot F12 - 0.40 \cdot F13 - 1.14 \cdot F19 + 54.67$, where F4 is component 4 of the PCA of HSI, F6 is component 6 of the PCA of HSI, F12 is component 12 of the PCA of HSI, F13 is component 13 of the PCA of HSI, and F19 is component 19 of the PCA of HSI. The RMSE was 0.07 kg/m^2 and the coefficient of determination was $R^2 = 0.27$. It is interesting to notice that both HSI models predict negative biomass results, while wLiDAR does not. This was attributed to (i) the wLiDAR being an inherently structural data source, and (ii) the narrow range and low herbaceous biomass field data.

Figure 3.36: Graph of predicted versus field measured data for PCA-based HSI data.

Fusion of wLiDAR and HSI

The model for obtaining predicted HB (y_p) from variables selected in the original fusion FSM is $y_p = 156.21 \cdot H16 + 0.001 \cdot C1 - 0.74 \cdot C2 + 0.85 \cdot C5 - 0.34 \cdot C7 + 0.25 \cdot C14 - 0.21 \cdot C15 + 0.01 \cdot C18 - 0.02 \cdot C22 + 2014.41 \cdot K20 + 0.03 \cdot D15$, where H16 is wLiDAR 0.3m below the assumed ground (multiple scattering), C1 is reflectance at 394nm, C2 is reflectance at 422nm, C5 is reflectance at 508nm, C7 is reflectance at 564nm, C14 is reflectance at 763nm, C15 is reflectance at 791nm, C18 is reflectance at 876nm, C22 is reflectance at 989nm, K20 is the second derivative of wLiDAR at 0.90m below assumed ground, D15 the first derivative of reflectance at 791nm. The RMSE was 0.07 kg/m^2 and the coefficient of determination was $R^2 = 0.35$.

Figure 3.37: Graph of predicted versus field measured data for original fusion data.

The model for obtaining predicted HB (y_p) from variables selected in the PCA-based fusion FSM is $y_p = -0.02 \cdot I1 + 4.74 \cdot I19 + 0.04 \cdot F4 - 0.31 \cdot F13 - 0.89 \cdot F18 - 0.88 \cdot F19 + 49.89$, where I1 is component 1 of the PCA of wLiDAR, I19 is component 19 of the PCA of wLiDAR, F4 is component 4 of the PCA of HSI, F13 is component 13 of the PCA of HSI, F18 is component 18 of the PCA of HSI, and F19 is component 19 of the PCA of HSI. The RMSE was 0.07 kg/m^2 and the coefficient of determination was $R^2 = 0.29$.

Figure 3.38: Graph of predicted versus field measured data for PCA-based fusion data.

The resulting R^2 values and RMSE showed parallel results. Keeping in mind that a high R^2 and a low RMSE are indicative of a better model, then the best results occurred with original fusion variables, with the highest R^2 of 0.35 and the lowest RMSE of 0.07 kg/m^2 . This was followed by the modeling of original HSI variables with an R^2 of 0.32 and an RMSE of 0.07 kg/m^2 . This could be due to the fact that those FSMs are the only ones for which a higher number of variables was selected for the modeling. For PCA-based variables, the fusion was only marginally better than HSI with an R^2 of 0.29 versus 0.27, respectively, and an RMSE of 0.07 kg/m^2 for both. The results were the lowest with wLiDAR modeling with an R^2 of 0.07 for original variables and 0.12 for PCA-based variables. The RMSE was 0.08 kg/m^2 for both wLiDAR classifications, which was lower than for HSI and the fusion modeling.

Although the modeling with original fusion variables was better than HSI and wLiDAR, the results could not be employed to accurately predict biomass in this savanna environment. One source of error with this data set lies with the limited biomass range values that made it difficult to notice a trend in the result. Also, the outgoing pulse width that was chosen for the wLiDAR instrument was too broad for this application (15 ns) as the resulting waveform signal was too coarse and prevented the capture of intricate details in the vegetation [J. Wu, unpb]. The fact that

grass was measured in a senescent state furthermore did not help the process of correctly modeling this vegetation type, especially with HSI. This was due to the collection of data in the fall season in South Africa. Not only does senescence dampen the vegetation signal due to low pigment content, but moisture levels are also lower than in the peak growing season. These facts effectively resulted in a vague herbaceous biomass signal. However, it was interesting to note that close-to-the-ground metrics (lower than 1m) were selected as significant variables in the original wLiDAR FSM as well as the first PCA of wLiDAR and that half of the spectral features from original HSI FSM were from the red-edge and near infrared regions, which are indicators of vegetation structural and spectral signals. This shows that wLiDAR and HSI have the potential for herbaceous biomass modeling, but results would be more conclusive under ideal physiological conditions.

Different transformations of the data were attempted in order to improve modeling results, such as log and square transformations, but the results were once again very comparable to those found in this section. Another source of error on the computing side resides in the fact that calculation of composite waveforms may smooth the already coarse waveform information too much, especially as the ground contribution could vary significantly between waveforms. The biomass modeling therefore was also performed without building composite waveforms, so that each exact GPS point would define the waveform used for feature extraction. The results were very similar to those obtained in this section. It could be argued that small footprint data are too precise, considering the positional error in GPS location, and considering the compositional variation on the ground.

The level of precision of wLiDAR for herbaceous biomass modeling as an application may depend on too many exacerbating negative factors to effectively accomplish our task. Future research could focus on modeling herbaceous biomass with wLiDAR, with some amendments most notably the use of a narrower pulse width, a region with a broader range of biomass, and non-senescent grass.

Chapter 4

Conclusion

This research evaluated the feature-level fusion of hyperspectral imagery (HSI) and waveform light detection and ranging (wLiDAR) with species classification and herbaceous biomass modeling as use cases. The remote sensing data were collected in three landuses of the Kruger National Park (KNP) and the Bushbuckridge communal settlement in South Africa, and for the purpose of this project, field reference data consisted of GPS locations of different tree species and plot-level herbaceous biomass. The first study focused on species-level classification, which consisted of preprocessing the imagery, segmenting tree regions from the area of interest, forming composite waveforms for each region of interest (ROI), and followed by salient feature extraction for each modalities, which were finally used as inputs into a stepwise discriminant analysis (SDA) performed in the statistical analysis software (SAS) (version 9.2). The results were evaluated in terms of overall classification accuracy, producer's and user's accuracies, *kappa* statistic, and z-score evaluation, which is a comparison of accuracy based on the "proportion of correct allocations". The objective was to determine if fusion of wLiDAR and HSI data achieved better classification results than single modality classification. The second study evaluated herbaceous biomass modeling, also using preprocessed data from wLiDAR and HSI, from which salient features were extracted at the plot-level for input into the forward selection model (FSM) performed in SAS (version 9.2). The end results were evaluated against field herbaceous biomass in terms of the coefficient of determination (R^2) and the root mean square error (RMSE), which are indicative of our ability to properly model dependent variable and the associated error, respectively.

The preprocessing of wLiDAR data was performed according to novel techniques, which are not currently considered as mandatory steps before extracting structural information. It is believed that those steps increase the fidelity of waveform data deconvolution and angle correction therefore were applied as part of preprocessing. This was followed by ground alignment of each waveform in the study site, which normalized data for potential topographical variations.

The features that were extracted in context of species classification consisted of the original intensity for each return in the wLiDAR signal, the reflectance values of the HSI, the first and second derivatives of the original data, along with the principal component analysis (PCA) of the original data sets. Additional tree structural metrics were calculated, such as the height and the crown thickness for every waveform. The features that were included in the SDA were the mean and maximum height for each tree region grouped as ROIs, the mean crown thickness, and the crown thickness to height ratio. The SDAs were performed separately for wLiDAR, HSI and last

but not least, the fused data set, which included a mix of features from both modalities. Two separate SDAs were performed for each modality including the fused data set, since the original intensity and reflectance values were most likely correlated to the PCA values. The first SDA on wLiDAR was performed with the original intensity data, the first and second derivatives, and the tree-based wLiDAR features. The second wLiDAR SDA was performed with PCA values, and the wLiDAR features. For the first HSI SDA, original reflectance values and first and second derivatives were included. The second HSI SDA consisted of PCA values only. The first fusion SDA included the feature outputs from the stepwise regression procedures from both wLiDAR and HSI. The second fusion SDA included the same outputs, but from the PCA-based SDAs.

The overall classification accuracies results were similar for HSI and fusion and considerably lower for wLiDAR. The best result, 74%, was obtained with the fusion of PCA-based data, followed closely by the original HSI data set with 73%. Waveform LiDAR achieved poorer results with a final classification accuracy of 53%. The *kappa* coefficients corroborated these results with the best score of 0.64 for the PCA-based fusion data set, followed by the original HSI data with 0.63 and lastly wLiDAR with 0.44. These results were not particularly good. However, they show that classification with HSI achieves results close enough to a fused classification approach and as such negates the need to include costly wLiDAR, especially considering the effort required for preprocessing and structural extraction. When one considered the user's accuracies, which reflect the percentage of correctly classified trees, the best results were almost always obtained by the fusion of PCA-based data, and in one case with the PCA-based HSI data set. The user's accuracies for wLiDAR were always the lowest, implying that our structural features were not well correlated with individual species characteristics. We concluded that data collection during the beginning of fall season in South Africa unduly influenced the within class variability and between class separability due to changing vegetation phenology. The other factor that could have negatively affected results is the broad pulse width (15ns), which is believed to influence the precision of the return signal [J. Wu, unpb]. The producer's accuracies, which describe the number of correctly classified trees per species according to the training data, delivered best results with the fused data sets (both original and PCA-based). In one case results were better with the HSI data set, but they were worse for wLiDAR in all cases. However, improvements in individual species classifications (producer's accuracies) were regarded as evidence of the potential of fusion, albeit for specific species or use cases.

The features collected for herbaceous biomass modeling consisted of the original intensity for each return of the wLiDAR signal, the reflectance values in the HSI, the first and second derivatives of the original data, and the PCA of the original data set. The PCA was slightly different than the previous use case, as different sites were used and the waveforms were cropped to limit the amount of information to approximately 2m above ground. Other features were extracted from the wLiDAR as well, namely those that related to possible herbaceous biomass indicators. These included the difference between each plot-level composite waveform and a dirt road composite waveform, the rise time of the leading edge and the fall time of the trailing edge, and finally Gaussian modeling parameters such as amplitude, standard deviation, and area under the curve. The additional HSI features that were added consisted of the linearly unmixed "grass signal" results.

Six different FSMs were applied, the first one included original wLiDAR data, such as the intensity data from each timescale of the composite waveforms, the first and second derivatives of the original intensity data, the rise time of the leading edge and the fall time of the trailing edge, the differences between each composite waveform and a dirt composite waveform, and the modeled Gaussian parameters. The second FSM included PCA-based wLiDAR data instead of

the original intensity, and included the other wLiDAR features, except for the first and second derivatives. The third FSM included original HSI data, the fourth included PCA-based HSI data, both of which also included the unmixing results from the "grass signal". The fifth FSM contained the features from both modalities associated with the original intensity and reflectance data, while the last included variables from both modalities used in the PCA-based modeling.

The FSM results showed that herbaceous biomass was only marginally correlated with the selected features. The coefficient of determination results were the best for the fusion of original data ($R^2 = 0.35$), followed by the modeling of original HSI features ($R^2 = 0.32$). The worst coefficient of determination was obtained from the original wLiDAR data ($R^2 = 0.07$). These values are low, as we would rather see values close to 1, indicating a high correlation between the selected variables. The RMSE was also calculated for each modeling case and results were the same for HSI and the fusion of both modalities (0.07), while the poorest results were obtained with wLiDAR (0.08). This performance of the six different cases of modeling exhibited a similar outcome to the species classification. The results are worse than those obtained by [Wu et al., 2009a], which was attributed to multiple factors. First, additional sites were used compared to the previous study. The herbaceous biomass that was measured as part of the field data consisted of a narrow range, such that it would be difficult to find a solid trend, especially considering the associated positional error (1-2m) due to GPS instruments. This error can affect results by up to four pixels on each side of the targeted position, which could ultimately dramatically affect the herbaceous biomass readings. South Africa's KNP is known for its savanna wildlife, which contributes to grazing of vegetation. This can result in considerable differences in the biomass assessment from one pixel to the next. The idea behind building composite waveforms was to average the biomass in order to compensate for such situations. But at such fine scales, the smoothing may have unduly affected the results. Also, there were improvements in the preprocessing algorithm from the time when [Wu et al., 2009a] performed their study and the one used during this research, especially in the how the angle correction was applied. Another difference between the two studies lies in the fact that [Wu et al., 2009a] originally retained only the single-peak waveforms before deconvolution. This study retained two-peak waveforms after deconvolution, and then modeled a cropped version of each composite waveform. The reasoning behind this method was that two peaks provided returns from the herbaceous components above ground, while the second peak could be attributed to the ground itself. [Wu et al., 2009a]'s reasoning was based on the fact that the single peak before deconvolution would carry that herbaceous information in that peak alone. In our case, the waveforms, having previously been aligned at the ground peak, were cropped at 2m above ground, where the bulk of herbaceous biomass was believed to lie. It is possible that "first peak" information, higher than 2m, was removed in this process. Another difference in this study is that a generic Gaussian modeling algorithm was used to perform composite waveform modeling of two-peaks or fewer waveforms, compared to the expectation maximization algorithm that was used in [Wu et al., 2009a]'s research. However, the generic Gaussian modeling algorithm performed reasonably well considering the 5% error rate.

Waveform LiDAR is a promising new technology that requires a better understanding by the community, both from the preprocessing and structural metric extraction point of view. It would be interesting to evaluate results from a similar approach and field data collected at the end of spring season, in order to capture wLiDAR and HSI of vegetation during the peak growing season. It could also be interesting to perform the same modeling in a controlled environment, in order to study how physical changes in slope, vegetation, and instrument settings directly affect waveform shape and amplitude. This experiment could allow the user to understand the physical interactions that

result in a specific waveform. Waveform LiDAR technology has significant potential, especially for assessment of vegetation structure as it directly measures these structural interactions, but sustained research is required to better understand the physical interactions that determine a waveform return and to make sense of these.

Appendices

Appendix A

Species Classification Tables

A.1 Stepwise Selection Procedure

Tables [A.1](#), [A.2](#), [A.3](#), [A.4](#) show summaries of the stepwise procedure produced by SAS for wLiDAR, and HSI variables with the original data metrics and with PCA variables. The features obtained in each procedure will form the variables for the discriminant analysis of the fusion data set.

A.1.1 Waveform LiDAR

Table A.1: Summary of stepwise procedure for original wLiDAR variables.

Step	Number In	Entered	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Canonical Squared Correlation	Pr > ASCC
1	1	wLiDAR Bd 97	0.2917	26.22	<.0001	0.70830841	<.0001	0.09723053	<.0001
2	2	wLiDAR Bd 102	0.1357	9.94	<.0001	0.61222546	<.0001	0.13844105	<.0001
3	3	wLiDAR Bd 48	0.1278	9.23	<.0001	0.53400581	<.0001	0.17257052	<.0001
4	4	wLiDAR Bd 124	0.1163	8.25	<.0001	0.4718979	<.0001	0.20745828	<.0001

Table A.2: Summary of stepwise procedure for PCA-based wLiDAR.

Step	Number In	Entered	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Canonical Squared Correlation	Pr > ASCC
1	1	wLiDAR PCA Bd 6	0.2554	21.83	<.0001	0.7446482	<.0001	0.08511727	<.0001
2	2	MAX H	0.1769	13.61	<.0001	0.61295265	<.0001	0.13958019	<.0001
3	3	wLiDAR PCA Bd 36	0.1265	9.13	<.0001	0.53539953	<.0001	0.17789184	<.0001
4	4	Mean H	0.1101	7.75	<.0001	0.47644112	<.0001	0.20455279	<.0001

A.1.2 HSI

Table A.3: Summary of stepwise procedure for original HSI variables.

Step	Number In	Entered	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Canonical Squared Correlation	Pr > ASCC
1	1	HSI Bd 2	0.453	52.73	<.0001	0.54696704	<.0001	0.15101099	<.0001
2	2	HSI Bd 1	0.3708	37.33	<.0001	0.34413678	<.0001	0.27036936	<.0001
3	3	HSI Bd 10	0.3108	28.41	<.0001	0.23716761	<.0001	0.3495947	<.0001
4	4	HSI FX2 Bd 14	0.1328	9.6	<.0001	0.2056601	<.0001	0.3765919	<.0001

Table A.4: Summary of stepwise procedure for PCA-based HSI variables.

Step	Number In	Entered	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Canonical Squared Correlation	Pr > ASCC
1	1	HSI PCA 5	0.4701	56.48	<.0001	0.52991717	<.0001	0.15669428	<.0001
2	2	HSI PCA 14	0.3468	33.63	<.0001	0.34612505	<.0001	0.26855409	<.0001
3	3	HSI PCA 2	0.2896	25.69	<.0001	0.24587804	<.0001	0.33611954	<.0001
4	4	HSI PCA 20	0.1607	12	<.0001	0.20635872	<.0001	0.3771149	<.0001

A.2 Discriminant Analysis

Tables A.5, A.6, A.7, A.8, A.9, and A.10 show the linear combinations of coefficients generated by the DA model in order to obtain species classification from wLiDAR, HSI, and combined wLiDAR and HSI variables with the original and PCA-based data metrics.

A.2.1 Waveform LiDAR

Table A.5: Linear combination of coefficients for species classification - original wLiDAR variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-8.71718	-4.46267	-5.84951	-7.72932
wLiDAR Bd 97	0.21688	0.11953	0.18647	0.18659
wLiDAR Bd 102	0.06627	0.04864	0.0419	0.06731
wLiDAR Bd 48	0.07617	0.15272	0.05088	0.03329
wLiDAR Bd 124	-0.04241	0.09125	0.45033	0.16396

Table A.6: Linear combination of coefficients for species classification - PCA-based wLiDAR variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-5.56124	-4.53455	-3.87681	-5.4552
wLiDAR PCA 6	-0.15201	-0.06804	-0.10257	-0.1459
Max H	0.39969	0.68433	0.16848	0.36247
wLiDAR PCA 36	0.09594	0.22489	0.52341	0.30457
Mean H	0.11834	-0.38059	0.41286	0.09811

A.2.2 HSI

Table A.7: Linear combination of coefficients for species classification - original HSI variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-12.86749	-19.83666	-25.23093	-11.24873
HSI Bd 2	-0.10039	0.11442	-0.10562	-0.1047
HSI Bd 1	0.17688	0.00488	0.26672	0.21667
HSI Bd 10	0.01398	-0.00711	-0.00452	-0.00289
HSI FX2 Bd 14	-0.00813	0.00139	-0.01041	-0.00495

Table A.8: Linear combination of coefficients for species classification - PCA-based HSI variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-8.1078	-18.42015	-22.07525	-10.32164
HSI PCA 5	0.05463	0.07863	0.09411	0.06514
HSI PCA 14	-0.07466	-0.20681	-0.06695	-0.0985
HSI PCA 2	-0.00143	-0.0047	-0.00365	-0.0000225
HSI PCA 20	0.74054	0.72183	1.03836	0.56836

A.2.3 Fusion of wLiDAR and HSI

Table A.9: Linear combination of coefficients for species classification - original fusion variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-29.87960	-32.15119	-41.36927	-28.12683
wLiDAR Bd 97	0.29022	0.21378	0.28420	0.26086
wLiDAR Bd 102	0.11116	0.10022	0.10955	0.12101
wLiDAR Bd 48	0.07765	0.13637	0.04523	0.02494
wLiDAR Bd 124	-0.59279	-0.48118	-0.20730	-0.34504
HSI Bd 2	-0.16613	0.05372	-0.16134	-0.17028
HSI Bd 1	0.31956	0.13128	0.39594	0.36182
HSI Bd 10	0.0008080	-0.01841	-0.01826	-0.01770
HSI FX2 Bd 14	-0.01089	-0.00267	-0.01375	-0.00884

Table A.10: Linear combination of coefficients for species classification - PCA-based fusion variables.

Variable	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
Constant	-19.25741	-27.95652	-34.06917	-20.03948
HSI PCA Bd 5	0.07465	0.09834	0.11512	0.08379
HSI PCA Bd 14	0.08715	-0.02404	0.10879	0.04591
HSI PCA Bd 2	-0.00362	-0.00647	-0.00534	-0.00175
HSI PCA Bd 20	1.07863	1.06149	1.40623	0.89177
wLiDAR PCA 6	-0.22102	-0.16202	-0.20376	-0.20592
MAX H	1.07242	1.24507	1.10967	0.91589
wLiDAR PCA 36	0.13193	0.16713	0.55636	0.36558
Mean H	-1.05126	-1.31784	-1.17470	-0.98783

A.3 Classification Results

The classification results are shown in the following tables for original and PCA-based wLiDAR, HSI, and fusion of both modalities. The tree IDs refer to positional GPS data of known tree species locations indicated in Appendix C. The classes are identified by the following numbers:

1. *Combretum hereroense*;
2. *Sclerocarya birrea*;
3. *Terminalia sericea*; and
4. *Peltophorum africanum*.

A.3.1 Waveform LiDAR

Table A.11: Classification results for species classification -original wLiDAR variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
1	2	1 *	0.356	0.085	0.2795	0.2796
3	1	2 *	0.2362	0.4677	0.113	0.1831
4	1	4 *	0.4257	0.052	0.0847	0.4377
6	2	3 *	0.1827	0.15	0.3394	0.328
7	3	1 *	0.6081	0.0187	0.1114	0.2617
9	2	3 *	0.2368	0.1045	0.3923	0.2664
10	1	3 *	0.2808	0.0774	0.4202	0.2215
13	3	2 *	0.2483	0.6921	0.0155	0.0442
14	3	1 *	0.4841	0.0475	0.0933	0.3751
15	3	1 *	0.5417	0.0283	0.0744	0.3556
16	2	1 *	0.6695	0.0015	0.1126	0.2164
20	1	4 *	0.3252	0.2775	0.0533	0.3439
22	1	4 *	0.2742	0.2218	0.0722	0.4318
23	4	1 *	0.4904	0.0161	0.0073	0.4862
26	2	4 *	0.2818	0.0806	0.0145	0.6231
33	3	1 *	0.4363	0.2899	0.0952	0.1786
37	2	3 *	0.2521	0.1446	0.3827	0.2207
38	1	3 *	0.137	0.0371	0.5802	0.2456
39	2	1 *	0.3923	0.0663	0.2121	0.3292
40	3	1 *	0.7751	0.002	0.0752	0.1478
41	2	3 *	0.1791	0.1393	0.4864	0.1951
43	4	2 *	0.2547	0.4699	0.1952	0.0802
46	4	3 *	0.2159	0.0708	0.4271	0.2862
47	1	3 *	0.3111	0.0773	0.3628	0.2488
49	4	1 *	0.4122	0.0283	0.3105	0.249
50	1	2 *	0.0546	0.9198	0.0157	0.01
52	1	2 *	0.1333	0.4557	0.2117	0.1993
57	2	1 *	0.4538	0.034	0.0792	0.433
60	2	4 *	0.1735	0.2405	0.2288	0.3572
61	1	4 *	0.3509	0.0791	0.0365	0.5335
62	1	4 *	0.393	0.0327	0.031	0.5433
63	1	4 *	0.3631	0.0585	0.1475	0.4309

Table A.12: Classification results for species classification - original wLiDAR variables (continued).

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
64	2	4 *	0.2987	0.2621	0.1404	0.2988
67	1	2 *	0.2243	0.5126	0.1021	0.161
69	1	4 *	0.2221	0.1921	0.1582	0.4276
70	4	1 *	0.5609	0.0427	0.0016	0.3948
71	1	4 *	0.1783	0.3028	0.0715	0.4473
74	2	3 *	0.2626	0.0318	0.3687	0.3368
80	4	2 *	0.1828	0.3001	0.2292	0.288
84	1	4 *	0.2059	0.1735	0.2983	0.3223
85	1	4 *	0.2114	0.1495	0.0865	0.5526
87	1	4 *	0.3599	0.1301	0.124	0.386
94	1	4 *	0.3162	0.1219	0.1119	0.4499
98	1	4 *	0.3461	0.1134	0.1598	0.3807
101	2	4 *	0.2443	0.2037	0.0571	0.4949
103	1	4 *	0.3443	0.1585	0.1184	0.3788
104	1	2 *	0.0406	0.7715	0.1217	0.0662
107	1	4 *	0.4086	0.0713	0.0752	0.4449
108	1	4 *	0.2734	0.1505	0.0817	0.4944
110	1	4 *	0.3123	0.1419	0.1426	0.4033
112	1	4 *	0.3908	0.0865	0.0996	0.4232
114	1	2 *	0.2968	0.4375	0.1107	0.155
121	2	1 *	0.5211	0.0083	0.0132	0.4575
123	1	2 *	0.2645	0.4885	0.0651	0.1819
125	1	3 *	0.2814	0.0707	0.4107	0.2371
128	4	3 *	0.0377	0.0151	0.8587	0.0885
129	2	3 *	0.2984	0.2022	0.3319	0.1675
130	3	1 *	0.6552	0.019	0.1761	0.1497
132	4	3 *	0.2216	0.0022	0.645	0.1312
135	1	4 *	0.4148	0.0748	0.0911	0.4193
136	4	1 *	0.4037	0.0757	0.275	0.2456
139	3	2 *	0.0994	0.5233	0.1966	0.1807
146	1	3 *	0.1656	0.2203	0.4413	0.1727
147	3	2 *	0.2149	0.277	0.2343	0.2738
148	4	2 *	0.0907	0.4426	0.3717	0.0949
149	3	2 *	0.103	0.4692	0.266	0.1618
150	4	2 *	0.148	0.3742	0.3085	0.1693
151	3	2 *	0.1002	0.5094	0.2157	0.1747
152	4	2 *	0.0788	0.465	0.3741	0.0822
153	4	1 *	0.5036	0.0238	0.3399	0.1327
156	1	4 *	0.2921	0.1637	0.2399	0.3043
157	1	3 *	0.2851	0.0902	0.4291	0.1956
158	1	2 *	0.1114	0.408	0.3011	0.1795
160	1	4 *	0.3715	0.0923	0.0651	0.4712
179	3	4 *	0.1319	0.2425	0.2494	0.3763
183	3	1 *	0.317	0.1287	0.2751	0.2792
187	4	3 *	0.2913	0.0621	0.4402	0.2064
190	1	4 *	0.2198	0.2217	0.2499	0.3086
191	4	1 *	0.4136	0.0966	0.1269	0.3629
192	3	1 *	0.3885	0.1407	0.1956	0.2752
193	4	3 *	0.1226	0.0868	0.4825	0.3081
194	1	4 *	0.3075	0.1744	0.1026	0.4155

Table A.13: Classification results for species classification - PCA-based wLiDAR variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
3	1	3 *	0.22	0.0609	0.3848	0.3344
4	1	4 *	0.3406	0.0155	0.2216	0.4223
6	2	3 *	0.0413	0.0257	0.8019	0.1312
7	3	1 *	0.4743	0.3023	0.0061	0.2173
8	1	4 *	0.2541	0.0195	0.2387	0.4876
10	1	4 *	0.3602	0.0995	0.1092	0.4312
13	3	1 *	0.4488	0.1437	0.0659	0.3416
14	3	2 *	0.2037	0.6047	0.0359	0.1556
15	3	2 *	0.0151	0.947	0.0158	0.0221
16	2	4 *	0.3617	0.0162	0.1312	0.4908
20	1	2 *	0.1556	0.7057	0.018	0.1206
22	1	2 *	0.2511	0.4023	0.1524	0.1942
23	4	3 *	0.192	0.0459	0.5323	0.2298
26	2	1 *	0.2934	0.1847	0.2686	0.2533
27	2	3 *	0.092	0.3283	0.4024	0.1774
30	2	3 *	0.0121	0.1789	0.726	0.083
33	3	1 *	0.5425	0.1637	0.0124	0.2813
37	2	1 *	0.3248	0.2149	0.1456	0.3146
38	1	4 *	0.2949	0.1242	0.111	0.47
39	2	3 *	0.2174	0.0267	0.457	0.2989
40	3	4 *	0.1766	0.0136	0.3505	0.4593
43	4	1 *	0.6374	0.0366	0.0334	0.2925
44	1	4 *	0.2179	0.2277	0.1917	0.3627
45	1	2 *	0.0773	0.3871	0.3202	0.2154
46	4	2 *	0.2606	0.3287	0.142	0.2687
49	4	1 *	0.4338	0.0285	0.1202	0.4175
52	1	2 *	0.2349	0.4994	0.1059	0.1598
57	2	4 *	0.3814	0.0674	0.1054	0.4458
61	1	2 *	0.2755	0.3236	0.1596	0.2413
62	1	4 *	0.2525	0.1164	0.2924	0.3386
63	1	4 *	0.3697	0.0906	0.0799	0.4598
64	2	3 *	0.0555	0.2173	0.4312	0.2959
67	1	2 *	0.2477	0.4681	0.0545	0.2297
69	1	4 *	0.2765	0.226	0.1683	0.3293
70	4	3 *	0.2896	0.1737	0.3051	0.2316
71	1	2 *	0.1342	0.422	0.2073	0.2365
74	2	4 *	0.178	0.158	0.235	0.429
80	4	2 *	0.2271	0.4518	0.1046	0.2166
83	4	3 *	0.1297	0.1245	0.4849	0.2609
84	1	4 *	0.2745	0.1253	0.2603	0.3399
85	1	3 *	0.1995	0.1909	0.3372	0.2724
86	4	3 *	0.2055	0.0917	0.3529	0.3499
94	1	2 *	0.2305	0.4593	0.1101	0.2001
100	3	2 *	0.1113	0.3541	0.2039	0.3307

Table A.14: Classification results for species classification - PCA-based wLiDAR variables (continued).

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
102	2	4 *	0.1658	0.1807	0.2946	0.3588
104	1	2 *	0.0659	0.4147	0.3439	0.1756
108	1	2 *	0.1461	0.6332	0.0742	0.1465
110	1	2 *	0.0611	0.8042	0.0705	0.0643
121	2	1 *	0.6489	0.0409	0.058	0.2523
123	1	2 *	0.239	0.5394	0.0276	0.194
128	4	3 *	0.1185	0.0553	0.5118	0.3144
129	2	1 *	0.5453	0.0878	0.0618	0.3051
130	3	1 *	0.5556	0.0006	0.0712	0.3725
134	1	3 *	0.308	0.042	0.3344	0.3156
135	1	3 *	0.2362	0.0896	0.425	0.2492
136	4	1 *	0.576	0.0279	0.0785	0.3176
140	2	3 *	0.2469	0.0225	0.5799	0.1506
143	2	3 *	0.084	0.279	0.5231	0.1139
144	2	3 *	0.1738	0.2834	0.4042	0.1386
145	2	3 *	0.1598	0.0408	0.7014	0.0981
147	3	2 *	0.1353	0.3706	0.3246	0.1695
148	4	3 *	0.2477	0.2515	0.3185	0.1822
150	4	3 *	0.2271	0.1592	0.3779	0.2358
152	4	3 *	0.2784	0.2214	0.3008	0.1994
153	4	1 *	0.6249	0.0078	0.0725	0.2948
155	3	1 *	0.4521	0.0777	0.135	0.3351
158	1	2 *	0.2002	0.4539	0.1743	0.1716
160	1	3 *	0.0602	0.3025	0.5489	0.0884
164	1	3 *	0.2418	0.0781	0.4199	0.2602
165	1	4 *	0.3172	0.0322	0.3166	0.334
166	1	3 *	0.0719	0.0575	0.7046	0.1661
167	1	4 *	0.3239	0.0365	0.3133	0.3264
183	3	1 *	0.3213	0.0986	0.2716	0.3085
187	4	1 *	0.4425	0.1081	0.0692	0.3802
191	4	1 *	0.441	0.1678	0.0788	0.3124
192	3	1 *	0.3704	0.3514	0.0465	0.2317
193	4	2 *	0.1328	0.5058	0.1115	0.2498

A.3.2 HSI

Table A.15: Classification results for species classification -original HSI variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
3	1	4 *	0.1893	0.0252	0.0038	0.7817
4	1	4 *	0.3136	0.0392	0.003	0.6443
5	1	4 *	0.4614	0.0156	0.005	0.518
7	3	4 *	0.2374	0.1731	0.0418	0.5477
8	1	4 *	0.0601	0.0049	0.0011	0.9339
9	2	1 *	0.4318	0.224	0.0062	0.3379
11	1	4 *	0.3184	0.0024	0.0057	0.6735
14	3	4 *	0.0578	0.0911	0.0103	0.8407
15	3	4 *	0.041	0.1163	0.0072	0.8356
16	2	4 *	0.2236	0.0995	0.0023	0.6746
31	1	4 *	0.2359	0.0122	0.0054	0.7466
33	3	2 *	0.0001	0.9374	0.0622	0.0003
36	1	4 *	0.4349	0.0044	0.0153	0.5454
40	3	2 *	0.0516	0.852	0.0055	0.0909
44	1	4 *	0.201	0.0047	0.0265	0.7678
50	1	4 *	0.3717	0.158	0.0083	0.462
52	1	4 *	0.2989	0.2799	0.0553	0.366
57	2	4 *	0.2719	0.0258	0.0305	0.6718
64	2	1 *	0.4997	0.0008	0.0185	0.481
69	1	4 *	0.347	0.0497	0.0398	0.5635
74	2	1 *	0.4165	0.2397	0.0463	0.2975
77	2	4 *	0.1956	0.2602	0.0062	0.5379
80	4	2 *	0.0549	0.7626	0.0439	0.1387
97	2	4 *	0.0737	0.2607	0.1017	0.5639
104	1	3 *	0.0762	0.3443	0.3718	0.2077
114	1	4 *	0.4348	0.005	0.0074	0.5528
121	2	1 *	0.5728	0.1273	0.0051	0.2947
122	2	4 *	0.2218	0.0589	0.0097	0.7096
125	1	4 *	0.0722	0.1647	0.0487	0.7144
126	1	3 *	0.0078	0.4809	0.4837	0.0276
127	3	4 *	0.0646	0.111	0.1947	0.6296

Table A.16: Classification results for species classification - original HSI variables (continued).

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
128	4	1 *	0.4693	0.0076	0.2048	0.3183
134	1	4 *	0.3903	0.1231	0.0125	0.4741
135	1	4 *	0.1608	0.0391	0.0052	0.7949
137	2	3 *	0.0001	0.4585	0.5395	0.002
139	3	1 *	0.5295	0.0307	0.24	0.1998
140	2	4 *	0.1668	0.1252	0.1159	0.5921
142	2	3 *	0.1043	0.0015	0.8182	0.076
143	2	4 *	0.1209	0.0611	0.3931	0.4249
144	2	4 *	0.0633	0.0186	0.3402	0.5779
145	2	4 *	0.0638	0.0038	0.4404	0.4921
146	1	4 *	0.279	0.0046	0.1321	0.5843
147	3	4 *	0.2443	0.3042	0.1093	0.3422
149	3	1 *	0.4852	0.0005	0.1616	0.3527
152	4	1 *	0.4893	0.0115	0.0958	0.4033
155	3	2 *	0.0467	0.7998	0.1261	0.0275
156	1	3 *	0.0469	0.0194	0.9188	0.0149
157	1	4 *	0.4105	0.0306	0.0592	0.4997
159	1	3 *	0.3333	0.0168	0.4412	0.2088
163	1	3 *	0.3367	0.0011	0.4806	0.1816
190	1	4 *	0.4148	0.0682	0.044	0.4729
194	1	4 *	0.228	0.0128	0.0179	0.7414

Table A.17: Classification results for species classification -PCA-based HSI variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
2	1	4 *	0.3765	0.0152	0.0017	0.6065
3	1	4 *	0.1736	0.0589	0.1072	0.6603
4	1	4 *	0.1388	0.0203	0.3173	0.5236
7	3	4 *	0.1878	0.134	0.0365	0.6418
9	2	1 *	0.6276	0.0582	0.0169	0.2973
11	1	4 *	0.2848	0.0054	0.001	0.7088
13	3	2 *	0.0223	0.4815	0.2006	0.2957
14	3	4 *	0.0421	0.0384	0.0475	0.872
15	3	4 *	0.0417	0.1202	0.1224	0.7157
16	2	4 *	0.3037	0.1444	0.0952	0.4568
31	1	4 *	0.3937	0.0134	0.0508	0.5421
33	3	2 *	0.0002	0.9042	0.0941	0.0015
40	3	4 *	0.4191	0.0526	0.0482	0.4801
44	1	4 *	0.389	0.0117	0.011	0.5883
45	1	2 *	0.1196	0.8031	0.0064	0.071
50	1	4 *	0.0862	0.0029	0.0023	0.9086
52	1	4 *	0.2234	0.1122	0.0052	0.6592
57	2	1 *	0.5708	0.0503	0.0186	0.3602
64	2	1 *	0.586	0.0014	0.0047	0.408
69	1	4 *	0.1633	0.0105	0.0006	0.8256
70	4	1 *	0.8575	0.0185	0.0173	0.1067
76	1	4 *	0.3648	0.0031	0.0001	0.632
77	2	4 *	0.2437	0.0376	0.001	0.7177
79	4	1 *	0.5633	0.0136	0.0031	0.42
80	4	2 *	0.0911	0.6558	0.0648	0.1883
84	1	4 *	0.2609	0.0259	0.0273	0.6859
94	1	4 *	0.4803	0.0042	0.0014	0.5141
99	2	1 *	0.7727	0.0308	0.0113	0.1853
104	1	2 *	0.0217	0.821	0.1131	0.0442
121	2	4 *	0.2574	0.0041	0.0045	0.734
122	2	4 *	0.2476	0.0945	0.0211	0.6368
125	1	4 *	0.1527	0.4051	0.0301	0.412
126	1	2 *	0.0206	0.6441	0.2498	0.0855
127	3	2 *	0.0675	0.6195	0.1665	0.1465
134	1	4 *	0.2175	0.0175	0.1791	0.5859
136	4	1 *	0.6244	0.0625	0.0211	0.2921
137	2	3 *	0	0.3118	0.6881	0.0001
138	2	3 *	0.0005	0.4826	0.516	0.001
140	2	3 *	0.2635	0.0967	0.5728	0.067
141	2	3 *	0.0002	0.1897	0.8099	0.0002
142	2	3 *	0.2949	0.0884	0.5756	0.0411
143	2	1 *	0.4164	0.0525	0.4024	0.1287
144	2	3 *	0.0893	0.0862	0.8008	0.0237
145	2	3 *	0.2517	0.0518	0.6299	0.0667
146	1	4 *	0.082	0.0055	0.2856	0.6269
148	4	3 *	0.2363	0.009	0.4608	0.2939
149	3	4 *	0.1437	0.0024	0.4005	0.4534

Table A.18: Classification results for species classification - PCA-based HSI variables (continued).

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
150	4	3 *	0.1154	0.0246	0.6823	0.1778
152	4	1 *	0.5871	0.0986	0.1491	0.1652
153	4	1 *	0.3265	0.0991	0.2655	0.3089
156	1	3 *	0.0316	0.2242	0.7076	0.0365
157	1	3 *	0.2792	0.0301	0.4313	0.2594
161	1	4 *	0.2794	0	0.002	0.7186
165	1	3 *	0.1807	0.003	0.7993	0.017
166	1	3 *	0.3265	0.0017	0.6616	0.0102
190	1	4 *	0.2339	0.0909	0.0241	0.6511
194	1	4 *	0.4289	0.0977	0.0308	0.4426

A.3.3 Fusion of wLiDAR and HSI

Table A.19: Classification results for species classification -original fusion variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
3	1	4 *	0.2333	0.0818	0.0037	0.6812
4	1	4 *	0.2269	0.0123	0.0027	0.7581
7	3	1 *	0.5043	0.0251	0.0421	0.4285
8	1	4 *	0.0932	0.0006	0.0011	0.9051
9	2	1 *	0.5182	0.1782	0.0107	0.2929
13	3	2 *	0.0838	0.5839	0.159	0.1734
14	3	4 *	0.068	0.0182	0.0077	0.9061
15	3	4 *	0.0591	0.0151	0.0054	0.9203
16	2	4 *	0.4786	0.0021	0.0042	0.5152
22	1	4 *	0.484	0.0062	0.0102	0.4996
31	1	4 *	0.4125	0.0063	0.0052	0.576
33	3	2 *	0.0005	0.9533	0.046	0.0002
37	2	1 *	0.4883	0.3029	0.0988	0.11
40	3	1 *	0.6887	0.1329	0.0193	0.1591
43	4	1 *	0.5833	0.1607	0.0181	0.2379
44	1	4 *	0.2688	0.002	0.0281	0.7011
49	4	1 *	0.5778	0.0038	0.0776	0.3408
50	1	2 *	0.12	0.8656	0.0007	0.0138
52	1	2 *	0.2282	0.5174	0.0307	0.2237
57	2	4 *	0.2305	0.0033	0.0205	0.7457
64	2	4 *	0.4372	0.0007	0.0187	0.5434
69	1	4 *	0.2098	0.0342	0.0298	0.7262
74	2	1 *	0.4633	0.0547	0.1061	0.3758
80	4	2 *	0.0423	0.7886	0.0356	0.1335
84	1	4 *	0.3822	0.081	0.0182	0.5185
104	1	2 *	0.0246	0.7699	0.1344	0.071

Table A.20: Classification results for species classification - original fusion variables (continued).

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
108	1	4 *	0.4453	0.0012	0.0082	0.5453
121	2	1 *	0.5221	0.0087	0.0027	0.4665
122	2	4 *	0.2916	0.2249	0.0072	0.4764
125	1	4 *	0.1251	0.0786	0.0904	0.7059
126	1	3 *	0.0168	0.0357	0.9284	0.0192
127	3	4 *	0.0174	0.0617	0.2966	0.6243
128	4	3 *	0.0761	0.0009	0.7538	0.1692
132	4	3 *	0.1032	0	0.6766	0.2201
135	1	4 *	0.1292	0.0139	0.0037	0.8532
136	4	1 *	0.6164	0.0217	0.0255	0.3364
139	3	1 *	0.4738	0.1376	0.1468	0.2417
142	2	3 *	0.12	0.0085	0.804	0.0675
144	2	4 *	0.0925	0.2106	0.2688	0.4281
145	2	4 *	0.0807	0.0211	0.3856	0.5126
146	1	4 *	0.3488	0.0081	0.1548	0.4884
147	3	2 *	0.2189	0.3906	0.0772	0.3133
149	3	4 *	0.3745	0.0037	0.1091	0.5127
151	3	4 *	0.2345	0.0633	0.2589	0.4433
152	4	1 *	0.6315	0.0545	0.0988	0.2152
153	4	3 *	0.3046	0.0024	0.4462	0.2469
155	3	2 *	0.1045	0.7195	0.1533	0.0227
156	1	3 *	0.0862	0.0147	0.8799	0.0192
158	1	2 *	0.4149	0.4288	0.0158	0.1405
163	1	3 *	0.4491	0	0.4592	0.0916
190	1	4 *	0.4071	0.0754	0.0413	0.4762
194	1	4 *	0.1597	0.0079	0.0096	0.8228

Table A.21: Classification results for species classification -PCA-based fusion variables.

Tree ID	From Class	Classified into Class	<i>Combretum hereroense</i>	<i>Sclerocarya birrea</i>	<i>Terminalia sericea</i>	<i>Peltophorum africanum</i>
3	1	4 *	0.1101	0.0268	0.2008	0.6623
4	1	4 *	0.0861	0.0025	0.3285	0.5829
7	3	1 *	0.4186	0.2294	0.0133	0.3388
8	1	4 *	0.2213	0.0011	0.0083	0.7693
9	2	1 *	0.7345	0.077	0.0132	0.1753
11	1	4 *	0.3345	0.0004	0.0005	0.6647
13	3	2 *	0.0449	0.4596	0.2094	0.286
14	3	4 *	0.0543	0.1496	0.0491	0.7469
15	3	2 *	0.01	0.5656	0.0786	0.3457
16	2	4 *	0.2571	0.0106	0.0918	0.6405
33	3	2 *	0.0006	0.9553	0.0433	0.0008
40	3	4 *	0.1818	0.0038	0.0521	0.7624
43	4	1 *	0.7701	0.0016	0.0011	0.2273
44	1	4 *	0.2612	0.0108	0.0197	0.7083
45	1	2 *	0.0655	0.7657	0.0119	0.1568
50	1	4 *	0.1486	0.0061	0.0035	0.8417
52	1	4 *	0.234	0.3133	0.0016	0.4511
57	2	1 *	0.5749	0.0143	0.0149	0.3959
64	2	4 *	0.1926	0.0005	0.071	0.7359
69	1	4 *	0.1157	0.0087	0.0005	0.8751
70	4	1 *	0.8269	0.0183	0.0077	0.1471
77	2	4 *	0.3339	0.0645	0.0007	0.6008
80	4	2 *	0.0792	0.7527	0.0417	0.1264
84	1	4 *	0.2103	0.0184	0.0331	0.7382
99	2	1 *	0.7361	0.05	0.0187	0.1952
104	1	2 *	0.0067	0.7286	0.2349	0.0297
121	2	4 *	0.3983	0.0018	0.0006	0.5993
122	2	4 *	0.2776	0.1838	0.0271	0.5115
125	1	4 *	0.2679	0.1497	0.0091	0.5734
126	1	4 *	0.1308	0.1212	0.299	0.449
127	3	4 *	0.0674	0.2599	0.1002	0.5725
134	1	4 *	0.1037	0.0047	0.0528	0.8388
135	1	4 *	0.3048	0.0422	0.0615	0.5915
136	4	1 *	0.7864	0.0135	0.0043	0.1957
140	2	1 *	0.5852	0.0555	0.2853	0.0739
143	2	3 *	0.3042	0.128	0.419	0.1488
144	2	3 *	0.1361	0.2788	0.5578	0.0273
145	2	1 *	0.5192	0.0607	0.3431	0.077
146	1	4 *	0.0872	0.004	0.1616	0.7472
148	4	3 *	0.2542	0.0238	0.4033	0.3187
149	3	4 *	0.0802	0.0046	0.1484	0.7668
150	4	3 *	0.1031	0.0326	0.6686	0.1957
151	3	4 *	0.1395	0.1102	0.1879	0.5624
152	4	1 *	0.6387	0.1393	0.067	0.155
153	4	1 *	0.6646	0.0168	0.0842	0.2344
156	1	3 *	0.0867	0.3665	0.477	0.0699
161	1	4 *	0.3452	0	0.0009	0.6539
165	1	3 *	0.3536	0.0029	0.6223	0.0212
166	1	3 *	0.1578	0.0015	0.8296	0.0112
190	1	4 *	0.3079	0.0911	0.014	0.587

Appendix B

Herbaceous Biomass Modeling Tables

B.1 Forward Selection Model

This appendix groups the output of the SAS program. It contains the variables selected during each forward selection procedure, as well as the parameter estimates for each of those variables. Also included is DF, the degree of freedom associated with each feature, the standard error for each parameter estimate, the t-value, which is a statistic testing whether the selected feature is zero, calculated as the parameter estimate divided by it's standard error, $\Pr > |t|$ is the probability of obtaining an absolute t value greater than the one obtained, given that the true parameter is zero. A value close to zero is proof that the parameter itself should not be zero. The variance inflation factor is an index of how related the variables are to each other. [SAS Institute Inc., 2008]

B.1.1 Waveform LiDAR

Table B.1: Parameter estimates for herbaceous biomass modeling - original wLiDAR variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	$\Pr > t $	Variance Inflation
Intercept	1	22.64787	12.61382	1.8	0.0736	0
wLiDAR Bd 12	1	166.3025	67.03328	2.48	0.0137	1.17867
wLiDAR Bd 14	1	-85.87346	29.564	-2.9	0.004	1.93658
wLiDAR Bd 15	1	119.23812	46.38014	2.57	0.0106	1.96306
wLiDAR Bd 21	1	-1054.3143	367.36847	-2.87	0.0044	1.25547
wLiDAR FX Bd 15	1	133.75412	77.24983	1.73	0.0844	1.03054

Table B.2: Parameter estimates for herbaceous biomass modeling - PCA-based wLiDAR variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	Pr > t	Variance Inflation
Intercept	1	28.51251	6.89888	4.13	.0001	0
wLiDAR PC_1	1	-0.03224	0.00622	-5.18	.0001	1.04463
wLiDAR PC_{19}	1	4.53629	1.9829	2.29	0.0229	1.01148
wLiDAR PC_{28}	1	-19.07412	7.56906	-2.52	0.0123	1.01638
Amplitude Gauss 1	1	-0.44862	0.19983	-2.25	0.0255	1.05182

B.1.2 HSI

Table B.3: Parameter estimates for herbaceous biomass modeling - original HSI variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	Pr > t	Variance Inflation
Intercept	1	79.54259	7.58058	10.49	.0001	0
HSI Bd 1	1	0.00508	0.1261	0.04	0.9679	119.47556
HSI Bd 2	1	-0.72171	0.16956	-4.26	.0001	275.73859
HSI Bd 5	1	0.7691	0.10583	7.27	.0001	279.32976
HSI Bd 7	1	-0.29	0.05677	-5.11	.0001	121.48043
HSI Bd 14	1	0.25577	0.10026	2.55	0.0113	1219.86231
HSI Bd 15	1	-0.24207	0.12128	-2	0.0469	2052.38588
HSI Bd 18	1	0.0343	0.04998	0.69	0.493	405.71103
HSI Bd 22	1	-0.02645	0.01883	-1.4	0.1612	57.24839
HSI FX Bd 15	1	0.02222	0.00816	2.72	0.0069	1.05506

Table B.4: Parameter estimates for herbaceous biomass modeling - PCA-based HSI variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	Pr > t	Variance Inflation
Intercept	1	54.66634	4.60504	11.87	.0001	0
HSI PCA Bd 4	1	0.03908	0.01107	3.53	0.0005	1.01466
HSI PCA Bd 6	1	0.05977	0.02234	2.67	0.0079	1.03605
HSI PCA Bd 12	1	0.2385	0.07217	3.3	0.0011	1.04812
HSI PCA Bd 13	1	-0.39972	0.09047	-4.42	.0001	1.22652
HSI PCA Bd 19	1	-1.13604	0.26823	-4.24	.0001	1.25292

B.1.3 Fusion of wLiDAR and HSI

Table B.5: Parameter estimates for herbaceous biomass modeling - original fusion variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	Pr > t	Variance Inflation
Intercept	1	66.17298	8.48566	7.8	.0001	0
wLiDAR Bd 16	1	156.21134	53.68973	2.91	0.0039	1.25692
HSI Bd 1	1	0.00121	0.1233	0.01	0.9922	119.48352
HSI Bd 2	1	-0.74334	0.16609	-4.48	.0001	276.75428
HSI Bd 5	1	0.84814	0.10662	7.96	.0001	296.56589
HSI Bd 7	1	-0.34143	0.05745	-5.94	.0001	130.16666
HSI Bd 14	1	0.24508	0.09838	2.49	0.0133	1228.60586
HSI Bd 15	1	-0.20996	0.11994	-1.75	0.0811	2099.74358
HSI Bd 18	1	0.01165	0.04991	0.23	0.8156	423.19124
HSI Bd 22	1	-0.02049	0.01848	-1.11	0.2684	57.65016
wLiDAR FX2 Bd 20	1	2014.41227	659.07195	3.06	0.0024	1.05691
HSI FX Bd 15	1	0.02643	0.00805	3.28	0.0011	1.07409

Table B.6: Parameter estimates for herbaceous biomass modeling - PCA-based fusion variables.

Variable	DF	Parameter Estimate	Standard Error	t-Value	Pr > t	Variance Inflation
Intercept	1	49.88876	6.83778	7.3	0.0001	0
wLiDAR PCA Bd 1	1	-0.01693	0.00592	-2.86	0.0045	1.1524
wLiDAR PCA Bd 19	1	4.73926	1.81381	2.61	0.0094	1.03051
HSI PCA Bd 4	1	0.0378	0.01137	3.32	0.001	1.09761
HSI PCA Bd 13	1	-0.30681	0.09297	-3.3	0.0011	1.32628
HSI PCA Bd 18	1	-0.88755	0.2603	-3.41	0.0007	1.29892
HSI PCA Bd 19	1	-0.87944	0.27597	-3.19	0.0016	1.35795

Appendix C

GPS Points

The GPS points that were used for species-level classification and herbaceous biomass modeling are displayed in the following tables.

C.1 GPS Points for Species Classification

Table C.1: Tree Species Classification - GPS Points.

Tree ID	MAP X	MAP Y	Latitude - S	Longitude - E	Species
1	315162.96	7255344.32	-24.80556	31.171363	Sbir
2	315235.2	7255305.12	-24.805922	31.172072	Cher
3	315262.64	7255211.6	-24.80677	31.172331	Cher
4	315259.84	7255195.92	-24.806911	31.172301	Cher
5	315212.8	7255192.56	-24.806936	31.171836	Cher
6	315163.52	7255176.32	-24.807076	31.171346	Sbir
7	315102.48	7255128.16	-24.807504	31.170736	Tser
8	315092.96	7255124.24	-24.807538	31.170642	Cher
9	315089.6	7255126.48	-24.807517	31.170609	Sbir
10	315084	7255100.16	-24.807754	31.17055	Cher
11	315070.56	7255093.44	-24.807813	31.170416	Cher
12	315281.12	7255184.16	-24.80702	31.17251	Sbir
13	315290.64	7255253.6	-24.806394	31.172614	Tser
14	315319.76	7255268.16	-24.806266	31.172904	Tser
15	315325.36	7255276.56	-24.806191	31.17296	Tser
16	315272.16	7254970.24	-24.80895	31.172393	Sbir
17	314577.2	7256925.2	-24.791219	31.16578	Cher
18	314602.96	7256936.4	-24.791121	31.166037	Sbir
19	314644.96	7256921.28	-24.791262	31.16645	Sbir
20	314644.4	7256909.52	-24.791368	31.166443	Cher
21	314628.16	7256904.48	-24.791412	31.166281	Sbir
22	314626.48	7256914	-24.791326	31.166266	Cher

Tree ID	MAP X	MAP Y	Latitude - S	Longitude - E	Species
23	314565.44	7256902.24	-24.791424	31.165661	Pafr
24	314571.04	7256872	-24.791698	31.165712	Cher
25	314536.32	7256852.96	-24.791866	31.165367	Sbir
26	314544.16	7256829.44	-24.792079	31.165441	Sbir
27	314502.16	7256818.24	-24.792175	31.165024	Sbir
28	314490.4	7256794.16	-24.792391	31.164905	Sbir
29	314507.76	7256793.04	-24.792403	31.165076	Cher
30	314704.88	7256811.52	-24.79226	31.167028	Sbir
31	314696.48	7256846.24	-24.791946	31.166949	Cher
32	314702.08	7256852.4	-24.791891	31.167005	Sbir
33	314697.04	7256861.92	-24.791804	31.166957	Tser
34	314687.52	7256872	-24.791712	31.166864	Sbir
35	314677.44	7256855.2	-24.791863	31.166762	Sbir
36	314476.4	7256732	-24.79295	31.164758	Cher
37	314482	7256731.44	-24.792956	31.164813	Sbir
38	314500.48	7256728.08	-24.792989	31.164996	Cher
39	314494.32	7256720.24	-24.793059	31.164934	Sbir
40	314507.2	7256719.12	-24.79307	31.165061	Tser
41	314512.24	7256706.24	-24.793187	31.165109	Sbir
42	314532.96	7256691.12	-24.793326	31.165312	Sbir
43	314545.28	7256761.12	-24.792696	31.165443	Pafr
44	314548.64	7256758.32	-24.792722	31.165476	Cher
45	314567.12	7256718.56	-24.793083	31.165653	Cher
46	314468.56	7256633.44	-24.793839	31.164667	Pafr
47	314473.04	7256623.36	-24.793931	31.16471	Cher
48	314507.76	7256598.16	-24.794162	31.16505	Sbir
49	314511.68	7256589.2	-24.794244	31.165088	Pafr
50	314506.64	7256583.6	-24.794294	31.165037	Cher
51	314525.68	7256573.52	-24.794387	31.165224	Sbir
52	314545.28	7256568.48	-24.794435	31.165417	Cher
53	314681.92	7256478.32	-24.795265	31.166756	Sbir
54	314694.8	7256469.92	-24.795343	31.166883	Sbir
55	314704.32	7256465.44	-24.795384	31.166976	Cher
56	314721.12	7256505.2	-24.795027	31.167148	Sbir
57	314688.64	7256569.6	-24.794442	31.166835	Sbir
58	314716.64	7256582.48	-24.794329	31.167114	Sbir
59	314679.12	7256641.84	-24.793789	31.166751	Cher
60	314700.4	7256651.92	-24.7937	31.166962	Sbir
61	314703.76	7256648	-24.793736	31.166995	Cher
62	314715.52	7256648.56	-24.793733	31.167111	Cher
63	314714.4	7256685.52	-24.793399	31.167105	Cher
64	314680.8	7256697.84	-24.793284	31.166775	Sbir
65	314679.12	7256712.4	-24.793152	31.16676	Sbir
66	314680.8	7256707.92	-24.793193	31.166776	Sbir
67	314659.52	7256760.56	-24.792715	31.166572	Cher
68	314672.4	7256777.36	-24.792565	31.166702	Sbir
69	314597.92	7256508	-24.794987	31.16593	Cher
70	314595.68	7256531.52	-24.794775	31.165911	Pafr
71	314580	7256438.56	-24.795612	31.165743	Cher
72	314585.6	7256444.16	-24.795562	31.1658	Sbir
73	314567.68	7256425.68	-24.795727	31.16562	Sbir
74	314595.12	7256404.96	-24.795917	31.165888	Sbir
75	314624.24	7256458.72	-24.795435	31.166184	Sbir

Tree ID	MAP X	MAP Y	Latitude - S	Longitude - E	Species
76	314644.96	7256456.48	-24.795458	31.166388	Cher
77	314652.24	7256445.28	-24.79556	31.166459	Sbir
78	314680.24	7256431.28	-24.79569	31.166734	Sbir
79	314673.52	7256375.84	-24.796189	31.16666	Pafr
80	314638.24	7256366.32	-24.796271	31.16631	Pafr
81	314632.64	7256357.92	-24.796346	31.166253	Pafr
82	314645.52	7256289.04	-24.796969	31.166371	Sbir
83	314647.2	7256284.56	-24.79701	31.166387	Pafr
84	314680.8	7256252.64	-24.797302	31.166715	Cher
85	314691.44	7256262.16	-24.797218	31.166822	Cher
86	314707.12	7256245.36	-24.797371	31.166975	Pafr
87	314709.92	7256236.96	-24.797447	31.167001	Cher
88	314736.8	7256684.96	-24.793407	31.167327	Sbir
89	314723.36	7256670.96	-24.793531	31.167192	Sbir
90	314740.16	7256542.16	-24.794696	31.167341	Cher
91	314747.44	7256540.48	-24.794712	31.167413	Sbir
92	314727.84	7256501.28	-24.795064	31.167214	Sbir
93	314731.2	7256501.84	-24.795059	31.167247	Sbir
94	314719.44	7256463.2	-24.795406	31.167126	Cher
95	314928.32	7256691.12	-24.793374	31.169221	Sbir
96	314941.2	7256696.72	-24.793325	31.169349	Sbir
97	314760.88	7256425.12	-24.795755	31.16753	Sbir
98	314787.2	7256435.2	-24.795667	31.167792	Cher
99	314798.96	7256418.96	-24.795815	31.167906	Sbir
100	314799.52	7256394.32	-24.796038	31.167908	Tser
101	314793.36	7256385.36	-24.796118	31.167846	Sbir
102	314796.16	7256374.16	-24.796219	31.167872	Sbir
103	314797.84	7256369.12	-24.796265	31.167888	Cher
104	314781.6	7256360.72	-24.796339	31.167727	Cher
105	314776	7256342.24	-24.796505	31.167669	Sbir
106	314810.16	7256332.72	-24.796595	31.168005	Sbir
107	314823.6	7256321.52	-24.796698	31.168137	Cher
108	314819.68	7256315.92	-24.796748	31.168097	Cher
109	314732.88	7256212.88	-24.797667	31.167225	Cher
110	314737.36	7256200	-24.797784	31.167268	Cher
111	314745.2	7256193.28	-24.797846	31.167344	Cher
112	314796.16	7256178.16	-24.797988	31.167846	Cher
113	314806.8	7256195.52	-24.797833	31.167954	Sbir
114	314825.28	7256154.64	-24.798204	31.168131	Cher
115	314869.52	7256135.6	-24.798382	31.168566	Sbir
116	314923.28	7256132.24	-24.798418	31.169097	Cher
117	314772.64	7256183.76	-24.797935	31.167614	Sbir
118	314759.2	7256187.68	-24.797898	31.167482	Sbir
119	314893.04	7256131.68	-24.79842	31.168798	Sbir
120	314893.04	7256120.48	-24.798521	31.168797	Sbir
121	314904.24	7256122.72	-24.798502	31.168908	Sbir
122	314898.64	7256107.6	-24.798638	31.16885	Sbir
123	314773.76	7256428.48	-24.795726	31.167658	Cher
124	314952.4	7256155.2	-24.798215	31.169388	Sbir
125	314453.45	7257566.22	-24.785417	31.164642	Cher
126	314452.33	7257555.02	-24.785518	31.164629	Cher
127	314448.97	7257555.58	-24.785513	31.164596	Tser
128	314437.77	7257547.74	-24.785582	31.164484	Pafr
129	314674.09	7256773.26	-24.792602	31.166718	Sbir

Tree ID	MAP X	MAP Y	Latitude - S	Longitude - E	Species
130	314660.09	7256712.22	-24.793151	31.166572	Tser
131	314646.09	7256711.1	-24.79316	31.166433	Sbir
132	314622.57	7256720.06	-24.793076	31.166202	Pafr
133	314618.65	7256730.14	-24.792984	31.166164	Sbir
134	330634.64	7248980.51	-24.864801	31.323601	Cher
135	330646.96	7248972.67	-24.864873	31.323721	Cher
136	330656.48	7248954.75	-24.865036	31.323813	Pafr
137	330612.8	7248972.11	-24.864874	31.323383	Sbir
138	329793.52	7249606.03 -	24.85906	31.315354	Sbir
139	329787.92	7249581.95	-24.859277	31.315296	Tser
140	329814.24	7249570.75	-24.859381	31.315555	Sbir
141	329753.76	7249567.39 -	24.859405	31.314956	Sbir
142	330162.56	7248806.91	-24.866316	31.318908	Sbir
143	330159.2	7248767.15	-24.866674	31.31887	Sbir
144	330152.48	7248750.91	-24.86682	31.318802	Sbir
145	330181.04	7248752.03	-24.866813	31.319084	Sbir
146	329152.32	7249926.35	-24.856097	31.309049	Cher
147	329163.52	7249911.79	-24.85623	31.309159	Tser
148	329174.16	7249910.67	-24.856241	31.309264	Pafr
149	329170.8	7249909.55	-24.856251	31.30923	Tser
150	329184.24	7249895.55	-24.856379	31.309362	Pafr
151	329176.4	7249893.87	-24.856393	31.309284	Tser
152	329181.44	7249882.67	-24.856495	31.309332	Pafr
153	329174.16	7249880.99	-24.856509	31.30926	Pafr
154	329172.48	7249881.55	-24.856504	31.309243	Tser
155	329152.88	7249882.11	-24.856497	31.30905	Tser
156	329147.84	7249880.43	-24.856511	31.309	Cher
157	329137.2	7249892.75	-24.856399	31.308896	Cher
158	329133.28	7249902.27	-24.856312	31.308858	Cher
159	362349.68	7266295.84 -	24.711658	31.639149	Cher
160	362354.16	7266300.32 -	24.711618	31.639194	Cher
161	362392.8	7266302.56	-24.711602	31.639576	Cher
162	362379.92	7266294.16	-24.711676	31.639448	Cher
163	362381.04	7266293.6	-24.711681	31.639459	Cher
164	362377.68	7266277.92	-24.711823	31.639424	Cher
165	362383.84	7266267.84	-24.711914	31.639484	Cher
166	362382.16	7266263.36	-24.711954	31.639467	Cher
167	362355.28	7266275.12 -	24.711846	31.639202	Cher
168	362680.08	7264381.2	-24.728975	31.642227	Tser
169	362685.68	7264381.76	-24.72897	31.642282	Tser
170	362688.48	7264373.92	-24.729041	31.642309	Tser
171	362680.64	7264370	-24.729076	31.642231	Tser
172	362680.08	7264368.88	-24.729086	31.642226	Tser
173	362661.6	7264364.4	-24.729125	31.642043	Tser
174	362650.96	7264353.2	-24.729225	31.641936	Tser
175	362646.48	7264353.2	-24.729225	31.641892	Tser
176	362653.2	7264334.72	-24.729392	31.641957	Tser
177	362652.64	7264321.84	-24.729508	31.64195	Tser
178	362663.84	7264316.24	-24.72956	31.64206	Tser
179	362674.48	7264330.8	-24.72943	31.642167	Tser

Tree ID	MAP X	MAP Y	Latitude - S	Longitude - E	Species
180	362675.6	7264334.16	-24.729399	31.642178	Tser
181	362679.52	7264325.2	-24.729481	31.642216	Tser
182	362704.16	7264326.32	-24.729473	31.64246	Tser
183	362694.08	7264343.68	-24.729315	31.642362	Tser
184	362695.76	7264343.68	-24.729315	31.642378	Tser
185	315042.56	7256602.08	-24.794192	31.170339	Sbir
186	314965.28	7256524.24	-24.794885	31.169565	Sbir
187	314975.92	7256505.76	-24.795053	31.169667	Pafr
188	314984.32	7256523.68	-24.794892	31.169753	Sbir
189	315007.84	7256469.36	-24.795386	31.169978	Sbir
190	315068.88	7256411.12	-24.795919	31.170574	Cher
191	315104.16	7256429.6	-24.795756	31.170925	Pafr
192	315122.08	7256436.32	-24.795698	31.171103	Tser
193	315099.12	7256365.76	-24.796332	31.170867	Pafr
194	314967.52	7256182.08	-24.797974	31.169541	Cher
195	315126.56	7256214.56	-24.7977	31.171118	Sbir

C.2 GPS Points for Herbaceous Biomass Modeling

In the following table, the Landuse ID corresponds to site-specific areas used for this study, along with the point numbers or ID given during the GPS data collection.

Table C.2: Herbaceous biomass field data - GPS Points.

ID	Map X	Map Y	Latitude	Longitude - E	Landuse ID	Point
1	314622.01	7256693.18	-24.793318	31.166193	B26040819	1
2	314631.53	7256695.42	-24.793299	31.166287	B26040819	2
3	314640.49	7256698.78	-24.79327	31.166376	B26040819	3
4	314650.01	7256701.02	-24.793251	31.166471	B26040819	4
5	314658.97	7256704.38	-24.793222	31.16656	B26040819	5
6	314668.49	7256706.62	-24.793203	31.166654	B26040819	6
7	314666.81	7256716.7	-24.793112	31.166639	B26040819	7
8	314664.01	7256726.22	-24.793025	31.166612	B26040819	8
9	314662.89	7256735.74	-24.792939	31.166602	B26040819	9
10	314661.77	7256745.26	-24.792853	31.166593	B26040819	10
11	314659.53	7256754.78	-24.792767	31.166572	B26040819	11
12	314650.57	7256752.54	-24.792786	31.166483	B26040819	12
13	314653.37	7256742.46	-24.792877	31.166509	B26040819	13
14	314656.17	7256733.5	-24.792959	31.166536	B26040819	14
15	314659.53	7256723.98	-24.793045	31.166568	B26040819	15
16	314651.13	7256711.1	-24.79316	31.166483	B26040819	17
17	314647.21	7256719.5	-24.793084	31.166445	B26040819	18
18	314644.97	7256729.58	-24.792993	31.166424	B26040819	19
19	314642.17	7256737.98	-24.792917	31.166398	B26040819	20
20	314638.25	7256747.5	-24.79283	31.16636	B26040819	21
21	314629.85	7256745.82	-24.792844	31.166277	B26040819	22
22	314630.41	7256735.74	-24.792935	31.166281	B26040819	23
23	314631.53	7256725.1	-24.793031	31.166291	B26040819	24
24	314632.65	7256715.58	-24.793118	31.166301	B26040819	25
25	314634.89	7256707.18	-24.793194	31.166322	B26040819	26
26	314627.05	7256702.7	-24.793233	31.166244	B26040819	27
27	314622.57	7256712.78	-24.793142	31.166201	B26040819	28
28	314619.77	7256722.3	-24.793055	31.166174	B26040819	29
29	314615.85	7256731.82	-24.792969	31.166137	B26040819	30
30	314611.37	7256740.22	-24.792893	31.166094	B26040819	31
31	314604.65	7256736.3	-24.792927	31.166027	B26040819	32
32	314606.89	7256726.78	-24.793013	31.166048	B26040819	33
33	314609.69	7256717.26	-24.7931	31.166074	B26040819	34
34	314612.49	7256708.3	-24.793181	31.1661	B26040819	35
35	314614.73	7256698.22	-24.793272	31.166121	B26040819	36

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
36	314409.77	7257544.94	-24.785604	31.164207	H2A28050836	1
37	314418.17	7257548.86	-24.78557	31.164291	H2A28050836	2
38	314427.69	7257552.22	-24.785541	31.164385	H2A28050836	3
39	314437.21	7257555.02	-24.785516	31.16448	H2A28050836	4
40	314446.73	7257557.26	-24.785497	31.164574	H2A28050836	5
41	314455.69	7257559.5	-24.785478	31.164663	H2A28050836	6
42	314454.57	7257569.02	-24.785392	31.164653	H2A28050836	7
43	314451.77	7257578.54	-24.785306	31.164627	H2A28050836	8
44	314447.85	7257588.62	-24.785214	31.164589	H2A28050836	9
45	314445.61	7257597.02	-24.785138	31.164568	H2A28050836	10
46	314443.37	7257606.54	-24.785052	31.164548	H2A28050836	11
47	314433.85	7257604.3	-24.785071	31.164453	H2A28050836	12
48	314436.65	7257595.34	-24.785152	31.16448	H2A28050836	13
49	314439.45	7257585.26	-24.785244	31.164506	H2A28050836	14
50	314442.81	7257575.74	-24.78533	31.164538	H2A28050836	15
51	314445.61	7257566.78	-24.785411	31.164564	H2A28050836	16
52	314436.09	7257563.42	-24.785441	31.16447	H2A28050836	17
53	314433.29	7257573.5	-24.785349	31.164443	H2A28050836	18
54	314431.05	7257583.02	-24.785263	31.164423	H2A28050836	19
55	314421.53	7257580.78	-24.785282	31.164328	H2A28050836	24
56	314424.89	7257571.82	-24.785363	31.16436	H2A28050836	25
57	314426.57	7257562.86	-24.785444	31.164376	H2A28050836	26
58	314417.61	7257560.06	-24.785469	31.164287	H2A28050836	27
59	314414.25	7257569.02	-24.785387	31.164255	H2A28050836	28
60	314410.89	7257579.1	-24.785296	31.164223	H2A28050836	29
61	314406.97	7257588.06	-24.785215	31.164185	H2A28050836	30
62	314403.61	7257597.58	-24.785128	31.164153	H2A28050836	31
63	314392.41	7257592.54	-24.785172	31.164042	H2A28050836	32
64	314395.21	7257583.02	-24.785259	31.164068	H2A28050836	33
65	314396.89	7257574.06	-24.78534	31.164084	H2A28050836	34
66	314400.81	7257563.98	-24.785431	31.164121	H2A28050836	35
67	314403.05	7257553.9	-24.785522	31.164142	H2A28050836	36
68	330192.24	7248763.79	-24.866708	31.319197	B25040823	1
69	330182.72	7248759.31	-24.866747	31.319102	B25040823	2
70	330173.2	7248757.07	-24.866767	31.319007	B25040823	3
71	330163.12	7248754.83	-24.866786	31.318907	B25040823	4
72	330154.16	7248753.15	-24.8668	31.318819	B25040823	5
73	330144.64	7248749.79	-24.866829	31.318724	B25040823	6
74	330142.4	7248760.43	-24.866733	31.318703	B25040823	7
75	330140.72	7248770.51	-24.866642	31.318688	B25040823	8
76	330139.04	7248780.03	-24.866556	31.318672	B25040823	9
77	330137.36	7248788.99	-24.866474	31.318657	B25040823	10
78	330136.24	7248799.63	-24.866378	31.318647	B25040823	11
79	330145.76	7248800.75	-24.866369	31.318741	B25040823	12
80	330147.44	7248791.79	-24.86645	31.318757	B25040823	13
81	330150.24	7248782.27	-24.866537	31.318783	B25040823	14
82	330153.6	7248772.75	-24.866623	31.318815	B25040823	15
83	330156.96	7248763.79	-24.866704	31.318848	B25040823	16
84	330165.36	7248764.91	-24.866695	31.318931	B25040823	17
85	330163.12	7248773.31	-24.866619	31.31891	B25040823	18
86	330160.88	7248783.39	-24.866528	31.318889	B25040823	19

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
87	330158.64	7248792.91	-24.866441	31.318868	B25040823	20
88	330155.28	7248804.11	-24.86634	31.318836	B25040823	21
89	330163.12	7248804.11	-24.866341	31.318913	B25040823	22
90	330165.92	7248794.59	-24.866427	31.31894	B25040823	23
91	330167.6	7248785.07	-24.866513	31.318955	B25040823	24
92	330168.16	7248775.55	-24.866599	31.31896	B25040823	25
93	330169.28	7248767.15	-24.866675	31.31897	B25040823	26
94	330169.28	7248759.87	-24.866741	31.318969	B25040823	27
95	330180.48	7248763.79	-24.866707	31.31908	B25040823	28
96	330178.8	7248774.43	-24.866611	31.319065	B25040823	29
97	330176.56	7248783.39	-24.866529	31.319044	B25040823	30
98	330174.88	7248793.47	-24.866438	31.319028	B25040823	31
99	330185.52	7248805.79	-24.866328	31.319135	B25040823	32
100	330187.2	7248795.71	-24.866419	31.319151	B25040823	33
101	330188.88	7248786.19	-24.866506	31.319166	B25040823	34
102	330191.68	7248776.67	-24.866592	31.319193	B25040823	35
103	330194.48	7248767.71	-24.866673	31.319219	B25040823	36
104	329193.76	7249888.83	-24.85644	31.309455	A020508G2	1
105	329188.72	7249896.67	-24.856369	31.309406	A020508G2	2
106	329182.56	7249905.07	-24.856293	31.309346	A020508G2	3
107	329176.4	7249912.35	-24.856226	31.309286	A020508G2	4
108	329170.8	7249920.19	-24.856155	31.309232	A020508G2	5
109	329164.64	7249927.47	-24.856088	31.309172	A020508G2	6
110	329157.92	7249919.63	-24.856158	31.309104	A020508G2	7
111	329149.52	7249914.03	-24.856208	31.30902	A020508G2	8
112	329141.68	7249907.87	-24.856263	31.308942	A020508G2	9
113	329133.28	7249902.27	-24.856312	31.308858	A020508G2	10
114	329124.32	7249901.15	-24.856321	31.308769	A020508G2	11
115	329128.24	7249891.63	-24.856408	31.308807	A020508G2	12
116	329136.08	7249896.67	-24.856363	31.308885	A020508G2	13
117	329145.04	7249901.15	-24.856324	31.308974	A020508G2	14
118	329152.88	7249906.75	-24.856274	31.309053	A020508G2	15
119	329161.84	7249912.35	-24.856225	31.309142	A020508G2	16
120	329169.68	7249905.07	-24.856291	31.309219	A020508G2	17
121	329160.72	7249900.03	-24.856336	31.309129	A020508G2	18
122	329152.32	7249894.43	-24.856385	31.309046	A020508G2	19
123	329144.48	7249889.39	-24.85643	31.308967	A020508G2	20
124	329137.76	7249883.23	-24.856485	31.3089	A020508G2	21
125	329143.92	7249874.27	-24.856566	31.30896	A020508G2	22
126	329151.76	7249880.99	-24.856507	31.309038	A020508G2	23
127	329159.6	7249886.59	-24.856457	31.309117	A020508G2	24
128	329167.44	7249892.75	-24.856402	31.309195	A020508G2	25
129	329175.84	7249897.23	-24.856363	31.309279	A020508G2	26
130	329180.88	7249891.07	-24.856419	31.309328	A020508G2	27
131	329173.04	7249884.35	-24.856479	31.309249	A020508G2	28
132	329164.64	7249879.31	-24.856523	31.309166	A020508G2	29
133	329156.24	7249874.27	-24.856568	31.309082	A020508G2	30
134	329147.28	7249871.47	-24.856592	31.308993	A020508G2	31
135	329152.32	7249861.95	-24.856678	31.309042	A020508G2	32
136	329161.28	7249866.99	-24.856634	31.309131	A020508G2	33
137	329169.68	7249871.47	-24.856594	31.309215	A020508G2	34
138	329178.08	7249877.07	-24.856545	31.309298	A020508G2	35
139	329185.92	7249882.11	-24.8565	31.309376	A020508G2	36
140	330663.2	7248965.39	-24.864941	31.323881	A020508G6	1

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
141	330658.72	7248972.67	-24.864874	31.323838	A020508G6	2
142	330653.12	7248981.07	-24.864798	31.323783	A020508G6	3
143	330646.96	7248988.91	-24.864726	31.323723	A020508G6	4
144	330641.36	7248996.19	-24.86466	31.323669	A020508G6	5
145	330636.32	7249005.15	-24.864579	31.32362	A020508G6	6
146	330626.8	7249001.79	-24.864608	31.323526	A020508G6	7
147	330617.28	7248999.55	-24.864627	31.323431	A020508G6	8
148	330608.32	7248995.07	-24.864667	31.323342	A020508G6	9
149	330599.92	7248989.47	-24.864716	31.323258	A020508G6	10
150	330593.2	7248982.75	-24.864776	31.323191	A020508G6	11
151	330599.36	7248976.59	-24.864832	31.323251	A020508G6	12
152	330608.32	7248982.75	-24.864778	31.32334	A020508G6	13
153	330616.72	7248987.23	-24.864738	31.323424	A020508G6	14
154	330624.56	7248992.83	-24.864689	31.323502	A020508G6	15
155	330634.08	7248994.51	-24.864674	31.323597	A020508G6	16
156	330640.8	7248984.99	-24.864761	31.323662	A020508G6	17
157	330631.84	7248981.63	-24.86479	31.323573	A020508G6	18
158	330622.88	7248976.59	-24.864835	31.323484	A020508G6	19
159	330613.92	7248973.79	-24.864859	31.323395	A020508G6	20
160	330606.08	7248967.63	-24.864914	31.323316	A020508G6	21
161	330611.12	7248959.23	-24.86499	31.323365	A020508G6	22
162	330620.08	7248962.03	-24.864966	31.323454	A020508G6	23
163	330629.04	7248964.83	-24.864942	31.323543	A020508G6	24
164	330639.12	7248967.63	-24.864918	31.323643	A020508G6	25
165	330648.08	7248970.43	-24.864893	31.323732	A020508G6	26
166	330653.68	7248964.83	-24.864945	31.323787	A020508G6	27
167	330645.28	7248959.23	-24.864994	31.323703	A020508G6	28
168	330638	7248953.63	-24.865044	31.32363	A020508G6	29
169	330629.04	7248949.15	-24.865083	31.323541	A020508G6	30
170	330621.76	7248944.11	-24.865128	31.323469	A020508G6	31
171	330628.48	7248936.83	-24.865195	31.323534	A020508G6	32
172	330635.2	7248943.55	-24.865135	31.323602	A020508G6	33
173	330642.48	7248950.27	-24.865075	31.323674	A020508G6	34
174	330650.32	7248955.31	-24.86503	31.323753	A020508G6	35
175	330659.84	7248958.67	-24.865001	31.323847	A020508G6	36
176	362392.24	7266247.68	-24.712097	31.639565	A1A270408	1
177	362382.72	7266247.12	-24.712101	31.639471	A1A270408	2
178	362372.64	7266247.12	-24.7121	31.639371	A1A270408	3
179	362362.56	7266246	-24.712109	31.639271	A1A270408	4
180	362353.04	7266245.44	-24.712114	31.639177	A1A270408	5
181	362342.96	7266244.88	-24.712118	31.639077	A1A270408	6
182	362341.28	7266254.96	-24.712027	31.639062	A1A270408	7
183	362339.04	7266264.48	-24.711941	31.639041	A1A270408	8
184	362337.92	7266274.56	-24.711849	31.639031	A1A270408	9
185	362336.24	7266284.64	-24.711758	31.639015	A1A270408	10
186	362334	7266293.6	-24.711677	31.638994	A1A270408	11
187	362344.08	7266294.72	-24.711668	31.639093	A1A270408	12
188	362345.76	7266284.64	-24.711759	31.639109	A1A270408	13
189	362348	7266275.12	-24.711845	31.63913	A1A270408	14
190	362349.68	7266265.6	-24.711931	31.639146	A1A270408	15
191	362350.24	7266255.52	-24.712022	31.63915	A1A270408	16

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
192	362360.88	7266256.64	-24.712013	31.639256	A1A270408	17
193	362358.64	7266266.72	-24.711922	31.639235	A1A270408	18
194	362356.96	7266276.8	-24.711831	31.639219	A1A270408	19
195	362354.72	7266286.32	-24.711745	31.639198	A1A270408	20
196	362353.04	7266296.4	-24.711654	31.639182	A1A270408	21
197	362363.68	7266296.4	-24.711655	31.639287	A1A270408	22
198	362365.36	7266286.88	-24.711741	31.639303	A1A270408	23
199	362367.6	7266277.36	-24.711827	31.639324	A1A270408	24
200	362369.28	7266267.28	-24.711918	31.63934	A1A270408	25
201	362370.96	7266257.76	-24.712004	31.639355	A1A270408	26
202	362381.04	7266257.76	-24.712005	31.639455	A1A270408	27
203	362380.48	7266268.4	-24.711909	31.639451	A1A270408	28
204	362379.92	7266278.48	-24.711818	31.639446	A1A270408	29
205	362379.36	7266288.56	-24.711727	31.639441	A1A270408	30
206	362378.24	7266298.08	-24.711641	31.639431	A1A270408	31
207	362390	7266298.08	-24.711642	31.639548	A1A270408	32
208	362390	7266288.56	-24.711728	31.639547	A1A270408	33
209	362390.56	7266279.04	-24.711814	31.639551	A1A270408	34
210	362391.12	7266268.96	-24.711905	31.639556	A1A270408	35
211	362391.68	7266259.44	-24.711991	31.63956	A1A270408	36
212	362602.24	7265578.48	-24.718158	31.641575	A3A270408	1
213	362595.52	7265571.2	-24.718223	31.641508	A3A270408	2
214	362589.36	7265563.92	-24.718288	31.641446	A3A270408	3
215	362582.64	7265556.64	-24.718353	31.641379	A3A270408	4
216	362575.92	7265548.8	-24.718424	31.641312	A3A270408	5
217	362569.76	7265542.08	-24.718484	31.64125	A3A270408	6
218	362564.16	7265550.48	-24.718407	31.641196	A3A270408	7
219	362558	7265558.88	-24.718331	31.641136	A3A270408	8
220	362552.4	7265566.72	-24.71826	31.641081	A3A270408	9
221	362546.24	7265574.56	-24.718188	31.641021	A3A270408	10
222	362541.2	7265581.84	-24.718122	31.640972	A3A270408	11
223	362547.36	7265590.24	-24.718047	31.641034	A3A270408	12
224	362553.52	7265582.96	-24.718113	31.641094	A3A270408	13
225	362559.68	7265575.12	-24.718184	31.641154	A3A270408	14
226	362565.84	7265567.84	-24.718251	31.641214	A3A270408	15
227	362572	7265560	-24.718322	31.641274	A3A270408	16
228	362577.04	7265570.08	-24.718231	31.641325	A3A270408	17
229	362570.88	7265577.92	-24.71816	31.641265	A3A270408	18
230	362564.72	7265585.2	-24.718094	31.641205	A3A270408	19
231	362558.56	7265593.04	-24.718023	31.641145	A3A270408	20
232	362551.84	7265599.76	-24.717961	31.641079	A3A270408	21
233	362557.44	7265608.72	-24.717881	31.641135	A3A270408	22
234	362564.72	7265601.44	-24.717947	31.641206	A3A270408	23
235	362570.88	7265594.16	-24.718014	31.641266	A3A270408	24
236	362577.04	7265586.88	-24.71808	31.641327	A3A270408	25
237	362583.76	7265579.6	-24.718146	31.641392	A3A270408	26
238	362589.92	7265588	-24.718071	31.641454	A3A270408	27
239	362583.2	7265595.28	-24.718005	31.641388	A3A270408	28
240	362576.48	7265602.56	-24.717938	31.641323	A3A270408	29
241	362569.76	7265610.4	-24.717867	31.641257	A3A270408	30
242	362564.16	7265617.12	-24.717806	31.641202	A3A270408	31
243	362570.88	7265623.28	-24.717751	31.641269	A3A270408	32
244	362578.16	7265617.12	-24.717807	31.641341	A3A270408	33
245	362584.32	7265609.84	-24.717873	31.641401	A3A270408	34

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
246	362591.04	7265602	-24.717945	31.641466	A3A270408	35
247	362597.76	7265594.72	-24.718011	31.641532	A3A270408	36
248	362702.48	7264328	-24.729457	31.642443	A5A250408	1
249	362692.96	7264325.2	-24.729482	31.642349	A5A250408	2
250	362683.44	7264322.96	-24.729501	31.642254	A5A250408	3
251	362673.36	7264320.16	-24.729525	31.642155	A5A250408	4
252	362664.4	7264316.8	-24.729555	31.642066	A5A250408	5
253	362654.88	7264315.12	-24.729569	31.641971	A5A250408	6
254	362653.76	7264325.2	-24.729478	31.641961	A5A250408	7
255	362653.2	7264334.72	-24.729392	31.641957	A5A250408	8
256	362652.64	7264344.24	-24.729306	31.641952	A5A250408	9
257	362652.08	7264354.88	-24.72921	31.641948	A5A250408	10
258	362652.08	7264363.84	-24.729129	31.641948	A5A250408	11
259	362660.48	7264370	-24.729074	31.642032	A5A250408	12
260	362664.96	7264359.36	-24.729171	31.642075	A5A250408	13
261	362667.76	7264350.4	-24.729252	31.642102	A5A250408	14
262	362671.68	7264341.44	-24.729333	31.64214	A5A250408	15
263	362675.6	7264332.48	-24.729414	31.642178	A5A250408	16
264	362684.56	7264336.96	-24.729375	31.642267	A5A250408	17
265	362680.64	7264346.48	-24.729288	31.642229	A5A250408	18
266	362676.72	7264355.44	-24.729207	31.642191	A5A250408	19
267	362672.8	7264364.4	-24.729126	31.642153	A5A250408	20
268	362668.32	7264373.92	-24.72904	31.64211	A5A250408	21
269	362675.6	7264378.4	-24.729	31.642182	A5A250408	22
270	362681.2	7264369.44	-24.729081	31.642237	A5A250408	23
271	362685.12	7264360.48	-24.729162	31.642275	A5A250408	24
272	362689.04	7264351.52	-24.729244	31.642313	A5A250408	25
273	362692.96	7264342.56	-24.729325	31.64235	A5A250408	26
274	362700.8	7264344.8	-24.729305	31.642428	A5A250408	27
275	362696.88	7264353.76	-24.729224	31.64239	A5A250408	28
276	362692.96	7264363.28	-24.729138	31.642353	A5A250408	29
277	362690.16	7264372.8	-24.729052	31.642326	A5A250408	30
278	362685.68	7264381.2	-24.728975	31.642282	A5A250408	31
279	362694.64	7264384	-24.728951	31.642371	A5A250408	32
280	362698	7264373.92	-24.729042	31.642403	A5A250408	33
281	362701.92	7264364.96	-24.729124	31.642441	A5A250408	34
282	362705.84	7264356	-24.729205	31.642479	A5A250408	35
283	362709.2	7264346.48	-24.729291	31.642511	A5A250408	36
284	362841.36	7263925.36	-24.733105	31.643777	A6A290508	1
285	362850.32	7263928.16	-24.733081	31.643866	A6A290508	2
286	362859.84	7263930.4	-24.733061	31.64396	A6A290508	3
287	362869.36	7263933.76	-24.733032	31.644054	A6A290508	4
288	362879.44	7263936.56	-24.733007	31.644154	A6A290508	5
289	362888.4	7263939.92	-24.732978	31.644243	A6A290508	6
290	362892.88	7263930.96	-24.733059	31.644287	A6A290508	7
291	362896.8	7263922	-24.73314	31.644324	A6A290508	8
292	362899.6	7263911.92	-24.733232	31.644351	A6A290508	9
293	362902.96	7263902.96	-24.733313	31.644383	A6A290508	10
294	362906.88	7263894	-24.733394	31.644421	A6A290508	11
295	362897.36	7263891.2	-24.733419	31.644327	A6A290508	12
296	362894	7263900.72	-24.733332	31.644295	A6A290508	13
297	362890.64	7263909.68	-24.733251	31.644262	A6A290508	14

ID	Map X	Map Y	Latitude - S	Longitude - E	Landuse ID	Point
298	362887.84	7263919.76	-24.73316	31.644236	A6A290508	15
299	362884.48	7263928.72	-24.733079	31.644203	A6A290508	16
300	362874.96	7263925.92	-24.733103	31.644109	A6A290508	17
301	362877.76	7263916.4	-24.733189	31.644136	A6A290508	18
302	362880.56	7263907.44	-24.73327	31.644162	A6A290508	19
303	362883.36	7263897.36	-24.733362	31.644189	A6A290508	20
304	362887.28	7263888.4	-24.733443	31.644227	A6A290508	21
305	362877.2	7263885.04	-24.733472	31.644127	A6A290508	22
306	362873.84	7263894.56	-24.733386	31.644095	A6A290508	23
307	362869.92	7263904.08	-24.7333	31.644057	A6A290508	24
308	362867.12	7263913.04	-24.733219	31.64403	A6A290508	25
309	362863.76	7263922.56	-24.733132	31.643998	A6A290508	26
310	362853.68	7263920.32	-24.733152	31.643898	A6A290508	27
311	362857.04	7263910.8	-24.733238	31.64393	A6A290508	28
312	362859.84	7263901.28	-24.733324	31.643957	A6A290508	29
313	362863.2	7263891.2	-24.733415	31.643989	A6A290508	30
314	362866	7263882.8	-24.733492	31.644016	A6A290508	31
315	362855.92	7263879.44	-24.733521	31.643916	A6A290508	32
316	362853.12	7263888.4	-24.73344	31.643889	A6A290508	33
317	362849.76	7263897.36	-24.733359	31.643857	A6A290508	34
318	362847.52	7263906.88	-24.733273	31.643836	A6A290508	35
319	362844.72	7263916.96	-24.733181	31.643809	A6A290508	36

Appendix D

Species Classification Algorithm

D.1 Main Algorithm

The following Matlab (version 7.8) algorithm was generated to form a matrix containing feature values for each tree ROI.

Contents

- Tree Identification
- Initialization
- PCA of HSI
- wLiDAR Data - Normalize each waveform by the total energy (area)
- wLiDAR Features
- Extract info from ROIs
- Tree Matrix
- Eliminate unwanted species
- Z-Test, dependent samples
- Kappa Accuracy

Tree Identification

```
%by Diane Sarrazin  
% 24 June 2010
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
%Creates a matrix with feature variables related to each tree ROI  
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Initialization

```
clc; clear all; close all;  
  
SPECIES = [7 5 4 0 6 11 1 1 14 16 1 4 5 7 0 5 ...
```



```

    2 7 0 0 12 15 1 0 2 0 1 1 0 3 1 1 2 1 0 1 0 3 1 0 ...
    0 4 0 0 4 0 5 4 9 0 0 0 0 0 17 0 0 4 0 1 2 1 1 2]; %collated species values
    %for subsets 1-16
tot_specie = sum(SPECIES);
s=623;
tree_matrix = zeros(tot_specie, s);
temp3 = 1; %tree count
mean_ground = zeros(16,1);

[pca_hsic1, pca_hsic2, pca_hsic3, pca_hsic5, pca_hsic6...
pca_hsid1, pca_hsid2, pca_hsid3, pca_hsid4, pca_hsid5, pca_hsid6, pca_hsid7,...
pca_hsid9, pca_hsic8a, pca_hsic9a, increment_hsi] = pca_hsi_landuse();

load pca_all_w_reduc; %includes pca for each subset and incremnt number

% figure,
% subplot(1,3,1), imshow(pca_wd1(:,:,1),[])
% subplot(1,3,2), imshow(pca_wd1(:,:,2),[])
% subplot(1,3,3), imshow(pca_wd1(:,:,3),[])

%For each image, extract features
for i = 1:16

    if i == 1
        load roib_mat.txt %col 4-27 are hsi values
        roi =roib_mat;
        subset = [158 49 88 17 48 36 65 13 36 54 105 88 66 47 42 76];
        %# of pixels in each ROI
        class = [2 1 1 1 1 2 3 1 2 1 1 2 3 3 3 2]; % class of each ROI
        species = [7 5 4 0]; %# of tree per species (cher, sbir, tser, pafr)

        hsi = enviread('subsetb'); hsi_pca = pca_hsic;
        load peakb; w = w_reduc; w_pca = pca_wb; load peakb;
    end

    if i == 2
        clear pca_hsic, clear pca_wb
        load roic1_mat.txt %col 4-27 are hsi values
        roi =roic1_mat;
        subset = [68 65 104 54 68 16 30 135 46 9 43 28 36 25 121 73 61 47 64];
        class = [1 2 2 1 2 1 4 1 2 2 2 2 1 2 1 2 3 2 2];
        species = [6 11 1 1]; %# of tree per species (cher, sbir, tser, pafr)

        hsi = enviread('subsetc1'); hsi_pca = pca_hsic1;
        load peakc1; w = w_reduc; w_pca = pca_wc1; load peakc1;
    end
end

```

```

end

if i == 3
    clear pca_hsic1, clear pca_wc1
    load roic2_mat.txt %col 4-27 are hsi values
    roi =roic2_mat;
    subset = [125 11 34 34 10 52 112 24 18 19 51 22 45 18 39 40 18 39 37 ...
              49 51 40 35 32 36 32 21 43 58 22 32 134 139 28 4];
    class = [1 2 1 2 3 2 2 4 1 1 4 1 2 4 1 2 1 2 2 1 2 2 2 1 2 1 1 1 2 2 2 1 2 1 4];
    species = [14 16 1 4]; %# of tree per species (cher, sbir, tser, pafr)

    hsi = enviread('subetc2'); hsi_pca = pca_hsic2;
    load peakc2; w = w_reduc; w_pca = pca_wc2; load peakc2;
end

if i == 4
    clear pca_hsic2, clear pca_wc2
    load roic3_mat.txt %col 4-27 are hsi values
    roi =roic3_mat;
    subset = [66 138 27 42 262 77 89 64 32 94 56 104 33 47 36 20 51];
    class = [1 2 2 2 2 1 2 2 4 4 4 2 4 1 1 4 1];
    species = [5 7 0 5]; %# of tree per species (cher, sbir, tser, pafr)

    hsi = enviread('subetc3'); hsi_pca = pca_hsic3;
    load peakc3; w = w_reduc;w_pca = pca_wc3; load peakc3;
end

if i == 5
    clear pca_hsic3, clear pca_wc3
    load roic5_mat.txt %col 4-27 are hsi values (no need for subetc4, no vectors)
    roi =roic5_mat;
    subset = [293 114 51 148 64 129 27 119 103];
    class = [2 2 1 2 2 2 1 2 2];
    species = [2 7 0 0]; %# of tree per species (cher, sbir, tser, pafr)

    hsi = enviread('subetc5'); hsi_pca = pca_hsic5;
    load peakc5; w = w_reduc;w_pca = pca_wc5; load peakc5;
end

if i == 6
    clear pca_hsic5, clear pca_wc5
    load roic6_mat.txt %col 4-27 are hsi values

```

```

roi =roic6_mat;
subset = [208 29 45 31 32 26 26 110 70 244 49 27 ...
          79 50 16 22 46 96 344 109 119 52 131 64 20 274 96 236];
class = [2 1 2 3 2 2 1 1 2 2 1 1 1 1 1 2 1 2 1 2 2 2 2 2 1 2];
species = [12 15 1 0]; %# of tree per species (cher, sbir, tser, pafr)

hsi = enviread('subsetc6'); hsi_pca = pca_hsic6;
load peakc6; w = w_reduc; w_pca = pca_wc6; load peakc6;
end

if i == 7
    clear pca_hsic6, clear pca_wc6
    load roid1_mat.txt %col 4-27 are hsi values
    roi =roid1_mat;
    subset = [26 6 4 12];
    class = [1 1 3 4];
    species = [2 0 1 1]; %# of tree per species (cher, sbir, tser, pafr)

    hsi=freadenvi2('L8_H2_A280508_spectral_subset',...
                  'E:\Wu\Data\Landuse8\Spectral data\');
    hsi_pca = pca_hsid1; load peakd1; w = w_reduc;w_pca = pca_wd1; load peakd1;
end

if i == 8
    clear pca_hsid1, clear pca_wd1
    load roid2_mat.txt %col 4-27 are hsi values
    roi =roid2_mat;
    subset = [155 11 61 17 54];
    class = [2 3 2 4 2];
    species = [0 3 1 1]; %# of tree per species (cher, sbir, tser, pafr)

    hsi =freadenvi2('L8_B260408_19_middle_spectral_subset',...
                  'E:\Wu\Data\Landuse8\Spectral data\');
    hsi_pca = pca_hsid2; load peakd2; w = w_reduc;w_pca = pca_wd2; load peakd2;
end

if i == 9
    clear pca_hsid2, clear pca_wd2
    load roid3_mat.txt %col 4-27 are hsi values
    roi =roid3_mat;
    subset = [5 6 26 27];
    class = [1 1 4 2];
    species = [2 1 0 1]; %# of tree per species (cher, sbir, tser, pafr)

```

```

    hsi=freadenvi2('L7_G6_A020508_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    hsi_pca = pca_hsid3; load peakd3; w = w_reduc;w_pca = pca_wd3; load peakd3;
end

if i == 10
    clear pca_hsid3, clear pca_wd3
    load roid4_mat.txt %col 4-27 are hsi values
    roi =roid4_mat;
    subset = [152 10 51 60];
    class = [2 3 2 2];
    species = [0 3 1 0]; %# of tree per species (cher, sbir, tser, pafr)

    hsi=freadenvi2('L7_B250408_22_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    hsi_pca = pca_hsid4; load peakd4; w = w_reduc;w_pca = pca_wd4; load peakd4;
end

if i == 11
    clear pca_hsid4, clear pca_wd4
    load roid5_mat.txt %col 4-27 are hsi values
    roi =roid5_mat;
    subset = [20 27 67 224];
    class = [2 2 2 2];
    species = [0 4 0 0]; %# of tree per species (cher, sbir, tser, pafr)

    hsi=freadenvi2('L7_B250408_23_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    hsi_pca = pca_hsid5; load peakd5; w = w_reduc; w_pca = pca_wd5; load peakd5;
end

if i == 12
    clear pca_hsid5, clear pca_wd5
    load roid6_mat.txt %col 4-27 are hsi values
    roi =roid6_mat;
    subset = [9 12 12 6 11 4 18 3 3 7 6 14 17];
    class = [1 3 4 3 4 3 4 4 3 3 1 1 1];
    species = [4 0 5 4]; %# of tree per species (cher, sbir, tser, pafr)

    hsi=freadenvi2('L7_G2_A020508_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    hsi_pca = pca_hsid6; load peakd6; w = w_reduc;w_pca = pca_wd6; load peakd6;
end

if i == 13
    clear pca_hsid6, clear pca_wd6

```

```

load roid7_mat.txt %col 4-27 are hsi values
roi =roid7_mat;
subset = [6 6 9 8 8 12 9 8 21];
class = [1 1 1 1 1 1 1 1 1];
species = [9 0 0 0]; %# of tree per species (cher, sbir, tser, pafr)

hsi=freadenvi2('L2_A1_A270408_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');
hsi_pca = pca_hsid7; load peakd7; w = w_reduc; w_pca = pca_wd7; load peakd7;
end

if i == 14
clear pca_hsid7, clear pca_wd7
load roid9_mat.txt %col 4-27 are hsi values
roi =roid9_mat;
subset = [7 9 14 7 12 6 10 13 41 27 16 2 6 44 16 8 5];
class = [3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3];
species = [0 0 17 0]; %# of tree per species (cher, sbir, tser, pafr)

hsi=freadenvi2('L2_A5_A250408_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');
hsi_pca = pca_hsid9; load peakd9; w = w_reduc; w_pca = pca_wd9; load peakd9;
end

if i == 15
clear pca_hsid9, clear pca_wd9
load roic8a_mat.txt %col 4-27 are hsi values (no need for subsetc4, no vectors)
roi =roic8a_mat;
subset = [91 78 58 68 208];
class = [2 2 4 2 2];
species = [0 4 0 1]; %# of tree per species (cher, sbir, tser, pafr)

hsi = enviread('subsetc8a');
hsi_pca = pca_hsic8a; load peakc8a; w = w_reduc; w_pca = pca_wc8a; load peakc8a;
end

if i == 16
clear pca_hsic8a, clear pca_wc8a
load roic9a_mat.txt %col 4-27 are hsi values (no need for subsetc4, no vectors)
roi =roic9a_mat;
subset = [33 51 41 34 29 67];
class = [1 4 3 4 1 2];
species = [2 1 1 2]; %# of tree per species (cher, sbir, tser, pafr)

hsi = enviread('subsetc9a');
```

```

        hsi_pca = pca_hsic9a; load peakc9a; w = w_reduc; w_pca = pca_wc9a; load peakc9a;
    end

```

PCA of HSI

```

% Extract features out of HSI
[fx_hsi] = gradient(hsi);
[fx2_hsi] = gradient(gradient(hsi));

```

wLiDAR Data - Normalize each waveform by the total energy (area)

```

w_r = reshape (w, [size(w,1)*size(w,2), size(w,3)]);
w_r_norm = zeros(size(w_r));
A1 = cumsum(w,3);
A = sum(w,3); A_r = reshape(A, [size(A,1)*size(A,2),1]);

for j = 1:size(w_r,1)
    w_r_norm(j,:) = w_r(j,:) ./ A_r(j);
end

% x=1:size(w,3);
% figure, subplot(1,2,1), plot(x, w_r(10,:)), title('Cropped waveform')
% subplot(1,2,2), plot(x,w_r_norm(10,:)); title('Cropped, normalized, waveform')
w_norm = reshape(w_r_norm, [size(w,1), size(w,2), size(w,3)]);

clear last_peak_loc %now it is 100 for all wf

```

wLiDAR Features

```

[height, crown, HA] = height_function(w,A,A1);
%to know where ground is in majority of subsets
%mean_ground(i) = mean(mean(last_peak_loc));

[fx_w] = gradient(w);
[fx2_w] = gradient(gradient(w));

% for i=1:10
%     figure, subplot(3,1,1), plot(squeeze(w(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
%     subplot(3,1,2), plot(squeeze(fx_w(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
%     subplot(3,1,3), plot(squeeze(fx2_w(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
% end

```

Extract info from ROIs

```

temp = 1; temp1 = 0;

for iter = 1: sum(species)

    %Make a variable out of each tree with its HSI values
    temp1 = temp1 + subset(iter); %x pixels in each tree
    tree = zeros(subset(iter),s-1);
    tree(:,1:27) = roi(temp:temp1,:); %hsi

    %Add the wLiDAR info at the end of each variable
    for ii = 1:subset(iter); %x pixels in each tree

        temp2 = (temp - 1) + ii; %var to keep going down in roi file
        tree(ii,28:51) = squeeze(hsi_pca(roi(temp2,3),roi(temp2,2),:))';
        %one pixel per line, in all bands
        tree(ii,52:75) = squeeze( fx_hsi( roi(temp2,3), roi(temp2,2),:))';
        tree(ii,76:99) = squeeze( fx2_hsi( roi(temp2,3), roi(temp2,2),:))';

        tree(ii,100:229) = squeeze( w(roi(temp2,3),roi(temp2,2),:))';
        tree(ii,230:359) = squeeze( w_pca(roi(temp2,3),roi(temp2,2),:))';
        %first 148 are important
        tree(ii,360:489) = squeeze( fx_w( roi(temp2,3), roi(temp2,2),:))';
        tree(ii,490:619) = squeeze( fx2_w( roi(temp2,3), roi(temp2,2),:))';
        tree(ii,620) = squeeze( height( roi(temp2,3), roi(temp2,2),:))';
        tree(ii,621) = squeeze( crown( roi(temp2,3), roi(temp2,2),:))';

        %tree(ii,622) = CT/H defined in tree_matrix;
        %tree(ii,623) = max h defined in tree_matrix;

    end
end

```

Tree Matrix

Put subset and ID in each tree

```

tree_matrix(temp3,1) = i; % subset #
tree_matrix(temp3,2) = temp3; % tree ID

%Assign class to each tree
tree_matrix(temp3,3) = class(iter); % class

% Mean per band of pixels of HSI, zscores_hsi, WLiDAR,
%zscores_wf_out for each tree
tree_matrix(temp3, [4:621]) = mean(tree(:, [4:621])); %mean by column/band

```

```

tree_matrix(temp3, 622) = tree_matrix(temp3, 621) / tree_matrix(temp3, 620);
%CT/H
tree_matrix(temp3, 623) = max(tree(:,620));
%max height per column/band per tree

%update variables before other tree in same subset
temp = temp + subset(iter); %moving to other set of pixels in a tree
temp3 = temp3 + 1; %moving to other tree
clear tree;

end

%update variables before other subset
clear roi; clear species; clear subset; clear temp; clear temp1; clear temp2;
clear class; clear w; clear w_r; clear w_pca; clear hsi_r; clear hsi
clear hsi_pca; clear height; clear crown; clear A; clear A1; clear HA;
clear fx_w; clear fx2_w; clear fx_hsi; clear fx2_hsi
clear fy_w; clear fy2_w; clear fy_hsi; clear fy2_hsi
clear fx_w_pca; clear fx2_w_pca; clear fx_hsi_pca; clear fx2_hsi_pca
clear fy_w_pca; clear fy2_w_pca; clear fy_hsi_pca; clear fy2_hsi_pca
%clear zscores_r1; clear zscores_r2; clear coeff;

end

tree_matrix(isnan(tree_matrix) == 1) = 0;
tree_matrix = tree_matrix';

%clear all variables not useful for future computation
clear i; clear ii; clear iter; clear temp3; clear roib_mat, clear roic1_mat
clear roic2_mat; clear roic3_mat, clear roic5_mat, clear roic6_mat
clear roid1_mat, clear roid2_mat, clear roid3_mat; clear roid4_mat
clear roid5_mat, clear roid6_mat, clear roid7_mat, clear roid9_mat
clear roic8a_mat; clear roic9a_mat, clear s, clear zscores_wf_out
clear zscores_hsi, clear pca_hsic9a, clear pca_wc9a

% For better reading in excel (limited by col, but not rows)

tree_matrix = tree_matrix';

```

Eliminate unwanted species

In the case where some species give bad results because of small sample size

$S4 = \text{sum}(\text{tree_matrix}(:,3) == 4)$; $S5 = \text{sum}(\text{tree_matrix}(:,3) == 5)$; $S = \text{tot_specie} - (S4 + S5)$;

$S123 = \text{zeros}(S, \text{size}(\text{tree_matrix}, 2))$; $a = 1$; for $i = 1:\text{tot_specie}$


```

    if tree_matrix(i,3) == 1
        S123(a,:) = tree_matrix(i,:);
        a = a+1;
    elseif tree_matrix(i,3) == 2
        S123(a,:) = tree_matrix(i,:);
        a = a+1;
    elseif tree_matrix(i,3) == 3
        S123(a,:) = tree_matrix(i,:);
        a = a+1;
    end
end

S123 = S123'; S123(isnan(S123)) = 0;

```

Z-Test, dependent samples

```

%from not transposed matrix

%take out classes 4&5
%tree_matrix_new = zeros(195, size(tree_matrix,2));
% a=1;
%
% for i = 1:size(tree_matrix,1)
%     if tree_matrix(i,3) == 1 | tree_matrix(i,3) == 2 | tree_matrix(i,3) == 3
%
%         tree_matrix_new(a,:) = tree_matrix(i,:);
%         a = a+1;
%     end
% end

%load tree_matrix2
tree_matrix_new = tree_matrix';

%Build the basics of z_matrix
z_matrix = zeros(size(tree_matrix,2), 8);
z_matrix(:, [1:2]) = tree_matrix_new(:, [2:3]); %tree ID and known class
z_matrix(:,3) = tree_matrix_new(:,3); % wLiDAR class - orig
z_matrix(:,4) = tree_matrix_new(:,3); % wLiDAR class - PCA
z_matrix(:,5) = tree_matrix_new(:,3); %HSI class - orig
z_matrix(:,6) = tree_matrix_new(:,3); %HSI class - orig
z_matrix(:,7) = tree_matrix_new(:,3); %Fusion class - orig
z_matrix(:,8) = tree_matrix_new(:,3); %Fusion class - PCA

```

```

% Everything that's not well classified is told here
%wLiDAR - orig
z_matrix([1,7,14,15,16,23,33,39,40,49,57,70,121,130,136,153,183,191,192],3) = 1;
z_matrix([3,13,43,50,52,67,80,104,114,123,139,147,148,149,150,151,152,158],3) = 2;
z_matrix([6,9,10,37,38,41,46,47,74,125,128,129,132,146,157,187,193],3) =3;
z_matrix([4,20,22,26,60,61,62,63,64,69,71,84,85,87,94,98,101,103,107,108,...
110,112,135,156,160,179,190,194],3) = 4;

%wLiDAR - PCA
z_matrix([7,13,26,33,37,43,49,121,129,130,136,153,155,183,187,191,192],4) = 1;
z_matrix([14,15,20,22,45,46,52,61,67,71,80,86,100,104,108,110,121,...
147,158,193],4) = 2;
z_matrix([3,6,23,27,30,39,64,70,83,85,86,128,134,135,140,143,144,145,...
148,150,152,160,164,166],4) =3;
z_matrix([4,8,10,16,38,40,44,57,62,63,69,74,84,102,165,167],4) = 4;

%HSI - orig
z_matrix([9,64,74,121,128,139,149,152],5) = 1;
z_matrix([33,40,80,155],5) = 2;
z_matrix([104,126,137,142,156,159,163],5) =3;
z_matrix([3,4,5,7,8,11,14,15,16,31,36,44,50,52,57,69,77,97,114,122,125,127,...
134,135,140,143,144,145,146,147,157,190,194],5) =4;

%HSI - PCA
z_matrix([9,57,64,70,79,99,136,143,152,153],6) = 1;
z_matrix([13,33,45,80,104,126,127],6) = 2;
z_matrix([137,138,140,141,142,144,145,148,150,156,157,165,166],6) =3;
z_matrix([2,3,4,7,11,14,15,16,31,40,44,50,52,69,76,77,84,94,121,122,125,...
134,146,149,161,190,194],6) =4;

%FUSION - orig
z_matrix([7,9,37,40,43,49,74,121,136,139,152],7) = 1;
z_matrix([13,33,50,52,80,104,147,155,158],7) = 2;
z_matrix([126,128,132,142,153,156,163],7) =3;
z_matrix([3,4,8,14,15,16,22,31,44,57,64,69,84,108,122,125,127,135,...
144,145,146,149,151,190,194],7) =4;

%FUSION - PCA
z_matrix([7,9,43,57,70,99,136,140,145,152,153],8) = 1;
z_matrix([13,15,33,45,80,104],8) = 2;
z_matrix([143,144,148,150,156,165,166],8) =3;
z_matrix([3,4,8,11,14,16,40,44,50,52,64,69,77,84,121,122,125,126,127,...
134,135,146,149,151,161,190],8) =4;

%z-test wLiDAR orig vs FUSION pca

```

```

cc = numel (find (z_matrix(:,3) == z_matrix(:,2) & z_matrix(:,8) == z_matrix(:,2)) );
ci = numel (find (z_matrix(:,3) == z_matrix(:,2) & z_matrix(:,8) ~= z_matrix(:,2)) );
ic = numel (find (z_matrix(:,3) ~= z_matrix(:,2) & z_matrix(:,8) == z_matrix(:,2)) );
ii = numel (find (z_matrix(:,3) ~= z_matrix(:,2) & z_matrix(:,8) ~= z_matrix(:,2)) );

z1 = (ci-ic)\sqrt((ci+ic))
% chi1 = (ci-ic)^2\sqrt((ci+ic))

%z-test HSI orig and FUSION pca

cc = numel (find (z_matrix(:,5) == z_matrix(:,2) & z_matrix(:,8) == z_matrix(:,2)) );
ci = numel (find (z_matrix(:,5) == z_matrix(:,2) & z_matrix(:,8) ~= z_matrix(:,2)) );
ic = numel (find (z_matrix(:,5) ~= z_matrix(:,2) & z_matrix(:,8) == z_matrix(:,2)) );
ii = numel (find (z_matrix(:,5) ~= z_matrix(:,2) & z_matrix(:,8) ~= z_matrix(:,2)) );

z2 = (ci-ic)\sqrt((ci+ic))
% chi1 = (ci-ic)^2\sqrt((ci+ic))

%z-test wLiDAR and HSI -orig

cc = numel (find (z_matrix(:,3) == z_matrix(:,2) & z_matrix(:,5) == z_matrix(:,2)) );
ci = numel (find (z_matrix(:,3) == z_matrix(:,2) & z_matrix(:,5) ~= z_matrix(:,2)) );
ic = numel (find (z_matrix(:,3) ~= z_matrix(:,2) & z_matrix(:,5) == z_matrix(:,2)) );
ii = numel (find (z_matrix(:,3) ~= z_matrix(:,2) & z_matrix(:,5) ~= z_matrix(:,2)) );

z3 = (ci-ic)\sqrt((ci+ic))
% chi1 = (ci-ic)^2\sqrt((ci+ic))

```

Kappa Accuracy

```

%CLASS 1
xii = 149; n = 195;
xx = 65*53+77*73+33*30+20*39;
k1 = (n*xii-xx)/(n^2-xx);

%CLASS 2
xii = 125; n = 195;
xx = 65*56+77*77+33*21+20*41;
k2 = (n*xii-xx)/(n^2-xx);

%CLASS 3
xii = 151; n = 195;
xx = 65*57+77*68+33*26+20*44;
k3 = (n*xii-xx)/(n^2-xx);

```

```
%CLASS 6
xii = 44+65+24+12; n = 195;
xx = 65*55+77*71+33*31+20*38;
k6 = (n*xii-xx)/(n^2-xx);
```

The following Matlab (version 7.8) algorithm was generated to find the most accurate height and crown thickness measures from wLiDAR data. It is a sub-algorithm from the main species classification algorithm.

- Height, crown thickness and herbaceous area function
- Find the peaks in each waveform and their bin location
- Make sense of peak and loc and remove the small peaks
- Find first detected peak in each waveform
- Height: FWHM of first peak to last peak
- Find Crown Thickness - FWHM of first peak to first valley
- Find values of crown thickness at each pixel
- Find HA

Height, crown thickness and herbaceous area function

[illegible]

Find the peaks in each waveform and their bin location

```

%clc;close all;clear all
last_peak_loc = 100.*ones(size(w,1),size(w,2));
%for tree class, ground peak is always at bin 100
w_r = reshape (w, [size(w,1)*size(w,2), size(w,3)]); %ground last peak

% Peak normalize intensities of the waveforms
% w_max = max(max(w_r));
% w_r_norm = w_r ./ w_max;

for i=1:size(w_r,1)

    [locmax binloc]=lmax(w_r(i,:),3 );
    peak{i} = locmax; %peak value
    loc{i} = binloc; %bin number

end

% Convert peak and bin values from cell to matrix
numpeaks = cellfun(@numel,peak);

peak = cell2mat(peak)';
loc = cell2mat(loc)';

clear locmax; clear binloc; clear i;

last_peak_loc = 100.*ones(size(w,1),size(w,2));
%for tree class, ground peak is always at bin 100

```

Make sense of peak and loc and remove the small peaks

```

a = 0;
peak_val = zeros(max(size(numpeaks)), max(numpeaks));
loc_peak = zeros(max(size(numpeaks)), max(numpeaks));
numpeaks_new = zeros(size(w_r,1),1);

for i = 1:max(size(w_r))

    if i == 1
        a = 0; %no numpeaks(i-1) if i=1
    else
        a=a+numpeaks(i-1);
    end

    % Put each peak associated with a waveform in its associated line of

```

```

% peak_val
if isempty(a) == 1
    a = 0;
else
    inc2 = peak(a+1:a+numpeaks(i),1);
    inc3 = loc(a+1:a+numpeaks(i),1);

    % Remove the small peaks, under normalized values 0.05
    c = find(inc2 >= 0.02); %indices of where the peak is in inc2
    numpeaks_new(i,1) = numel(c); % numel that are peaks

    for ii = 1:numel(c)
        peak_val( i,ii ) = inc2(c(ii));
        loc_peak( i,ii ) = inc3(c(ii));
    end

end

clear inc2; clear inc3; clear c;

end
clear a, clear i,

numpeaks_r = numpeaks_new;
numpeaks = reshape(numpeaks_r, [size(w,1),size(w,2)]);

```

Find first detected peak in each waveform

```

% first_peak_r = zeros(size(w_r,1),1);
first_peak_loc_r = zeros(size(w_r,1),1);

% last_peak_r = zeros(size(w_r,1),1);
% last_peak_loc_r = zeros(size(w_r,1),1);

for i = 1:max(size(w_r))

    % Exclude null waveforms
    if numpeaks_r(i) ~= 0 %if not matrix,

%         first_peak_r(i,1) = peak_val(i,1);
        first_peak_loc_r(i,1) = loc_peak(i,1);

%         last_peak_r(i,1) = peak_val(i,numpeaks_r(i));
%         last_peak_loc_r(i,1) = loc_peak(i,numpeaks_r(i));

    end

end

```

```

clear incl

numpeaks = reshape(numpeaks_r, [size(w,1),size(w,2)]);

% first_peak = reshape(first_peak_r,[size(w,1),size(w,2)]);
first_peak_loc = reshape(first_peak_loc_r,[size(w,1),size(w,2)]);

% last_peak = reshape(last_peak_r,[size(w,1),size(w,2)]);
% last_peak_loc = reshape(last_peak_loc_r,[size(w,1),size(w,2)]);
last_peak_loc_r = reshape(last_peak_loc,[size(w,1)*size(w,2),1]);

Height: FWHM of first peak to last peak

height_r = zeros(size(w_r,1),1);

for i = 1:size(w_r,1)

    if numpeaks_r(i) >= 2

        HM = peak_val(i,1)/2; %half max intensity of first peak
        %array_position is bin number corresponding to FWHM
        %finding closest value of a band with HM
        if peak_val(i,1) > 0.04*max(peak_val(i,:)); %if peak is super small
            [min_difference, array_position] = ...
                min(abs( w_r(i,1:loc_peak(i,1)) - HM ));

            %If all bands before first peak and HM are zero, choose band
            %before last peak,
            if w_r(i,1:loc_peak(i,1)-1) == 0
                [array_position] = [loc_peak(i,1)-1];
            end

            %If all bands before -1 band before first peak- are zero and
            %the HM is closer to the band with zero value, choose two bands
            %before last peak
            if (w_r(i,1:loc_peak(i,1)-2) == 0) & ...
                ( w_r(i,loc_peak(i,1)-1) - HM) > (w_r(i,loc_peak(i,1)-2) - HM))
                [array_position] = [loc_peak(i,1)-2];
            end

            height_r(i) = (last_peak_loc_r(i) - array_position)*0.15;
        end
    end
end
end

```

```
clear HM; clear min_difference; clear array_position
```

```
height = reshape(height_r, [size(w,1),size(w,2)]);
%figure, imshow(height,[],colorbar, title('Height image'))
```

Find Crown Thickness - FWHM of first peak to first valley

Find location of First Valley just before last peak (as per bins, not absolute gaussian fit)

```
valley_r = zeros(size(w_r,1),1);
valley_loc_r = zeros(size(w_r,1),1);

for i = 1:size(w_r,1)

    if numpeaks_r(i) >= 2

        a = numpeaks_r(i);
        b = numpeaks_r(i)-1;
        e = min (w_r(i, loc_peak(i,b):loc_peak(i,a) ));
        c = find( w_r(i,:) == e ); %indice of where the valley is

        %There can be two loc with same value in w_r
        %indices of c in range: second to last peak:last peak
        d = find( (c >= loc_peak(i,b)) & (c <= loc_peak(i,a)) );
        valley_loc_r(i) = c(round(median(d))); %choose middle bin if c all the same
        valley_r(i) = w_r(i,valley_loc_r(i));

    end

end

clear a; clear b; clear c; clear d; clear e
```

Find values of crown thickness at each pixel

```
crown_r = zeros(size(w_r,1),1);

for i = 1:size(w_r,1)

    if numpeaks_r(i) >= 2

        CT = peak_val(i,1)/2; %half max intensity of first peak
        %array_position is bin number corresponding to FWHM
        %finding closest value of a band with CT
```



```

    if peak_val(i,1) > 0.04*max(peak_val(i,:));
        [min_difference, array_position] = min(abs( w_r(i,1:loc_peak(i,1)) - CT ));

        %If all bands before first peak and HM are zero, choose band
        %before last peak,
        if w_r(i,1:loc_peak(i,1)-1) == 0
            [array_position] = [loc_peak(i,1)-1];
        end

        %If all bands before -1 band before first peak- are zero and
        %the HM is closer to the band with zero value, choose two bands
        %before last peak
        if (w_r(i,1:loc_peak(i,1)-2) == 0) & ...
            ( w_r(i,loc_peak(i,1)-1) - CT) > (w_r(i,loc_peak(i,1)-2) - CT))
            [array_position] = [loc_peak(i,1)-2];
        end

        %INSTEAD OF LAST_PEAK_LOC_R, FIRST VALLEY before last peak
        crown_r(i) = (valley_loc_r(i) - array_position)*0.15;

    end
end
end

clear CT; clear min_difference; clear array_position

```

Find HA

```

%Sum or area under the curve from valley before last peak
%and 10% of the end of falling edge

A_r = reshape (A,[size(A,1)*size(A,2),1]);
A1_r = reshape (A1,[size(A1,1)*size(A1,2),size(A1,3)]);
HA_r = zeros(size(numpeaks_r));

for i = 1:size(w_r,1)

    %Find first loc between last_peak_loc:end that is lower than
    %intensity = 0.1

    % First evaluate if there is a noticeable peak (>0.05)

    if last_peak_loc_r(i) == 0
        HA_r(i) = A_r(i);
        %if no noticeable peak, HA = CUMSUM of areas under curve.
    else

```

```

    a1 = find( w_r(i,last_peak_loc_r(i):end) < 0.01 );

    %Conditions regarding a1
    if isempty(a1) == 1
        a1 = size(w_r,2); %if no value <= 0.01, chose last bin
    else
        a1 = last_peak_loc_r(i) + a1(1) - 1;
        %if many bins, chose first one
    end

    %Assign a0
    a0 = valley_loc_r(i);

    if a0 == 0
        a0 = 1; %if not other peak, assign a0 to first time bin
    end

    HA_r(i) = A1_r(i,a1) - A1_r(i,a0);

end

clear a1; clear a0;

end

HA = reshape(HA_r, [size(A)]);
% figure, imshow(HA,[])

crown = reshape(crown_r, [size(w,1),size(w,2)]);
% figure, subplot(1,2,2), imshow(crown,[]), title('Crown Thickness')
% subplot(1,2,1), imshow(height,[]), title('Tree Height')

```

Appendix E

Herbaceous Biomass Modeling Algorithm

E.1 Main Algorithm

Contents

The following Matlab (version 7.8) algorithm was generated to form a matrix containing feature values for each herbaceous biomass plot from the Bushbuckridge communal settlement and the KNP. The plot locations can be found in Appendix C.

- Herbaceous biomass modeling
- Figure of plot locations
- Mean of 3x3 HSI values for each GPS point
- Normalize wLiDAR Data by the total energy (area)
- Build composite waveform for 3x3 pixels of wLiDAR data for each GPS point
- Mean of 3x3 wLiDAR gradients (x,y) values for each GPS point
- Perform PCA on wLiDAR and extract values for each GPS point
- Get A, Sigma and Mu from each modeled gaussian
- Build matrix of different wLiDAR and HSI variables
- Group in terms of 5 g interval in biomass
- Linear regression

Herbaceous biomass modeling

by Diane Sarrazin 24 June 2010

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Creates a matrix with feature variables related to each biomass plot
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clc; close all; clear all;
```

```

a = 230; % # of variables
inc = 0;
model_input = zeros(36*9-5,a); %no biomass for points 20-23 (site1) and 16 (site2)
[pca_hsid1, pca_hsid2, pca_hsid3, pca_hsid5, pca_hsid6, pca_hsid7,...
    pca_hsid8, pca_hsid9, pca_hsid10, increment_hsi] = pca_hsi_landuse_bio();
load pca_all_w_bio_reduc;
% figure, imshow(pca_hsid1(:,:,2),[])

for iter = 1:9

    if iter==1
        hsi=freadenvi2('L8_H2_A280508_spectral_subset',...
            'E:\Wu\Data\Landuse8\Spectral data\');
        unmix = freadenvi2('unmix1','E:\Sarrazin\Data\ENVI\Unmixing\');
        %first band is road, second is tree, third is grass
        w_pca = pca_wd1; hsi_pca = pca_hsid1;
        load GPS_points_L8_H2.txt
        GPS_points = GPS_points_L8_H2;
        load peakd1;
    end

    if iter==2
        clear pca_wd1, clear pca_hsid1
        hsi = freadenvi2('L8_B260408_19_middle_spectral_subset',...
            'E:\Wu\Data\Landuse8\Spectral data\');
        unmix = freadenvi2('unmix2','E:\Sarrazin\Data\ENVI\Unmixing\');
        %first band is road, second is tree, third is grass
        w_pca = pca_wd2; hsi_pca = pca_hsid2;
        load GPS_points_L8_19_mid.txt
        GPS_points=GPS_points_L8_19_mid;
        load peakd2;
    end

    if iter==3
        clear pca_wd2, clear pca_hsid2
        hsi=freadenvi2('L7_G6_A020508_spectral_subset',...
            'E:\Wu\Data\Landuse7\Spectral data\');
        unmix = freadenvi2('unmix3','E:\Sarrazin\Data\ENVI\Unmixing\');
        %first band is road, second is tree, third is grass
        w_pca = pca_wd3; hsi_pca = pca_hsid3;
        load GPS_points_L7_G6.txt
        GPS_points=GPS_points_L7_G6;
        load peakd3;
    end
end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%4 - no biomass data

if iter==4
    clear pca_wd3, clear pca_hsid3
    hsi=freadenvi2('L7_B250408_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    unmix = freadenvi2('unmix5','E:\Sarrazin\Data\ENVI\Unmixing\');
    %first band is road, second is tree, third is grass
    w_pca = pca_wd5; hsi_pca = pca_hsid5;
    load GPS_points_L7_23.txt
    GPS_points=GPS_points_L7_23;
    load peakd5;
end

if iter==5
    clear pca_wd5, clear pca_hsid5
    hsi=freadenvi2('L7_G2_A020508_spectral_subset',...
    'E:\Wu\Data\Landuse7\Spectral data\');
    unmix = freadenvi2('unmix6','E:\Sarrazin\Data\ENVI\Unmixing\');
    %first band is road, second is tree, third is grass
    w_pca = pca_wd6; hsi_pca = pca_hsid6;
    load GPS_points_L7_G2.txt
    GPS_points=GPS_points_L7_G2;
    load peakd6;
end

if iter==6
    clear pca_wd6, clear pca_hsid6
    hsi=freadenvi2('L2_A1_A270408_spectral_subset',...
    'E:\Wu\Data\Landuse2\Spectral data\');
    unmix = freadenvi2('unmix7','E:\Sarrazin\Data\ENVI\Unmixing\');
    %first band is road, second is tree, third is grass
    w_pca = pca_wd7; hsi_pca = pca_hsid7;
    load GPS_points_L2_A1.txt
    GPS_points=GPS_points_L2_A1;
    load peakd7;
end

if iter==7
    clear pca_wd7, clear pca_hsid7
    hsi=freadenvi2('L2_A3_A270408_spectral_subset',...
    'E:\Wu\Data\Landuse2\Spectral data\');
    unmix = freadenvi2('unmix8','E:\Sarrazin\Data\ENVI\Unmixing\');
    %first band is road, second is tree, third is grass
    w_pca = pca_wd8; hsi_pca = pca_hsid8;
    load GPS_points_L2_A3.txt

```

```

        GPS_points=GPS_points_L2_A3;
        load peakd8;
    end

    if iter==8
        clear pca_wd8, clear pca_hsid8
        hsi=freadenvi2('L2_A5_A250408_spectral_subset',...
            'E:\Wu\Data\Landuse2\Spectral data\');
        unmix = freadenvi2('unmix9','E:\Sarrazin\Data\ENVI\Unmixing\');
        %first band is road, second is tree, third is grass
        w_pca = pca_wd9; hsi_pca = pca_hsid9;
        load GPS_points_L2_A5.txt
        GPS_points=GPS_points_L2_A5;
        load peakd9;
    end

    if iter==9
        clear pca_wd9, clear pca_hsid9
        hsi=freadenvi2('L2_A6_A290508_spectral_subset',...
            'E:\Wu\Data\Landuse2\Spectral data\');
        unmix = freadenvi2('unmix10','E:\Sarrazin\Data\ENVI\Unmixing\');
        %first band is road, second is tree, third is grass
        w_pca = pca_wd10; hsi_pca = pca_hsid10;
        load GPS_points_L2_A6.txt
        GPS_points=GPS_points_L2_A6;
        load peakd10;
    end

    w = w_reduc(:, :, 87:116);
    clear w_reduc

```

Figure of plot locations

```

%   [RGB] =hsi_rgb(hsi);
%   figure,subplot,imagesc(RGB),axis image,axis off
%   hold on
%   for i=1:size(GPS_points,1)
%       num=num2str(i);
%       scatter(GPS_points(i,2),GPS_points(i,3),[],[1 0 0],'filled')
%       text(GPS_points(i,2)+3,GPS_points(i,3),num,'fontsize',8,'color','yellow');
%       hold on
%   end
%
%   clear i

```

Mean of 3x3 HSI values for each GPS point

put in val3, [GPS_pointsx24]

```

val3 = GPS(hsi,GPS_points);

% Mean of 3x3 HSI gradients (x,y) values for each GPS point
% put in val1 and val2, [GPS_pointsx24]
[fx_hsi] = gradient(hsi);
[fx2_hsi] = gradient(gradient(hsi));

val1 = GPS(fx_hsi,GPS_points);
val1a = GPS(fx2_hsi,GPS_points);

% Mean of 3x3 BG/HSI values for each GPS point
% put in val3a-d, [GPS_pointsx4]
road = unmix(:,:,1); grass= unmix(:,:,3);
val3a = GPS(road, GPS_points);
val3b = GPS_add(road, GPS_points);
val3c = GPS(grass, GPS_points);
val3d = GPS_add(grass, GPS_points);

% put in val8, [GPS_pointsx24]
% first four bands of PCA include 99.5% of the variance
val8 = GPS(hsi_pca,GPS_points);

```

Normalize wLiDAR Data by the total energy (area)

```

w_r = reshape(w,[size(w,1)*size(w,2), size(w,3)]); %30 bands
w_r_norm = zeros(size(w_r));
%   A1 = cumsum(w,3);
A = sum(w,3); A_r = reshape(A, [size(A,1)*size(A,2),1]);

for j = 1:size(w_r,1)
    w_r_norm(j,:) = w_r(j,:) ./ A_r(j);
end

%   x=1:size(w,3);
%   figure, subplot(1,2,1), plot(x, w_r(10,:)), title('Cropped waveform')
%   subplot(1,2,2), plot(x,w_r_norm(10,:)); title('Cropped, normalized, waveform')

w_norm = reshape(w_r_norm, [size(w,1), size(w,2), size(w,3)]);

```

Build composite waveform for 3x3 pixels of wLiDAR data for each GPS point

align comp wf peak to bin 14, area normalize, find diff with road wf put in val4, [GPS_points x variable band number]

```
[val4 hbdiff] = GPScomp_bio(w_norm, GPS_points, numpeaks);

A1_diff = sum(hbdiff,2);

% x = 1:size(w_r,2);
% lmval = zeros(size(val4,1),2);
% for i = 1:size(val4,1)
%     figure, plot(x, val4(i,:))
%     lmax(val4(i,:));
% end
```

Mean of 3x3 wLiDAR gradients (x,y) values for each GPS point

put in val10 and val11, [GPS_pointsx30]

```
[fx_w] = gradient(w_norm);
[fx2_w] = gradient(gradient(w_norm));

val10a = GPS(fx_w,GPS_points);
val10c = GPS(fx2_w,GPS_points);

% for i=1:20:100
%     figure, subplot(3,1,1), plot(squeeze(w_norm(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
%     subplot(3,1,2), plot(squeeze(fx_w(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
%     subplot(3,1,3), plot(squeeze(fx2_w(i,i+4,:)))
%     xlabel('Time (ns)'), ylabel('Intensity')
% end
```

Perform PCA on wLiDAR and extract values for each GPS point

put in val9, [GPS_points x 56 first pca bands]

```
val9 = GPS(w_pca,GPS_points);
```


Get A, Sigma and Mu from each modeled gaussian

```

[M2] = em_function(GPS_points, val4);


```

Build matrix of different wLiDAR and HSI variables

for each GPS points

```

var = zeros(size(GPS_points,1),a);

var(:,1) = GPS_points(:,1); %+inc; %ID
var(:,2) = GPS_points(:,4); %Field HB
var(:,3:26) = val3; %mean of 3x3 hsi
var(:,27:50) = val1; %mean of 3x3 fx_hsi
var(:,51:74) = val1a; % mean of 3x3 fx2_hsi
var(:,75:98) = val8; %mean of 3x3 hsi_pca
var(:,99) = val3c; %mean of 3x3 grass/unmix
var(:,100) = val3d; %sum of 3x3 grass/unmix

var(:,101:130) = val4; %mean of 3x3 - composite waveforms
var(:,131:130+size(val9,2)) = val9; %mean of 3x3 w_pca
var(:,161:190) = val10a; %mean of 3x3 fx_w
var(:,191:220) = val10c; %mean of 3x3 fx2_w
var(:,221) = A1_diff; % comp wf - road wf

var(:,222) = val14a; %amplitude 1
var(:,223) = val14b; %sigma 1

```

```

var(:,224) = val14c; %sum 1
var(:,225) = val14f; % amplitude 2
var(:,226) = val14g; %sigma 2
var(:,227) = val14h; %sum 2
var(:,228) = val18a; %rt
var(:,229) = val18b; %ft

model_input(inc+1:inc+size(GPS_points,1) ,:) = var;
inc = inc+size(GPS_points,1);

clear var; clear GPS_points; clear hsi; clear HA; clear band;
clear w; clear w_r; clear w_norm; clear w_r_norm; clear last_peak_loc
clear height; clear crown; clear fx_hsi; clear fy_hsi
clear A; clear A1; clear numpeaks; clear fx_w; clear fy_w
clear fx2_hsi; clear fy2_hsi; clear fx2_w; clear fy2_w; clear i
clear j; clear increment_w; clear hsi_pca; clear w_pca
clear bg; clear ft; clear rt; clear M2; clear x; clear hsi_r; clear A_r;

clear val1; clear val2; clear val1a; clear val2a; clear val3;
clear val3a; clear val3b; clear val4; clear val5; clear val6;
clear val6a; clear val7; clear val8; clear val9; clear val10a;
clear val10b; clear val10c; clear val10d; clear val14e; clear val14a;
clear val14b; clear val14c; clear val14d; clear val18a; clear val18b;
clear val14e; clear val14f; clear val14g; clear val14h;
clear val14i; clear val14j; clear coeff; clear zscores_r;
clear pcvars, clear variance; clear threshold

end

clear pca_wd10, clear pca_hsid10

model_input(isnan(model_input)) = 0;
model_input(isinf(model_input)) = 0;

Group in terms of 5 g interval in biomass

for j = 1:22 %max HB/5
    index = find(model_input(:,2) >= 1+5*j & model_input(:,2)<=5+5*j);
    if(1-isempty(index))
        model_input([index],230) = round(mean(model_input(index,2),1));
    end
end

%Remove waveforms that were badling modeled
model_input = model_input([1:3,5,7:34,36:51,53,55:57, 59:79,81:108,...

```

```
110:123,125,127:165,167:168,170:174,176:183,185:239,241:250,252:end],:);
```

```
%Transpose for complete viewing in Excel
model_input_trans = model_input';
```

Linear regression

Quick view of the correlation between chosen variables

```
M2 = model_input(:,[221:230]); %230 is grouped biomass
R2 = corrcoef(log(M2+eps));
%eps takes care of the zeros in M2, prevents errors later on

%Try linear modeling
d = size(M2,1);
% sigma1, s1, a2, sigma2 (discarded a1 and s2)
%M2_s=[M2(:,2),M2(:,3),M2(:,4),ones(d,1)];
%B = M2(:,1);

%Try log approach
M2_s=[log(M2(:,1)),log(M2(:,2)),log(M2(:,3)),log(M2(:,4)),...
log(M2(:,5)),log(M2(:,8)),log(M2(:,9)),ones(d,1)];
M2_s(isinf(M2_s)) = 0;
B = log(M2(:,10) + eps); %B(B<0)=0;

coeff = M2_s\B; %inv not div
%coeff(isinf(coeff)) = 0; coeff(isnan(coeff)) = 0;
HB1 = M2_s*coeff; %HB1(isnan(HB1)) = 0;
%ind = find(HB1>=0);

RR = corrcoef([HB1,B]);

%RR = corrcoef([HB1(ind),B(ind)]);
R = RR(2,1)*RR(1,2);
%x = lsqlin(M2_s, B );

figure
%scatter(HB1(ind),B(ind),25)
scatter(HB1,B,25)
hold on
plot(0:10,0:10,'--r')
%axis([0 7 0 6])
text(0.2,5.25,['R^2=',num2str(R)] % title(['PC',num2str(b+1),' of wLiDAR']));
xlabel('Estimated biomass ln(g)')
ylabel('Measured biomass ln(g)')
```

```
title('Measured biomass vs. Estimated biomass')
grid on
```

E.2 PCA HSI Calculations

The following Matlab (version 7.8) algorithm was generated to find the PCA of concatenated HSI, and is a sub-algorithm to the herbaceous biomass modeling algorithm.

Contents

- PCA on many concatenated HSI images
- PCA the long way...
- Matlab PCA (the fast way!) and variance analysis
- Plot eigenvectors ordered by eigenvalues to see where bands contributions are
- Find which bands contributes most to each PC
- Form original size PCAs for each subset

```
function [pca_hsid1, pca_hsid2, pca_hsid3, pca_hsid5, pca_hsid6, pca_hsid7,...
        pca_hsid8, pca_hsid9, pca_hsid10, increment] = pca_hsi_landuse_bio()
```

```
% Performs PCA on the landuses of interest, as one image
%Reshaping them first to minimize padding
```

```
% Exports the pca corresponding to each landuse here under
```

```
%Diane Sarrazin
%24 June 2010
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

PCA on many concatenated HSI images

```
clear all; clc; close all;
```

```
%Read HSI images
hsid1 = freadenvi2('L8_H2_A280508_spectral_subset',...
'E:\Wu\Data\Landuse8\Spectral data\');
hsid2 = freadenvi2('L8_B260408_19_middle_spectral_subset',...
'E:\Wu\Data\Landuse8\Spectral data\');
hsid3 = freadenvi2('L7_G6_A020508_spectral_subset',...
'E:\Wu\Data\Landuse7\Spectral data\');
hsid5 = freadenvi2('L7_B250408_23_spectral_subset',...
'E:\Wu\Data\Landuse7\Spectral data\');
hsid6 = freadenvi2('L7_G2_A020508_spectral_subset',...
'E:\Wu\Data\Landuse7\Spectral data\');
hsid7 = freadenvi2('L2_A1_A270408_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');
```

```

hsid8 = freadenvi2('L2_A3_A270408_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');
hsid9 = freadenvi2('L2_A5_A250408_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');
hsid10 = freadenvi2('L2_A6_A290508_spectral_subset',...
'E:\Wu\Data\Landuse2\Spectral data\');

%Reshape each image in many lines, but 24 col
b = size(hsid1,3);
hsid1_r = reshape(hsid1,[size(hsid1,1)*size(hsid1,2),b]);
hsid2_r = reshape(hsid2,[size(hsid2,1)*size(hsid2,2),b]);
hsid3_r = reshape(hsid3,[size(hsid3,1)*size(hsid3,2),b]);
hsid5_r = reshape(hsid5,[size(hsid5,1)*size(hsid5,2),b]);
hsid6_r = reshape(hsid6,[size(hsid6,1)*size(hsid6,2),b]);
hsid7_r = reshape(hsid7,[size(hsid7,1)*size(hsid7,2),b]);
hsid8_r = reshape(hsid8,[size(hsid8,1)*size(hsid8,2),b]);
hsid9_r = reshape(hsid9,[size(hsid9,1)*size(hsid9,2),b]);
hsid10_r = reshape(hsid10,[size(hsid10,1)*size(hsid10,2),b]);

%Concatenate the reshaped files, many lines, 24 col
file_reshape = zeros(size(hsid1_r,1)+size(hsid2_r,1)+...
size(hsid3_r,1)+size(hsid5_r,1)+size(hsid6_r,1)...
+size(hsid7_r,1)+size(hsid8_r,1)+size(hsid9_r,1)+size(hsid10_r,1), b);
temp = 0;

file_reshape(1:size(hsid1_r,1),:) = hsid1_r;
temp = temp + size(hsid1_r,1);

file_reshape(temp+1:temp+size(hsid2_r,1),:) = hsid2_r;
temp = temp + size(hsid2_r,1);

file_reshape(temp+1:temp+size(hsid3_r,1),:) = hsid3_r;
temp = temp + size(hsid3_r,1);

file_reshape(temp+1:temp+size(hsid5_r,1),:) = hsid5_r;
temp = temp + size(hsid5_r,1);

file_reshape(temp+1:temp+size(hsid6_r,1),:) = hsid6_r;
temp = temp + size(hsid6_r,1);

file_reshape(temp+1:temp+size(hsid7_r,1),:) = hsid7_r;
temp = temp + size(hsid7_r,1);

file_reshape(temp+1:temp+size(hsid8_r,1),:) = hsid8_r;
temp = temp + size(hsid8_r,1);

```

```

file_reshape(temp+1:temp+size(hsid9_r,1),:) = hsid9_r;
temp = temp + size(hsid9_r,1);

file_reshape(temp+1:temp+size(hsid10_r,1),:) = hsid10_r;
temp = temp + size(hsid10_r,1);

sl=size(file_reshape,1); b1 = b;

```

PCA the long way...

```

image_cov = cov(file_reshape);
[v_original,d] = eig(image_cov); %v are eigenvectors(columns)
%d is the diagonal of generalized eigenvectors(eigenvalues)
[d_new,indices] = sort(diag(d),'descend');
v_orig = v_original(:,indices); v = v_orig';

PC_map = zeros(sl,b1);

for num = 1:sl

    x = file_reshape(num,:)' ;
    PC = v*x;
    PC_map(num,:) = PC';
    %brings the imp info in first PC

end

```

Matlab PCA (the fast way!) and variance analysis

```

[coeff, zscores_reshape, pcvars] = princomp(file_reshape);
%[v_orig,PC_map_HRPI,d_new]

%Find the amount of variance (in %) in each PC band
variance=zeros(1,b1);threshold=0;increment=0;
for band=1:b1
    if (threshold <= 99.5) %set threshold here
        increment=increment+1;
        variance(1,band)=pcvars(band)/(sum(pcvars))*100;
        threshold=sum(variance);
    end
end
end

```

Plot eigenvectors ordered by eigenvalues to see where bands contributions are

```
% x=394:28.5:1049.5;
% figure, plot(x,v_orig(:,[1,2,3]),'-');title('EIGENVECTORS vs. SPECTRAL RANGE');
% xlabel('Wavelength (nm)');ylabel('Eigenvectors');
%really height for lidar and spectral info for HSI
% legend('PC 1','PC 2','PC 3');
```

Find which bands contributes most to each PC

```
%band_contribution=zeros(b1,b1);
% [s_eig,indices]=sort(v_orig,'descend'); %or coeff instead of v_orig
% %indices(1:6,1:6)
%interesting to see how close numbers are together in same column...
% clear s_eig
```

Form original size PCAs for each subset

```
%for easier manipulation in biomass algorithm
temp = 0;

pca_hsid1_r = PC_map(1:size(hsid1_r,1),:);
pca_hsid1 = reshape(pca_hsid1_r, [size(hsid1,1), size(hsid1,2), b]);
temp = temp + size(hsid1_r,1);

pca_hsid2_r = PC_map(temp+1:temp+size(hsid2_r,1),:);
pca_hsid2 = reshape(pca_hsid2_r, [size(hsid2,1), size(hsid2,2), b]);
temp = temp + size(hsid2_r,1);

pca_hsid3_r = PC_map(temp+1:temp+size(hsid3_r,1),:);
pca_hsid3 = reshape(pca_hsid3_r, [size(hsid3,1), size(hsid3,2), b]);
temp = temp + size(hsid3_r,1);

pca_hsid5_r = PC_map(temp+1:temp+size(hsid5_r,1),:);
pca_hsid5 = reshape(pca_hsid5_r, [size(hsid5,1), size(hsid5,2), b]);
temp = temp + size(hsid5_r,1);

pca_hsid6_r = PC_map(temp+1:temp+size(hsid6_r,1),:);
pca_hsid6 = reshape(pca_hsid6_r, [size(hsid6,1), size(hsid6,2), b]);
temp = temp + size(hsid6_r,1);

pca_hsid7_r = PC_map(temp+1:temp+size(hsid7_r,1),:);
pca_hsid7 = reshape(pca_hsid7_r, [size(hsid7,1), size(hsid7,2), b]);
temp = temp + size(hsid7_r,1);
```

```

pca_hsid8_r = PC_map(temp+1:temp+size(hsid8_r,1),:);
pca_hsid8 = reshape(pca_hsid8_r, [size(hsid8,1), size(hsid8,2), b]);
temp = temp + size(hsid8_r,1);

pca_hsid9_r = PC_map(temp+1:temp+size(hsid9_r,1),:);
pca_hsid9 = reshape(pca_hsid9_r, [size(hsid9,1), size(hsid9,2), b]);
temp = temp + size(hsid9_r,1);

pca_hsid10_r = PC_map(temp+1:temp+size(hsid10_r,1),:);
pca_hsid10 = reshape(pca_hsid10_r, [size(hsid10,1), size(hsid10,2), b]);
temp = temp + size(hsid10_r,1);

%Display PCA

for bd=1:3 %may show more than increment1 number of images (6 img per figure)

    figure,
    subplot(3,3,1),imshow(pca_hsid1(:,:,bd),[])
    subplot(3,3,2),imshow(pca_hsid2(:,:,bd),[])
    subplot(3,3,3),imshow(pca_hsid3(:,:,bd),[])
    subplot(3,3,4),imshow(pca_hsid5(:,:,bd),[])
    subplot(3,3,5),imshow(pca_hsid6(:,:,bd),[])
    subplot(3,3,6),imshow(pca_hsid7(:,:,bd),[])
    subplot(3,3,7),imshow(pca_hsid8(:,:,bd),[])
    subplot(3,3,8),imshow(pca_hsid9(:,:,bd),[])
    subplot(3,3,9),imshow(pca_hsid10(:,:,bd),[])

end

```

E.3 PCA wLiDAR Calculations

The following Matlab (version 7.8) algorithm was generated to find the PCA of concatenated wLiDAR, and is a sub-algorithm to the herbaceous biomass modeling algorithm.

Contents

- PCA on many concatenated w images
- PCA BY MY OWN
- Matlab PCA and variance analysis
- Plot eigenvectors ordered by eigenvalues to see where bands contributions are
- Find which bands contributes most to each PC
- Form original size PCAs for each subset
- Clear a few variables


```
% Performs PCA on the landuses of interest, as one image - wLiDAR
```

```
% Performs PCA on the landuses of interest, as one image
%Reshaping them first to minimize padding
```

```
%Diane Sarrazin
```

```
%24 June 2010
```

```
%%%%%%%%%%
```

PCA on many concatenated w images

```
clear all; clc; close all;
```

```
% originally the following for 30 band wf:
```

```
%wd1_r = w_reduc_r(:,87:116);
```

```
% pos part = 87:100, neg part = 100:116 (ground at 100)
```

```
load peakd1; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd1_r = w_reduc_r(:,87:116);
load peakd2; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd2_r = w_reduc_r(:,87:116);
load peakd3; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd3_r = w_reduc_r(:,87:116);
load peakd5; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd5_r = w_reduc_r(:,87:116);
load peakd6; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd6_r = w_reduc_r(:,87:116);
load peakd7; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd7_r = w_reduc_r(:,87:116);
load peakd8; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd8_r = w_reduc_r(:,87:116);
load peakd9; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd9_r = w_reduc_r(:,87:116);
load peakd10; w_reduc_r = reshape(w_reduc,[size(w_reduc,1)*...
size(w_reduc,2),size(w_reduc,3)]); wd10_r = w_reduc_r(:,87:116);
```

```
clear w_reduc
```

```
file_reshape = zeros(size(wd1_r,1)+size(wd2_r,1)+...
```

```
size(wd3_r,1)+size(wd5_r,1)+size(wd6_r,1)...
```

```
+size(wd7_r,1)+size(wd8_r,1)+size(wd9_r,1)+size(wd10_r,1), size(wd1_r,2));
```

```
temp = 0; big_bd = size(file_reshape,2);
```

```
file_reshape(1:size(wd1_r,1),:) = wd1_r;
```

```
temp = temp + size(wd1_r,1);
```

```

file_reshape(temp+1:temp+size(wd2_r,1),:) = wd2_r;
temp = temp + size(wd2_r,1);

file_reshape(temp+1:temp+size(wd3_r,1),:) = wd3_r;
temp = temp + size(wd3_r,1);

file_reshape(temp+1:temp+size(wd5_r,1),:) = wd5_r;
temp = temp + size(wd5_r,1);

file_reshape(temp+1:temp+size(wd6_r,1),:) = wd6_r;
temp = temp + size(wd6_r,1);

file_reshape(temp+1:temp+size(wd7_r,1),:) = wd7_r;
temp = temp + size(wd7_r,1);

file_reshape(temp+1:temp+size(wd8_r,1),:) = wd8_r;
temp = temp + size(wd8_r,1);

file_reshape(temp+1:temp+size(wd9_r,1),:) = wd9_r;
temp = temp + size(wd9_r,1);

file_reshape(temp+1:temp+size(wd10_r,1),:) = wd10_r;
temp = temp + size(wd10_r,1);

sl=size(file_reshape,1); b1 = big_bd;
clear temp

```

PCA BY MY OWN

```

image_cov = cov(file_reshape);
[v_original,d] = eig(image_cov); %v are eigenvectors(columns)
%d is the diagonal of generalized eigenvectors(eigenvalues)
[d_new,indices] = sort(diag(d),'descend');
v_orig = v_original(:,indices); v = v_orig';

PC_map = zeros(sl,b1);

for num = 1:sl

    x = file_reshape(num,:);
    PC = v*x;
    PC_map(num,:) = PC';
    %brings the imp info in first PC

end

```

```
clear image_cov, clear v_original, clear d, clear d_new, clear indices
clear v, clear x, clear PC, clear PC, clear num
```

Matlab PCA and variance analysis

```
[coeff, zscores_reshape, pcvars] = princomp(file_reshape);
%[v_orig,PC_map_HRPI,d_new]

%Find the bands with most data
variance=zeros(1,b1);threshold=0;increment=0;
for band=1:b1
    if (threshold <= 99.5) %set threshold here
        increment=increment+1;
        variance(1,band)=pcvars(band)/(sum(pcvars))*100;
        threshold=sum(variance);
    end
end

%clear threshold, clear variance, clear band
%clear zscores_reshape, clear coeff
```

Plot eigenvectors ordered by eigenvalues to evaluate band contributions

```
x=1:big_bd;
figure, plot(x,v_orig(:,[1,2,3]),'-');title('EIGENVECTORS vs. TIME');
xlabel('Time (ns)');ylabel('Eigenvectors');
%really height for lidar and spectral info for w
%legend('PC 3','PC 6')
legend('Eig vect 1','Eig vect 2','Eig vect 3');
```

Find which bands contributes most to each PC

```
%band_contribution=zeros(b1,b1);
% [s_eig,indices]=sort(v_orig,'descend'); %or coeff instead of v_orig
% %indices(1:6,1:6) %interesting to see how close
%numbers are together in same column...
% clear s_eig
```

Form original size PCAs for each subset

```
load peakd1; wd1 = w_reduc; load peakd2; wd2 = w_reduc; load peakd3; wd3 = w_reduc;
load peakd5; wd5 = w_reduc; load peakd6; wd6 = w_reduc; load peakd7; wd7 = w_reduc;
load peakd8; wd8 = w_reduc; load peakd9; wd9 = w_reduc; load peakd10; wd10 = w_reduc;
clear w_reduc
```

```

temp = 0;

pca_wd1_r = PC_map(1:size(wd1_r,1),:);
pca_wd1 = reshape(pca_wd1_r, [size(wd1,1), size(wd1,2), big_bd]);
temp = temp + size(wd1_r,1);

pca_wd2_r = PC_map(temp+1:temp+size(wd2_r,1),:);
pca_wd2 = reshape(pca_wd2_r, [size(wd2,1), size(wd2,2), big_bd]);
temp = temp + size(wd2_r,1);

pca_wd3_r = PC_map(temp+1:temp+size(wd3_r,1),:);
pca_wd3 = reshape(pca_wd3_r, [size(wd3,1), size(wd3,2), big_bd]);
temp = temp + size(wd3_r,1);

pca_wd5_r = PC_map(temp+1:temp+size(wd5_r,1),:);
pca_wd5 = reshape(pca_wd5_r, [size(wd5,1), size(wd5,2), big_bd]);
temp = temp + size(wd5_r,1);

pca_wd6_r = PC_map(temp+1:temp+size(wd6_r,1),:);
pca_wd6 = reshape(pca_wd6_r, [size(wd6,1), size(wd6,2), big_bd]);
temp = temp + size(wd6_r,1);

pca_wd7_r = PC_map(temp+1:temp+size(wd7_r,1),:);
pca_wd7 = reshape(pca_wd7_r, [size(wd7,1), size(wd7,2), big_bd]);
temp = temp + size(wd7_r,1);

pca_wd8_r = PC_map(temp+1:temp+size(wd8_r,1),:);
pca_wd8 = reshape(pca_wd8_r, [size(wd8,1), size(wd8,2), big_bd]);
temp = temp + size(wd8_r,1);

pca_wd9_r = PC_map(temp+1:temp+size(wd9_r,1),:);
pca_wd9 = reshape(pca_wd9_r, [size(wd9,1), size(wd9,2), big_bd]);
temp = temp + size(wd9_r,1);

pca_wd10_r = PC_map(temp+1:temp+size(wd10_r,1),:);
pca_wd10 = reshape(pca_wd10_r, [size(wd10,1), size(wd10,2), big_bd]);
temp = temp + size(wd10_r,1);

%Display PCA

for bd=1:3 %may show more than increment1 number of images (6 img per figure)

    figure,
    subplot(3,3,1),imshow(pca_wd1(:,:,bd),[]),title(['PC',num2str(bd),' of w 1']);
    subplot(3,3,2),imshow(pca_wd2(:,:,bd),[]),title(['PC',num2str(bd),' of w 2']);

```

```

subplot(3,3,3),imshow(pca_wd3(:,:,bd),[]),title(['PC',num2str(bd),' of w 3']);
subplot(3,3,4),imshow(pca_wd5(:,:,bd),[]),title(['PC',num2str(bd),' of w 4']);
subplot(3,3,5),imshow(pca_wd6(:,:,bd),[]),title(['PC',num2str(bd),' of w 5']);
subplot(3,3,6),imshow(pca_wd7(:,:,bd),[]),title(['PC',num2str(bd),' of w 6']);
subplot(3,3,7),imshow(pca_wd8(:,:,bd),[]),title(['PC',num2str(bd),' of w 7']);
subplot(3,3,8),imshow(pca_wd9(:,:,bd),[]),title(['PC',num2str(bd),' of w 8']);
subplot(3,3,9),imshow(pca_wd10(:,:,bd),[]),title(['PC',num2str(bd),' of w 9']);

end

```

Clear a few variables

```

clear pca_wd1_r, clear pca_wd2_r, clear pca_wd3_r, clear pca_wd5_r, clear pca_wd6_r
clear pca_wd7_r, clear pca_wd8_r, clear pca_wd9_r, clear pca_wd10_r, clear big_bd
clear bd, clear b1; clear wd1, clear wd2, clear wd3, clear wd5, clear wd6, clear wd7
clear wd8, clear wd9, clear wd10, clear wd1_r, clear wd2_r, clear wd3_r, clear wd5_r
clear wd6_r, clear wd7_r, clear wd8_r, clear wd9_r, clear wd10_r; clear temp, clear sl
clear file_reshape, clear PC_map, clear bd, clear band, clear x, clear zscores_reshape

```

E.4 Smoothing of imagery and feature metrics based on a 3x3 window

The following Matlab (version 7.8) algorithm was generated to find the average value in a 3x3 region for each band of a specified feature variable, and is a sub-algorithm to the herbaceous biomass modeling algorithm.

Contents

- Returns the mean of 3x3 region around each GPS_points for an image

```
function [val] = GPS(fname,GPS_points);
```

Returns the mean of 3x3 region around each GPS_points for an image

```
% fname is variable from which to extract mean 3x3 values for each GPS point
% GPS_points is variable from which to extract GPS_points

```

```

%Diane Sarrazin
%24 June 2010
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

temp = zeros(size(fname,3),9);
val = zeros(size(GPS_points,1),size(fname,3));

```

```

for i=1:size(GPS_points,1)

    temp(:,1) = squeeze(fname(GPS_points(i,3),GPS_points(i,2),:));
    temp(:,2) = squeeze(fname(GPS_points(i,3)+1,GPS_points(i,2)+1,:));
    temp(:,3) = squeeze(fname(GPS_points(i,3)-1,GPS_points(i,2)-1,:));
    temp(:,4) = squeeze(fname(GPS_points(i,3)+1,GPS_points(i,2),:));
    temp(:,5) = squeeze(fname(GPS_points(i,3)-1,GPS_points(i,2),:));
    temp(:,6) = squeeze(fname(GPS_points(i,3),GPS_points(i,2)-1,:));
    temp(:,7) = squeeze(fname(GPS_points(i,3)+1,GPS_points(i,2)-1,:));
    temp(:,8) = squeeze(fname(GPS_points(i,3),GPS_points(i,2)+1,:));
    temp(:,9) = squeeze(fname(GPS_points(i,3)-1,GPS_points(i,2)+1,:));

    val(i,:) = mean(temp,2)';
    %the mean value of a 3x3 for each point
    %in each band in each row of val then transpose

    %      x=1:size(fname,3);
%      figure,
%      subplot(3,3,1), plot(x,temp(:,1)')
%      subplot(3,3,2), plot(x,temp(:,2)')
%      subplot(3,3,3), plot(x,temp(:,3)')
%      subplot(3,3,4), plot(x,temp(:,4)')
%      subplot(3,3,5), plot(x,temp(:,5)')
%      subplot(3,3,6), plot(x,temp(:,6)')
%      subplot(3,3,7), plot(x,temp(:,7)')
%      subplot(3,3,8), plot(x,temp(:,8)')
%      subplot(3,3,9), plot(x,temp(:,9)')

end
clear i; clear temp

```

E.5 Building Plot-level Composite Waveforms

The following Matlab (version 7.8) algorithm was generated to find the 3x3 composite waveforms of a specified wLiDAR variable, and is a sub-algorithm to the herbaceous biomass modeling algorithm. It also finds the signal difference between a each individual composite waveforms and a dirt road composite waveform.

```

function [val, hbdiff] = GPScomp_bio(fname,GPS_points,numpeaks);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Build composite waveforms
% Returns the composite waveform w for each plot, and the difference between
% comp wf and road.

```

```

% fname is variable containing composite waveforms in each row
% GPS_points is variable from which to extract GPS_points
% numpeaks is matrix of number of peaks in each waveform

%Waveforms that have more than 2 peaks are removed from the composite
%waveform calculation.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

val = zeros(size(GPS_points,1),size(fname,3));
hbdiff = zeros(size(GPS_points,1),size(fname,3));

%Form road cropped waveform
load road_w_norm;
[max_road ind_road] = max(road_w_norm,[],2);
bin_h = 13; be = 17;
temp_road = road_w_norm(1,ind_road-bin_h:ind_road+be-1);

x1 = 1:size(road_w_norm,2); x = 1:size(fname,3);
% figure, subplot(1,2,1), plot(x1, road_w_norm), title('Road wf')
% subplot(1,2,2), plot(x,temp_road); title('Cropped road wf')

%Identify each waveform in the 3x3 and eliminate those with more than 2 peaks
for i=1:size(GPS_points,1)

    temp = zeros(size(fname,3),9);

    b = GPS_points(i,3)-1;
    c = GPS_points(i,2)-1;
    if numpeaks(b,c) <= 2
        temp(:,1) = squeeze(fname(b,c,:));
    end

    b = GPS_points(i,3)-1;
    c = GPS_points(i,2);
    if numpeaks(b,c) <= 2
        temp(:,2) = squeeze(fname(b,c,:));
    end

    b = GPS_points(i,3)-1;
    c = GPS_points(i,2)+1;
    if numpeaks(b,c) <= 2
        temp(:,3) = squeeze(fname(b,c,:));
    end

    b = GPS_points(i,3);
    c = GPS_points(i,2)-1;

```

```

if numpeaks(b,c) <= 2
    temp(:,4) = squeeze(fname(b,c,:));
end

b = GPS_points(i,3);
c = GPS_points(i,2);
if numpeaks(b,c) <= 2
    temp(:,5) = squeeze(fname(b,c,:));
end

b = GPS_points(i,3);
c = GPS_points(i,2)+1;
if numpeaks(b,c) <= 2
    temp(:,6) = squeeze(fname(b,c,:));
end

b = GPS_points(i,3)+1;
c = GPS_points(i,2)-1;
if numpeaks(b,c) <= 2
    temp(:,7) = squeeze(fname(b,c,:));
end

b = GPS_points(i,3)+1;
c = GPS_points(i,2);
if numpeaks(b,c) <= 2
    temp(:,8) = squeeze(fname(b,c,:));
end

b = GPS_points(i,3)+1;
c = GPS_points(i,2)+1;
if numpeaks(b,c) <= 2
    temp(:,9) = squeeze(fname(b,c,:));
end

val(i,:) = mean(temp,2)';
clear tempo

%Re-align peak of comp with bin 14
[max_val max_ind] = max(val(i,:),[],2); %stats of comp wf for next phase
tempo = zeros(1,be+bin_h);

b = max_ind-bin_h; %2m AG
c = max_ind+be-1; %multiple scattering bellow ground peak (after)

empty_matrix = zeros(size(val(i,:)));
if val(i,:) == empty_matrix

```



```

        tempo = empty_matrix;

elseif b <= 0
    diff = size(val,2) - (c);
    tempo(1,abs(b)+1:end) = padarray(val(i,1:c),[0 diff],'pre');

elseif c > size(val,2)
    diff = c-size(val,2);
    tempo(1,1:end-diff) = val(i,b:end);

else
    tempo(1,:) = val(i,b:c);
end

val(i,:) = tempo;

%Re-normalize by area cause area might have changed
A=sum(val(i,:),2);
val(i,:) = val(i,:)./ A;

%Find difference between composite wf and road wf
hbdiff1 = val(i,)-temp_road;
hbdiff1(hbdiff1 < 0) = 0; hbdiff1(isnan(hbdiff1)) = 0;
hbdiff(i,:) = hbdiff1;

%    figure, plot(x,val(i,:), 'r'), hold on
%    plot(x,temp_road), hold on
%    plot(x,hbdiff(i,:), 'g'), legend('Comp wf','Road wf','Diff')
%    title('Cropped composite wf, road wf and difference between the two')
%    ylabel('Intensity'), xlabel('Time bins')

% COMPARE TO NOT DECONVOLVED WF
% compare waveforms (pre and post noise filtering)
%    x1 = 1:size(wf,3);
%    b = GPS_points(i,3);
%    c = GPS_points(i,2);
%    figure
%    subplot(2,1,1)
%    plot(squeeze(wf(b,c,:))), title('Before deconvolution')
%    xlabel('Time (ns)')
%    ylabel('Intensity')
%    subplot(2,1,2)
%    plot(squeeze(fname(b,c,:))),
%    title('After deconvolution and area normalization')
%    xlabel('Time (ns)')

```

```

%       ylabel('Intensity')
%
%
%
%       x=1:size(fname,3);
%       figure,
%       subplot(3,6,1), plot(x,temp(:,1))
%       subplot(3,6,2), plot(x,temp(:,2)),
%       title(['3x3 wf region']); % around GPS point',num2str(i)])
%       subplot(3,6,3), plot(x,temp(:,3))
%       subplot(3,6,7), plot(x,temp(:,4))
%       subplot(3,6,8), plot(x,temp(:,5))
%       subplot(3,6,9), plot(x,temp(:,6))
%       subplot(3,6,13), plot(x,temp(:,7))
%       subplot(3,6,14), plot(x,temp(:,8))
%       subplot(3,6,15), plot(x,temp(:,9))
%       subplot(3,6,[4 5 6 10 11 12 16 17 18]), plot(x,val(i,:)), hold on
%       plot(x,temp_road,'r'), hold on
%       plot(x,hbdiff(i,:), 'g'), legend('Comp wf','Road wf', 'Difference')
%       %ylabel('Intensity'), xlabel('Time bins')
%       title(['Composite wf']); % for plot',num2str(i)])
%       clear temp; clear temp1; clear ind_val, clear tempo

end

val(isnan(val)) = 0;
clear i; clear ii;

```

The following Matlab (version 7.8) algorithm was generated to find the Gaussian approximation from each composite waveform, and is a sub-algorithm to the herbaceous biomass modeling algorithm. The output of this function is the amplitude, the standard deviation (sigma), the sum over each band or area under the curve, and the peak location for a two-peaks or fewer composite waveform.

E.6 Gaussian Modeling of Waveforms

Contents

- Performs gaussian modeling of waveforms that have two peaks or less - ground
- EM decomposition

```
function [M2] = em_function(GPS_points, fname)
```

Performs gaussian modeling of waveforms that have two peaks or less - ground

```
%Thus no need to flipplr(fname_r)
```

```
% Inputs are: GPS_points, waveforms in each row of a matrix
```

```
%Outputs are: M2 [n x 8]
```

```
% col 1: A, amplitude of gaussian component 1
```

```
% col 2: Sigma, standard deviation of gaussian component 1
```

```
% col 3: S, Integrated area of gaussian curve 1
```

```
%col 4: mu or peak location of modeled gaussian
```

```
%col 5: amplitude of modeled gaussian 2
```

```
%col 6: sigma of modeled gaussian 2
```

```
%col 7: cumsum of 2nd modeled gaussian
```

```
%col 8 : mu or peak location of 2nd modeled gaussian
```

```
% Diane Sarrazin
```

```
% 24 June 2010
```

```
%%%%%%%%%%
```

EM decomposition

```
x=1:size(fname,2);
```

```
length = size(fname,2);
```

```
nbStates = 2;
```

```
Results = [];
```

```
curve_true = [];
```

```
curve_comp1 = [];
```

```
curve_comp2 = [];
```

```
%For each waveform, perform gaussian modeling
```

```
for i = 1:size(GPS_points,1)
```

```
    Data = [];
```

```
    pdf =fname(i,:).*100; %gaussian approx algorithm doesn't like small values -
```

```
    %so normalized waveforms *100
```

```
    [lmval,indd] = lmax(pdf,1);
```

```
    %Remove small peaks
```

```
    c = find(lmval >= 0.8); %%%
```

```
    lmval_temp = zeros(size(c));
```

```
    indd_temp = zeros(size(c));
```

```
    for ii = 1:size(c,2)
```

```
        lmval_temp(ii) = lmval(c(ii));
```

```
        indd_temp(ii) = indd(c(ii));
```

```
    end
```

```
    lmval = lmval_temp;
```

```
    indd = indd_temp;
```

```

if (sum(pdf)>0 && (numel(indd)==1))

    [Sigma,Mu,A] = mygaussfit(x,pdf,0.2);
    y=A*exp(-(x-Mu).^2/(2*Sigma^2))+rand(size(x))*0.3;
    temp(1) = A;
    temp(2) = Sigma^0.5;

    if (Mu+Sigma) < size(fname,2) && (Mu-Sigma) > size(fname,2)
        temp(3) = sum(y(1,floor(Mu-Sigma):ceil(Mu+Sigma)));
        %or sum(pdf(1:length,Mu,Sigma));
    else
        temp(3) = 0;
    end

    temp(4) = Mu;
    temp(5) = 0;
    temp(6) = 0;
    temp(7) = 0;
    temp(8) = 0;
    Results = [Results;temp];
    curve_true = [curve_true;pdf];
    curve_comp1 = [curve_comp1;y];
    curve_comp2 = [curve_comp2;zeros(size(pdf))];
    clear temp

%find first valley before second to last peak if more than 2 peaks
elseif size(lmval,2) >= 3 %

    a = size(lmval,2)-1; %before last peak
    b = size(lmval,2)-2; %two before last peak
    e = min (pdf(indd(b):indd(a)) );
    c = find( pdf == e ); %indices of where the valley is

    %There can be two loc with same value in pdf
    %indices of c in range: second to last peak:last peak
    d = find( (c >= indd(b)) & (c <= indd(a)) );
    valley_loc = c(round(median(d))); %choose middle bin if c all the same
    %valley = pdf(valley_loc,1); %value of valley

    %We want to get info from two last peaks (including ground peak)
    temp_pdf = zeros(size(pdf));
    temp_pdf(valley_loc:end) = pdf(valley_loc:end);
    pdf = temp_pdf;
    [lmval,indd] = lmax(pdf,1);

```

```

clear a; clear b; clear c; clear d; clear e

scale = 1/sum(pdf);
pdf = scale*pdf;
numbers = 10000*pdf;

for j = 1:length

    if round(numbers(j)) > 0
        Data = [Data j*ones(1,round(numbers(j)))];
    end

end

Mu0 = [indd]; %location of the peaks in the waveform
Sigma0 = ones(1,1,2); %standard deviation of each peak
Sigma0(1) = 2.5;
Sigma0(2) = 4;
Priors0 = [0.9 0.1];
%magnitude ratio between single scatt and multiple scatt
[Priors, Mu, Sigma, Pix] = EM(Data, Priors0, Mu0, Sigma0);
[CO,I0] = max(Priors);
[C1,I1] = min(Priors);
temp(1) = lmval(1); %value of first peak
temp(2) = Sigma(I0)^0.5; %max loc of priors
temp(3) = sum(1/scale*Priors(I0)*normpdf(1:length,Mu(I0),Sigma(I0)^0.5));
temp(4) = Mu(I0);
temp(5) = lmval(2); %value of second peak
temp(6) = Sigma(I1)^0.5; %min loc of priors
temp(7) = sum(1/scale*Priors(I1)*normpdf(1:length,Mu(I1),Sigma(I1)^0.5));
temp(8) = Mu(I1); %location of second peak
Results = [Results;temp];
curve_true = [curve_true;1/scale*pdf];
curve_comp1 = [curve_comp1;1/scale*Priors(I0)*...
normpdf(1:length,Mu(I0),Sigma(I0)^0.5)];
curve_comp2 = [curve_comp2;1/scale*Priors(I1)*...
normpdf(1:length,Mu(I1),Sigma(I1)^0.5)];
clear temp

elseif (sum(pdf)>0 && (numel(indd)==2))
    %build probability density fct of 10 000 points representing waveform
    scale = 1/sum(pdf);
    pdf = scale*pdf;
    numbers = 10000*pdf;

```

```

for j = 1:length

    if round(numbers(j)) > 0
        Data = [Data j*ones(1,round(numbers(j)))];
    end

end

Mu0 = [indd]; %location of the peaks in the waveform
Sigma0 = ones(1,1,2); %standard deviation of each peak
Sigma0(1) = 2.5;
Sigma0(2) = 4;
Priors0 = [0.9 0.1];
%magnitude ratio between single scatt and multiple scatt
[Priors, Mu, Sigma, Pix] = EM(Data, Priors0, Mu0, Sigma0);
[C0,I0] = max(Priors);
[C1,I1] = min(Priors);
temp(1) = lmval(1); %value of first peak
temp(2) = Sigma(I0)^0.5; %max loc of priors
temp(3) = sum(1/scale*Priors(I0)*normpdf(1:length,Mu(I0),Sigma(I0)^0.5));
temp(4) = Mu(I0);
temp(5) = lmval(2); %value of second peak
temp(6) = Sigma(I1)^0.5; %min loc of priors
temp(7) = sum(1/scale*Priors(I1)*normpdf(1:length,Mu(I1),Sigma(I1)^0.5));
temp(8) = Mu(I1); %location of second peak
Results = [Results;temp];
curve_true = [curve_true;1/scale*pdf];
curve_comp1 = [curve_comp1;1/scale*...
Priors(I0)*normpdf(1:length,Mu(I0),Sigma(I0)^0.5)];
curve_comp2 = [curve_comp2;1/scale*...
Priors(I1)*normpdf(1:length,Mu(I1),Sigma(I1)^0.5)];
clear temp

else

temp(1) = 0;
temp(2) = 0;
temp(3) = 0;
temp(4) = 0;
temp(5) = 0;
temp(6) = 0;
temp(7) = 0;
temp(8) = 0;

```

```

        Results = [Results;temp];
        curve_true = [curve_true;pdf];
        curve_comp1 = [curve_comp1;zeros(size(pdf))];
        curve_comp2 = [curve_comp2;zeros(size(pdf))];
        clear temp

    end

end

% for i=1:size(curve_true,1)
%     figure,plot(x,curve_true(i,:), 'k'), hold on
%     plot(x,curve_comp1(i,:), '--r'), hold on
%     plot(x,curve_comp2(i,:), '--b'),
%     title(['Gaussian modeling of GPS point',num2str(i)]);
%     xlabel('time (ns)'),ylabel('Intensity'),
%     legend('Composite waveform', 'Gaussian approx1', 'Gaussian approx2')
% end

Results(isnan(Results)) = 0;
M2 = Results;

```

E.7 Rise Time to Leading Edge and Fall Time of Trailing Edge of Waveforms

The following Matlab (version 7.8) algorithm was generated to find the rise time to the leading edge and fall time of the trailing edge of composite waveforms, and is a sub-algorithm to the herbaceous biomass modeling algorithm.

Contents

- Rise time to leading edge and fall time of falling edge

```
function [rt, ft] = rise_fall(fname)
```

Rise time to leading edge and fall time of falling edge

```
%Inspired by Joe McGlinchy's algorithm
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
rt = zeros(size(fname,1),1);
ft = zeros(size(fname,1),1);
```

```
for i = 1:size(fname,1)
```

```

top = find(fname(i,:)==max(fname(i,:)));
if isempty(top)
    rt(i,1)=0;
    ft(i,1)=0;
else
    top1 = top(1);
    topf=top(end);

    bot=find(fname(i,:)>=0.1*max(fname(i,:)));
    a=find(bot<125);
    bot=bot(a);
    bot1=bot(1);
    botf=bot(end);

    rt(i,1)=top1-bot1;
    ft(i,1)=botf-topf;
end

clear top topf bot botf top1 bot1
end

% x = 1:size(fname,2);
% figure, plot(x, fname(1,:))

```


Bibliography

- [Abdullah et al., 2001] Abdullah, M., Guan, L., and Azemi, B. M. (2001). Stepwise discriminant analysis for colour grading of oil palm using machine vision system. *Food and Bioproducts Processing*, 79(4):223–231.
- [Aldred and Bonner, 1985] Aldred, A. and Bonner, M. (62 pp., 1985). Application of airborne lasers to forest surveys. Information Report PI-X-51, Canadian Forestry Service, Petawawa National Forestry Centre.
- [Archer, 1995] Archer, S. (1995). Tree-grass dynamics in a prosopis-thornbush savanna parkland: reconstructing the past and predicting the future. *Journal of Ecoscience*, 2(1):83–99.
- [Asner et al., 2007] Asner, G., Knapp, D., Kennedy-Bowdoin, T., Jones, M., Martin, R., Boardman, J., and Field, C. (2007). Carnegie Airborne Observatory: in flight fusion of hyperspectral imaging and waveform light detection and ranging (wLIDAR) for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, 1(013536).
- [Asner et al., 2003] Asner, G. P., Borghi, C. E., and Ojeda, R. A. (2003). Desertification in central Argentina: changes in ecosystem carbon and nitrogen from imaging spectroscopy. *Ecological Applications*, 13(3):629–648.
- [CAO, 2008] CAO (2008). Carnegie Airborne Observatory, remote measurement of biospheric structure and function, <http://cao.stanford.edu>.
- [Carleer and Wolff, 2004] Carleer, A. and Wolff, E. (2004). Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering and Remote Sensing*, 70(1):135–140.
- [Cho et al., 2009] Cho, M. A., Debba, P., Mathieu, R., Naidoo, L., Aardt, J., and Asner, G. (2009). Multiple endmember spectral angle mapper (SAM) analysis improves discrimination of savanna tree species: canopy level analysis. In *1st Workshop on Hyperspectral Image and Signal processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4, Grenoble, France, 26–28 August.
- [Clark et al., 2005] Clark, M. L., Roberts, D. A., and Clark, D. B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Journal of Remote Sensing of Environment*, 96(3-4):375–398.
- [Cohen and Spies, 1992] Cohen, W. B. and Spies, T. A. (1992). Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, 41(1):1–17.

- [Congalton and Green, 1999] Congalton, R. G. and Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practices*. Lewis Publishers, Florida, first edition.
- [Dalponte et al., 2008] Dalponte, M., Bruzzone, L., and Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1416–1427.
- [Drake et al., 2002] Drake, J., Dubayah, R., Knox, R., Clark, D., and Blair, J. (2002). Sensitivity of large-footprint LIDAR to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, 81:378–392.
- [Dubayah and Drake, 2000] Dubayah, R. O. and Drake, J. B. (2000). LIDAR remote sensing for forestry. *Journal of Forestry*, 98:44–46.
- [Easton, 2008] Easton, R. (2008). *Linear mathematics with applications to imaging*. Cheston F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive.
- [ENVI, 2008] ENVI (2008). Envi help and documentation.
- [Fernandez, 2003] Fernandez, G. (2003). *Data mining using SAS applications*. Chapman and Hall CRC, Florida.
- [Flood and Gutelius, 1997] Flood, M. and Gutelius, B. (1997). Commercial implications of topographic terrain mapping using scanning airborne laser radar. *Photogrammetric Engineering and Remote Sensing*, LXIII(327-329):363–366.
- [Foody, 2004] Foody, G. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5):627–633.
- [Foody and Hill, 1996] Foody, G. M. and Hill, R. A. (1996). Classification of tropical forest classes from Landsat TM data. *International Journal of Remote Sensing*, 17(12):2353–2367.
- [Franklin, 1994] Franklin, S. (1994). Discrimination of subalpine forest species and canopy density using digital CASI, SPOT PLA and Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 60:1233–1241.
- [Gong et al., 1997] Gong, P., Pu, R., and Yu, B. (1997). Conifer species recognition: an exploratory analysis of in situ hyperspectral data. *Remote Sensing of Environment*, 62(2):189–200.
- [Harding et al., 2001] Harding, D., Lefsky, M., Parker, G., and Blair, J. (2001). Laser altimeter canopy height profiles: methods and validation for closed canopy, broadleaf forests. *Remote Sensing of Environment*, 76(3):283–297.
- [Holmgren and Persson, 2004] Holmgren, J. and Persson, A. (2004). Identifying species of individual trees using airborne laser scanner. *Journal of Remote Sensing of Environment*, 90(4):415–423.
- [Hyde et al., 2005] Hyde, P., Dubayah, R., Peterson, B., Blair, J. B., Hofton, M., Hunsaker, C., Knox, R., and Walker, W. (2005). Mapping of forest structure for wildlife habitat analysis using waveform LIDAR: validation of montane ecosystems. *Remote Sensing of Environment*, 96(3-4):427–437.
- [J. Wu, unpb] J. Wu, J. van Aardt, G. (unpb). A comparison of signal deconvolution algorithms based on small-footprint lidar waveform simulation. Unpublished TGARSS paper.

- [Lefsky et al., 1999a] Lefsky, M., Harding, D., Cohen, W., Parker, G., and Shugart, H. (1999a). Surface LIDAR remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, 67:83–98.
- [Lefsky et al., 1999b] Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., and Harding, D. (1999b). LIDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Journal of Remote Sensing of Environment*, 70(3):339–361.
- [Lillesand et al., 2008] Lillesand, T. M., Kiedfer, R. W., and Chipman, J. W. (2008). *Remote sensing and image interpretation*. John Wiley and Sons, Inc, New York, sixth edition.
- [Lim et al., 2003] Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. (2003). LIDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1):88–106.
- [Lucy, 1974] Lucy, L. (1974). An iterative technique for the rectification of observed distributions. *The Astronomy Journal*, 79(6):745–754.
- [Mallet and Bretar, 2008] Mallet, C. and Bretar, F. (2008). Full-waveform topographic LIDAR: State-of-the-art. *Journal of Photogrammetry and Remote Sensing*, 64:1–16.
- [Martin and Aber, 1997] Martin, M. and Aber, J. (1997). Estimation of forest canopy lignin and nitrogen concentration and ecosystem processes by high spectral resolution remote sensing. *Ecological Applications*, 7(2):431–443.
- [Martin et al., 1998] Martin, M., Newman, S., Aber, J., and Congalton, R. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, 65(3):249–254.
- [Means et al., 1999] Means, J. E., Acker, S. A., Harding, D. J., and Blair, J. B. (1999). Use of large-footprint scanning airborne LIDAR to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sensing of Environment*, 67:298–308.
- [Mirik et al., 2005a] Mirik, M., Norland, J. E., Crabtree, R. L., and Biondini, M. E. (2005a). Hyperspectral one-meter-resolution remote sensing in Yellowstone National Park, Wyoming: I. Forage nutritional values. *Rangeland Ecology and Management*, 58:452–458.
- [Mirik et al., 2005b] Mirik, M., Norland, J. E., Crabtree, R. L., and Biondini, M. E. (2005b). Hyperspectral one-meter-resolution remote sensing in Yellowstone National Park, Wyoming: II. Biomass. *Rangeland Ecology and Management*, 58(5):459–465.
- [Mutlu et al., 2008] Mutlu, M., Popescu, S., Stripling, C., and Spencer, T. (2008). Mapping surface fuel models using LIDAR and multispectral data fusion for fire behavior. *Remote Sensing of Environment*, 112(1):274–285.
- [Nagelkerke, 1991] Nagelkerke, N. (1991). *A note on a general definition of the coefficient of determination*, volume 78. Biometrika Trust.
- [Nayegandhi et al., 2006] Nayegandhi, A., Brock, J. C., Wright, C. W., and O’Connell, M. J. (2006). Evaluating a small footprint, waveform-resolving LIDAR over coastal vegetation communities. *Photogrammetric Engineering and Remote Sensing*, 72(12):1407–1417.

- [Neuenschwander et al., 2009] Neuenschwander, A., Magruder, L., and M. Tyler (2009). Landcover classification of small-footprint, full-waveform LIDAR data. *Journal of Applied Remote Sensing*, 3(033544).
- [Ozer, 1985] Ozer, D. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97(2):307–315.
- [Peterson et al., 2005] Peterson, B., Dubayah, R., Hyde, P., Hofton, M., Blair, J. B., and Fites-Kaufman, J. (2005). Use of lidar for forest inventory and forest management application. In *Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium*, pages 193–202.
- [Popescu et al., 2003] Popescu, S., Wynne, R., and Nelson, R. (2003). Measuring individual tree crown diameter with LIDAR and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5):564–577.
- [Popescu et al., 2004] Popescu, S., Wynne, R., and Scrivani, J. (2004). Fusion of small-footprint LIDAR and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. *Forest Science*, 50(4):551–565.
- [Reitberger et al., 2006] Reitberger, J., Krzystek, P., and Stilla, U. (2006). Analysis of full waveform LIDAR data for tree species classification. In *ISPRS Symposium of Commission III*, volume 7. International Society for Photogrammetry and Remote Sensing.
- [Roberts et al., 2008] Roberts, J. W., van Aardt, J., and Ahmed, F. (2008). Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(023522).
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G., and Williams, R. (1986). *Learning internal representations by error propagation*, volume 1. MIT Press, Cambridge, MA.
- [Sarrazin et al., 2010] Sarrazin, D., van Aardt, J., Asner, G., McGlinchy, J., Messinger, D., and Jiaying, W. (2010). Fusing waveform lidar and hyperspectral data for species-level structural assessment in savanna ecosystems. In Turner, M. D. and Kamerman, G., editors, *Laser Radar Technology and Applications XV*, volume 7684, pages 76841H–76841H10.
- [SAS Institute Inc., 2008] SAS Institute Inc., Cary, N. S. I. I. . (2008). Sas 9.2 help and documentation.
- [Schiever and Congalton, 1995] Schiever, J. R. and Congalton, R. (1995). Evaluating seasonal variability as an aid to cover-type mapping from landsat thematic mapper data in the northeast. *Photogrammetric Engineering and Remote Sensing*, 61(3):321–327.
- [Scholes, 1997] Scholes, R. J. (1997). *Vegetation of Southern Africa*, chapter 11, pages 258–277. The Press Syndicate of the University of Cambridge, Cape Town, South Africa.
- [Schott, 2007] Schott, J. (2007). *Remote sensing, the imaging chain approach*. Oxford University Press, second edition.
- [Seielstad and Queen, 2003] Seielstad, C. A. and Queen, L. (2003). Using airborne laser altimetry to determine fuel models for estimating fire behavior. *Journal of Forestry*, 101(4):10–15.
- [Steininger, 2000] Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing*, 21(6-7):1139–1157.

- [Sugumaran and Voss, 2007] Sugumaran, R. and Voss, M. (2007). Object-oriented classification of LIDAR-fused hyperspectral imagery for tree species identification in an urban environment. In *IEEE Proceedings of Urban Remote Sensing Joint Event 2007*, volume April 11-13, page 6pp, Paris, France.
- [Trodd, 1995] Trodd, N. M. (1995). Uncertainty in land cover mapping for modelling land cover change. In *Proceedings RSS95: Remote Sensing in Action*, pages 1138–1145. Remote Sensing Society, Nottingham, 11-14 September, Southampton, United Kingdom.
- [Underwood et al., 2003] Underwood, E., Ustin, S., and Dipietro, D. (2003). Mapping non-native plants using hyperspectral imagery. *Remote Sensing of Environment*, 86(2):150–161.
- [Ustin et al., 2004] Ustin, S., Roberts, D., Gamon, J., Asner, G., and Green, R. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, 54(6):523–534.
- [van Aardt and Wynne, 2001] van Aardt, J. and Wynne, R. (2001). Spectral separability among six southern tree species. *Photogrammetric Engineering and Remote Sensing*, 67:1367–1375.
- [van Aardt and Wynne, 2007] van Aardt, J. and Wynne, R. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing*, 28(1-2):431–436.
- [van Aardt et al., 2006] van Aardt, J., Wynne, R. H., and Oderwald, R. G. (2006). Forest volume and biomass estimation using small-footprint LIDAR-distributional parameters of a per-segment basis. *Forest Science*, 52(6):636–649.
- [Wagner et al., 2006] Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitizing airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(2):100–112.
- [Willmott and Matsuura, 2005] Willmott, C. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30:79–82.
- [Wu et al., 2009a] Wu, J., van Aardt, J., Asner, G., Mathieu, R., Kennedy-Bowdoin, T., Knapp, D., Wessels, K., Erasmus, B., and Smit, I. (2009a). Connecting the dots between laser waveforms and herbaceous biomass for assessment of land degradation using small-footprint waveform LIDAR data. *IGARSS*, II:334–337.
- [Wu et al., 2009b] Wu, J., van Aardt, J., Asner, G. P., Mathieu, R., Kennedy-Bowdoin, T., Knapp, D., Erasmus, B., Mathieu, R., Wessels, K., and Smit, I. P. J. (2009b). LIDAR waveform-based woody and foliar biomass estimation in savanna environments. In *Silvilaser 2009 - 9th International Conference on Lidar Applications for Assessing Forest Ecosystems*, page 10pp., College Station, TX.
- [Zhao et al., 2009] Zhao, K., Popescu, S., and Nelson, R. (2009). LIDAR remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, 113(1):182–196.