Use of Airborne LiDAR To Estimate Forest Stand Characteristics

Qi Li1, Wei Zhou1,*, and Chang Li2
1Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China; 2 College of Urban and Environmental Science, HuaZhong Normal University, 152 Luoyu Road, Wuhan 430079, China.

E-mail: qli@ceode.ac.cn

Abstract. Small-Footprint Airborne LiDAR(light detection and ranging) remote sensing is a breakthrough technology for deriving forest canopy structural characteristics. Because the technique is relatively new as applied to canopy measurement in China, there is a tremendous need for experiments that integrate field work, LiDAR remote sensing and subsequent analyses for retrieving the full complement of structural measures critical for forestry applications. Data storage capacity and high processing speed available today have made it possible to digitally sample and store the entire reflected waveform, instead of only extracting the discrete coordinates which form the so-called point clouds. Return waveforms can give more detailed insights into the vertical structure of surface objects, surface slope, roughness and reflectivity than the conventional echoes. In this paper, an improved Expectation Maximum (EM) algorithm is adopted to decompose raw waveform data. Derived forest biophysical parameters, such as vegetation height, subcanopy topography, crown volume, ground reflectivity, vegetation reflectivity and canopy closure, are able to describe the horizontal and vertical forest canopy structure.

1. Introduction
Development of airborne laser scanning goes back to the 1970s and 1980s. The first commercially available airborne laser scanners recorded the time of one backscattered pulse. The recording of only one pulse is sufficient if there is only one target within the laser footprint. Some commercial laser scanners typically measure first and last pulse; the latest LiDAR are able to measure up to five pulses. In order to derive digital terrain models (DTMs), laser pulses reflected by the ground surface must be distinguished from non-terrain points. This task can be achieved using various filtering techniques that classify the point cloud into terrain and off-terrain points just based on the spatial relationship of the 3D data. For many applications, this methodology has been considered an appropriate way. However, the user neither know how the electronics of LiDAR system actually determine the location of the returns, nor understand any distortions of the pulse shape that receiver electronics or surface structures may have influenced the pulse echo. LiDAR system manufacturers keep secret the pulse detection methods that their systems use. However, The choice of pulse detection methods has significant

* Corresponding author Tel.: +86 10 82178128.
E-mail addresses: qli@ceode.ac.cn, wzhou@ceode.ac.cn.
impact on accuracy in practice\cite{1}. Besides, if the output is range measurements merely, much of the information about structured surfaces is lost.

The solution is to digitally sample and store the entire echo waveform of reflected laser pulses. Digitizing and recording the complete backscattered waveform during the acquisition for later post-processing has the advantages that algorithms can be adjusted to tasks, intermediate results are respected, and neighborhood relations of pulses can be considered. The technical feasibility has been demonstrated by large-footprint airborne systems developed by NASA in the 1990s, namely the Scanning LiDAR Imager of Canopies by Echo Recover (SLICER) and the Laser Vegetation Imaging Sensor (LVIS). Recently, three commercial airborne systems have become available, namely RIEGL LMS-Q560, Toposys FalconIII, Leica ALS and Optech ALTM (Figure 1). An approach based on unsupervised learning was proposed where a mixture of Gaussian distributions are fitted to the waveforms to detect and parameterize the peaks. This approach uses the Expectation Maximization (EM) algorithm. A detection algorithm based on the expectation-maximization (EM) algorithm is used to estimate the number of echo pulses of the waveforms.

![Waveform Digitization](image1)

**Figure 1.** Waveform Digitization\cite{2}

### 2. Algorithm of decomposing waveform

As a first approximation we will assume that the laser output pulse shape or impulse response (i.e., the shape of the outgoing laser pulse after passing through the full detector and digitizer chain) is Gaussian. We further assume that the returning laser pulse is composed of a series of potentially-overlapping reflections similar in shape to the impulse response (i.e., in this case by a series of Gaussian-shaped reflections). As was shown in the theory part of the paper, the implicit assumption of Gaussian decomposition is that the cross-section profile can be represented by a series of Gaussian functions\cite{3}.

It is of interest to extract more than the first and the last echoes from each waveform. Also, the width of the echoes is of interest. A detection algorithm based on the improved expectation-maximization (EM) algorithm is used to estimate the number of echo pulses of the waveforms. The algorithm also outputs the width of the echo pulses. Unsupervised learning is a method of machine learning where a model is fitted to observations. An important part of the unsupervised learning problem is determining the number of components or classes which best describe the data\cite{4}.

Unsupervised learning will be used in this paper to detect echo pulses. This is done by fitting Gaussians to the waveforms. It will be assumed that the waveforms were generated from a distribution which is the sum of simpler distributions. That is, the samples of the waveforms are assumed to arise from the following distribution.

\[
f(x) = \sum_{j=1}^{k} p_j \times f_j(x)
\]

\[
f_j(x) \in N(\mu_j, \sigma_j^2)
\]

Where \( k \) is the number of Gaussians, \( f_j(x) \) is the Gaussian probability density function, \( p_j \) is the relative weight of \( f_j(x) \), \( \mu_j \) is the expected value and \( \sigma_j \) is the standard deviation of the fitted Gaussians.
The EM algorithm, which will be used to fit the Gaussians to the waveforms, is a widely used approach in learning the presence of unobserved variables. Original formula of EM algorithm:

\[ Q_j = \frac{p_j f_j(x_i)}{\sum_{j=1}^{k} p_j f_j(x_i)} \]  

(1)

\[ p_j = \frac{\sum_{i=1}^{n} Q_{ij}}{n} \]  

(2)

\[ \mu_j = \frac{\sum_{i=1}^{n} Q_{ij} i}{p_j \times n} \]  

(3)

\[ \sigma_j = \sqrt{\frac{\sum_{i=1}^{n} Q_{ij} (i - \mu_j)^2}{p_j \times n}} \]  

(4)

\( Q_{ij} \) is the probability that sample \( i \) belongs to component \( j \) and \( k \) is the number of components that are fitted to the waveforms. For each component \( j \), the mean value \( \mu_j \) and standard deviation \( \sigma_j \) are estimated. The mean value \( \mu_j \) will be used as the position of the echo and the standard deviation \( \sigma_j \) as the width of the echo. The likelihood estimates for \( \mu_j \) and \( \sigma_j \) are found by iterating through formula (1)-(4). The algorithm needs to be initialized with start values.

3. Experiments and discussion

3.1. Pre-processing

The original waveforms (Figure 2) have to be segmented using a threshold to remove noises before the EM algorithm is applied. The threshold is derived on a per shot basis by establishing the mean background noise occurring in the waveform record, the last 5% of the waveform record is used to calculate threshold \( \sigma_{\text{noise}} \). All samples of the waveform below the threshold \( \sigma_{\text{noise}} \) are then set to zero. Figure 3 is the pre-processed waveform.

3.2. Initial values

The correctness of the initial values of \( \mu_j \), \( p_j \) and \( \sigma_j \) is important for the EM algorithm to generate a good estimate. The initial value for \( \mu_j \), is found by smoothing the pre-processed signal and First derivative of waveform.
Waveform (solid) and Three gauss components (dashed) have been estimated from the waveform in Figure 4. The estimated parameter: $\mu_1=54$, $\sigma_1=9.0$, $\mu_2=89$, $\sigma_2=8.5$, $\mu_3=131$, $\sigma_3=7.2$.

Automated last return detection software is applied to the backscatter waveform to identify the start, peak, and end of the last return, inferred to be from the ground. In Figure 4 the laser pulse hits the canopy first and creates one echo pulse. A fraction of the laser pulse also hits the ground giving rise to a twice echo pulse. The vertical lines in the figure illustrate the positions of the ground extracted by the SLICER.

Figure 6. Waveform (solid) and two fitted Gauss components (dashed), $\mu_1=57$, $\sigma_1=7.3$, $\mu_2=83$, $\sigma_2=7.3$. Vertical lines show the positions of the echoes extracted by the Gaussian decomposition. The EM algorithm performs well in decomposing overlapping echo pulses.

In Figure 7, algorithm of SLICER has detected only one peak that is ground, but three gauss components have been estimated from the waveform by Gaussian decomposition in Figure 8.

3.3. Waveform for forestry application

LiDAR remote sensing has vast potential for the direct measurement and estimation of several key forest characteristics. The direct measurements of small-footprint LiDAR are canopy height, subcanopy topography, and the vertical distribution of intercepted surfaces between the canopy top and the ground. Other forest structural characteristics, such as aboveground biomass, are modeled or inferred from these direct measurements\(^{5,7}\).
3.3.1. Treeheight
With small-footprint systems, the first return above a noise threshold can be used to estimate the top of the canopy, and the midpoint of the last return represents the ground return.

![Figure 9. Treeheight](image1)

![Figure 10. Map of treeheight](image2)

Figure 11. Map of treeheight

3.3.2. Crown volume

![Figure 12. Sketch map of Crown volume](image3)

![Figure 13. Local Crown volume](image4)
4. Conclusion
It appears that full-wave systems will greatly enhance our capability to map natural and artificial objects, but this comes at a cost: Instead of having one or a few trigger pulses the whole discrete signal must be stored. Major research and development efforts will be needed in order to develop algorithms and software that can efficiently transform the recorded waveform clouds into geo-spatial data sets. The EM algorithm has been created to decompose a laser altimeter return waveform from simple and complex surfaces into a series of Gaussians. By analysis of the backscattered signal, it should also be possible to determine quality parameters for a given range measurement, which can be used as a direct input into further processing steps.

Acknowledgments
This study has been supported by funding from the National Natural Science Foundation, China (Grant No. 41001204) and the State Administration of the Cultural Heritage, China (Grant No. 20100306). We also thank Prof. Pinliang Dong from the Department of Geography University of North Texas for his kind help to revise the manuscript of this paper.

5. References
[1] Wagner W, Ullrich A, Melzer T, Briese C and Kraus K 2004 A From single-pulse to full waveform airborne laser scanners: potential and practical challenges ISPRS XXth Congress Istanbul, Vol XXXV Part B/3 pp 201-206
[2] Wagner W, Ullrich A, Ducic V, Melzer and tudnicka 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
[3] Hofton M A, Blair J B and Minster J 2000 Decomposition of Laser Altimeter Waveforms IEEE Transactions on Geoscience and Remote Sensing 38 (4) 1989-1996
[4] Oliver J J, Baxter R A and Wallace C S 1996 Unsupervised Learning Using MML Machine Learning Proceedings of the Thirteenth International Conference (Bari, Italy, 3-6 July 1996) ed Filippo Neri et al pp 364-372
[5] Hyyppä J 2004 Algorithms and methods of airborne laser scanning for forest measurement International Archives of Photogrammetry Remote Sensing and Spatial Information Science vol36 pp82-88
[6] Ralph O D and Jason B D 2000 LiDAR Remote Sensing for Forestry Applications Journal of Forestry 98 44-46
[7] Mean J E, Acker S A, Harding D J, Blair J B, Lefsky M A, Cohen W B, Harmon M E and McKee W A 1999 Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the Western Cascades of Oregon Remote Sensing of Environment vol67 pp298-308