CONNECTING THE DOTS BETWEEN LASER WAVEFORMS AND HERBACEOUS BIOMASS FOR ASSESSMENT OF LAND DEGRADATION USING SMALL-FOOTPRINT WAVEFORM LIDAR DATA


Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY
Carnegie Institution for Science, Stanford, CA
Council for Scientific and Industrial Research, Pretoria, South Africa
School of Animal, Plant and Environmental Science, University of the Witwatersrand, Johannesburg, South Africa
Kruger National Park Scientific Services, Skukuza, South Africa

ABSTRACT

Measurement and management of vegetation biomass accumulation in ecosystems typically involves extensive field data collection, which can be expensive and time consuming, while leaving the user with relatively crude inputs to intricate biomass models. Light detection and ranging (LiDAR) remote sensing, which provides extensive height measurements of terrain and vegetation, has become an effective alternative to characterize vegetation structure. In this paper, we report on ongoing efforts at developing signal processing approaches to model herbaceous biomass using a new generation of airborne laser scanners, namely full-waveform LiDAR systems. Structural and statistic-based feature metrics are directly derived from LiDAR waveforms at the pixel level and related to plot-level field data. Initial results reveal a definite correlation between the LiDAR waveform and herbaceous biomass. Ongoing research focuses on the links between fractional cover estimated from imaging spectroscopy and woody biomass.

Index Terms—Biomass, LiDAR, waveform, signal processing

1. INTRODUCTION

Information regarding global carbon sources (e.g., emissions) and sinks (e.g., carbon sequestration) is essential to our understanding of global energy flows and general carbon stock fluctuations. Such information also plays an important role in fine-scale dynamics, specifically those related to vegetation biomass and its link to land degradation, i.e., the loss of an ecosystem’s capability to provide services to communities. However, measurement and management of vegetation biomass (carbon) accumulation typically involves extensive field data collection, which includes parameters such as foliar area, crown volume, bare soil coverage, and vegetation height. Acquisition of these data can be expensive and time consuming. Traditional remote sensing technology, such as multi-spectral data (e.g., 1km² pixels in NOAA AVHRR data or 250m x 250m pixels MODIS data), has been applied to develop regional indicators of vegetation production. However, these spectrally-and-spatially coarse resolution data cannot unravel changes in the land surface at the scale at which fine scale plant physiological processes actually occur (a few meters). Nor can they identify vegetation composition and structure, especially in the vertical dimension. Light detection and ranging (LiDAR) remote sensing, which provides extensive height measurements of terrain and vegetation, has created novel opportunities for accurate characterization of vegetation structure. A LiDAR sensor typically emits a laser pulse and registers the return trip distance between the sensor and a reflective target, thereby enabling range measurements. A novel type of LiDAR sensor, called waveform LiDAR, capable of recording and digitizing the full-backscattered signal at high vertical resolution (~1ns), holds much promise for detailed vertical characterization of vegetation structure.

Full-waveform LiDAR data have been widely used for estimating forest parameters, e.g., canopy height, stem diameter, woody biomass, etc. [1-5]. However, these studies are constrained to tree characterization. In this paper, we explored the possibility of above-ground herbaceous biomass estimation via a signal-processing approach applied to small-footprint waveform LiDAR data. The entire waveform processing workflow consists of de-noising, signal deconvolution, Gaussian decomposition, statistical
feature extraction, and regression model development. The research goal is to eventually link woody-herbaceous biomass assessment and lidar-imaging spectroscopy approaches, even though this paper focuses mainly on herbaceous biomass modeling.

2. DATA

2.1. Study area

The study area is bounded by (22°8’00” S; 30°34’52”E) and (25°32’48”s; 32°2’50”E) in South Africa (Figure 1) and spans a conservation-subsistence farming land use gradient. This gradient is defined along a transect from the Bushbuckridge (communal range lands; high rural population density) to the Sabie Sands game reserve (private conservation area) and Kruger National Park (state-owned conservation area) areas.

Figure 1: Location of the study area that spans a land use gradient (west-to-east) from heavily exploited range- and farmland, a private game reserve, and the Kruger National Park, South Africa

2.2. Remote sensing and field data

Waveform LiDAR data (pixel size: 0.56x0.56 m; vertical resolution: 1ns) were acquired by Carnegie Airborne Observatory (CAO) during April 2008. Each scene pixel is represented by an incoming (received) waveform distribution with 256 bands at 1ns (0.15m) spacing. The associated waveform of the outgoing pulse was also available.

Field data for this research were collected from 36 sites in the study area, each 50 x 50 m in size. A total of 36 plots were laid out within each site at a 10 m spacing, resulting in a grid-like pattern (Figure 2). For each plot, the herbaceous biomass was weighed within a 0.5x0.5m grid, along with assessment of other variables, e.g., woody biomass, canopy density, etc.

3. HERBACEOUS BIOMASS MODELING

Figure 3 shows the workflow of the waveform LiDAR-based herbaceous biomass modeling procedure. It consists of two parts, namely signal preprocessing and modeling.

3.1. Signal preprocessing

The raw incoming (received) waveform typically exhibits a stretched and featureless character, attributed to a fixed time span allocated for detection, the sensor’s variable outgoing pulse signal, the receiver impulse response, and system noise. Signal preprocessing therefore is necessary to recover the true response distribution of optically active targets along the path of the LiDAR waveform.

First, system noise is typically present in the form of high frequency components of the raw signal in the frequency domain. Therefore, we smoothed the raw waveform by setting a cut-off frequency threshold for removal of noise components in the frequency domain (a similar effect as a low pass filter), followed by conversion back to the time domain. The subsequent noise-filtered waveform can be mathematically modeled as:

\[ P_i(t) = P_o(t) * \sigma(t) * \Gamma(t) \]  

where \( P_o(t) \) refers to the outgoing waveform (known), \( \sigma(t) \) represents the cross-section (the true response distribution of the target), and \( \Gamma(t) \) is the receiver impulse response (estimated by the return signal from a flat ground area). The true response of the target can be derived by sequentially...
deconvolving the incoming waveform from the outgoing waveform and receiver impulse response. We applied the Richardson-Lucy algorithm [6] for this purpose. Richardson-Lucy is an iterative deconvolution procedure, which is based on Bayes’ statistical theorem. The mathematical solution of \( \sigma(t) \) can be expressed as:

\[
\hat{\sigma}_{i+1}(t) = \hat{\sigma}_i(t) \left[ \frac{P(t)}{h(t) * \hat{\sigma}_i(t)} \right] * h(t) \quad (2)
\]

where \( h(t) = P(t) * \Gamma(t) \), and \( i \) denotes the iteration. The residual for each iteration is computed as:

\[
r_i(t) = P_i(t) - h(t) * \hat{\sigma}_i(t) \quad (3)
\]

The residual will converge as the iteration progresses. The user can terminate the iteration, either by selecting a specific residual threshold or by setting a constant iteration number. Harsdorf and Reuter [7] claimed that the one-dimensional Richardson-Lucy algorithm resulted in the most stable results when compared to Fourier transform and non-negative least squares approaches. This processing step enhances the vertical signal resolution, which facilitates extraction of target information from the waveform.

### 3.2. Modeling

![Figure 4](image-url)

**Figure 4:** Raw waveform (left) and Gaussian decomposition of the deconvolved waveform (right).

The last component of a return LiDAR waveform typically corresponds to the ground-level response, which may be composed of bare soil, grass, leaves, stones, etc. We hypothesized that the herbaceous biomass, directly associated with the grass abundance, can be linked to the properties of the last waveform component (e.g. width, height, area). Figure 4 (left) shows the raw return waveform (single peak) where there is no tree or shrub. Figure 4 (right) reveals a dual-peak intensity distribution after deconvolution of the raw waveform; this was hidden in the raw signal (left) due to the existence of an imperfect system response and variable outgoing “pulse”. An expectation-maximization (EM) algorithm was subsequently employed to decompose this deconvoluted waveform into two individual Gaussian curves [8]. It is evident from Figure 4 that the second Gaussian is mainly due to the asymmetric trailing edge, relative to the leading edge in the raw waveform. This asymmetric trailing edge typically results from the late return photons due to the structure of the ground layer (e.g., grass), leading to multiple scattering of the return signal. On the other hand, the first Gaussian was seen as corresponding mainly to the single scattering from the ground material (e.g., bare soil, grass, stone, etc). The mathematical description of this waveform as a mixed Gaussian model is expressed as:

\[
g(t) = a_1e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + a_2e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \quad (4)
\]

where \( a_1 \) and \( a_2 \) are the amplitudes of the Gaussian peaks and \( \sigma_1 \) and \( \sigma_2 \) are the standard deviation (related to width) of each Gaussian (\( \mu \) and \( \mu_1 \) are input and mean variables, respectively). The next step involved extraction of waveform metrics (independent variables) and linking these to the field biomass data. Since we have parameterized the waveform in terms of a Gaussian distribution, feature metrics can be directly extracted from Eq. 4 (e.g., \( a_1, a_2, \sigma_1, \) and \( \sigma_2 \)). We also added two additional metrics, namely \( s_1 \) and \( s_2 \), which correspond to the integration (area) of the two Gaussian curves. These six independent metrics are not necessarily uncorrelated, which led to the exclusion of highly-correlated (> 0.8) metrics after calculation of correlation coefficients. The herbaceous biomass model was finally retrieved based on a linear regression fit between the selected, independent feature metrics and field data in the form of:

\[
H_{\text{biomass}} = \sum_{n=1}^{n} c_n p_n + k \quad (5)
\]

where \( p_n \) refers to the \( n \)th feature metric, \( c_n \) represents the associated coefficient, and \( k \) is the regression intercept.

### 4. EXPERIMENTAL RESULTS

The proposed model was tested for 6 different sites, the only ones that contained waveform lidar data. Herbaceous biomass in these sites ranged between 0–90 gram/plot (216 plots in total). We only considered waveforms (before deconvolution) with a single peak, i.e., waveforms that did not exhibit multiple peaks due to tree canopy returns. This reduced the number of sample plots to 159. We also assumed that the GPS locations of the pixel-based (0.56x0.56m) waveform and the plot center (field sample) were both representative of the same plot. Herbaceous biomass samples were then grouped into 5g classes for the purposes of this study, which led to 18 weight-based biomass classes (e.g. 0–5, 5–10, ...85–90) in the 0-90g range. Waveform-derived metrics and measured biomass were averaged within each class.

Table 1 shows the correlation coefficient matrix for the field data and waveform-derived metrics, used to optimize the variable selection. All the metrics in Table 1 have been converted into “natural log” space to minimize the nonlinearity between the parameters. It is evident that pairs \( (a_1, s_1) \) and \( (a_2, s_2) \) exhibited high correlations. We therefore discarded \( a_1 \) and \( s_2 \) to ensure model robustness, since these correlated metrics also exhibited a lower correlation to the biomass, when compared with \( s_1 \) and \( a_2 \), respectively.
Figure 5 shows the results of herbaceous biomass estimation using feature metrics $\sigma_1$, $s_1$, $a_2$, and $\sigma_2$ (Eq. 6), where the coefficients were solved by least squares linear regression. We have concluded that the waveform approach has potential for estimating above-ground herbaceous biomass, given the model’s ability to explain almost 60% in herbaceous biomass variability. However, we feel that the small range in herbaceous biomass field values, limited structural information, and senescent state of the vegetation were detrimental to model performance.

$$\ln(H) = 6.3 \ln(a_1) + 5.2 \ln(s_1) + 0.3 \ln(a_2) + 0.4 \ln(\sigma_2) - 41.6 \quad (6)$$

### Table 1: Correlation coefficients between field data and waveform-derived feature metrics

<table>
<thead>
<tr>
<th></th>
<th>$a_1$</th>
<th>$\sigma_1$</th>
<th>$s_1$</th>
<th>$a_2$</th>
<th>$\sigma_2$</th>
<th>$s_2$</th>
<th>Bio (H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td></td>
<td>-0.12</td>
<td>0.98</td>
<td>0.79</td>
<td>0.35</td>
<td>0.59</td>
<td>0.69</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>-0.12</td>
<td>1</td>
<td>0.07</td>
<td>0</td>
<td>-0.05</td>
<td>-0.09</td>
<td>0.21</td>
</tr>
<tr>
<td>$s_1$</td>
<td>0.98</td>
<td>0.07</td>
<td>1</td>
<td>0.80</td>
<td>0.36</td>
<td>0.58</td>
<td>0.75</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.79</td>
<td>0</td>
<td>0.80</td>
<td>1</td>
<td>0.52</td>
<td>0.93</td>
<td>0.67</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.35 -0.05</td>
<td>0.36</td>
<td>0.52</td>
<td>1</td>
<td>0.57</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>$s_2$</td>
<td>0.59 -0.09</td>
<td>0.58</td>
<td>0.93</td>
<td>0.57</td>
<td>1</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Bio (H)</td>
<td>0.69</td>
<td>0.21</td>
<td>0.75</td>
<td>0.67</td>
<td>0.35</td>
<td>0.50</td>
<td>1</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

We successfully extracted waveform LiDAR feature metrics from the deconvolved waveform’s Gaussian responses to model plot-level herbaceous biomass - the coefficient of determination ($R^2$) indicated that our model could explain 60% of the variation in herbaceous biomass. Although this could be considered as relatively low, it is clear that significant potential exists for assessment of herbaceous biomass in savanna ecosystems at fine scales using waveform LiDAR. We mainly attributed the relatively poor model performance to a narrow range of field biomass values. Future research will focus on biomass estimation during the wet season, linking woody-herbaceous biomass assessment, and applying spectral-based mixture mapping to further explore the relative variation of LiDAR returns across different vegetation species, structures, biomass, etc., at the sub-pixel level.

6. ACKNOWLEDGEMENTS

The Carnegie Airborne Observatory is supported by the W.M. Keck Foundation and William Hearst III; the study was funded by the Andrew Mellon Foundation. We are grateful for field campaign funding from the Council for Scientific and Industrial Research (SA) and PhD student funding from the Rochester Institute of Technology (USA).

7. REFERENCES


