Remote Sens. **2012**, *4*, 377-403; doi:10.3390/rs4020377

OPEN ACCESS

Remote Sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Tree Species Detection Accuracies Using Discrete Point Lidar and Airborne Waveform Lidar

Nicholas R. Vaughn, L. Monika Moskal * and Eric C. Turnblom

School of Environmental and Forest Sciences, University of Washington, Box 352100, Seattle, WA 98195, USA; E-Mails: nrv2@uw.edu (N.R.V.); ect@uw.edu (E.C.T.)

* Author to whom correspondence should be addressed; E-Mail: lmmoskal@uw.edu; Tel.:+1-206-221-6391; Fax:+1-206-685-0790.

Received: 2 December 2011; in revised form: 30 January 2012 / Accepted: 30 January 2012 / Published: 2 February 2012

Abstract: Species information is a key component of any forest inventory. However, when performing forest inventory from aerial scanning Lidar data, species classification can be very difficult. We investigated changes in classification accuracy while identifying five individual tree species (Douglas-fir, western redcedar, bigleaf maple, red alder, and black cottonwood) in the Pacific Northwest United States using two data sets: discrete point Lidar data alone and discrete point data in combination with waveform Lidar data. Waveform information included variables which summarize the frequency domain representation of all waveforms crossing individual trees. Discrete point data alone provided 79.2 percent overall accuracy (kappa = 0.74) for all 5 species and up to 97.8 percent (kappa = 0.96) when comparing individual pairs of these 5 species. Incorporating waveform information improved the overall accuracy to 85.4 percent (kappa = 0.817) for five species, and in several two-species comparisons. Improvements were most notable in comparing the two conifer species and in comparing two of the three hardwood species.

Keywords: Support Vector Machine; fullwave lidar; discrete Fourier transform; forest inventory

1. Introduction

Information about individual tree species can be extremely beneficial when estimating many forest resource values, such as timber value, habitat quality, or susceptibility to loss. Unfortunately, detection of individual tree species using remote sensing (RS) data has proven to be a difficult task to accomplish. The species of a tree is only one of several factors that affect the realized shape and color of an individual tree crown. Other factors such as terrain, environment, competition, and genetic variation have large influences as well. As a result, for most variables that one can measure from RS data, there is significant distributional overlap between species, for instance (Figure 4 in [1]) and (Figure 3 in [2]). Due to the difficulty of obtaining sufficient classification accuracy, species information is commonly disregarded or alternative methods are found [3]. One such alternative is to impute species information from the ground data observations best matching each identified crown segment [4,5]. This technique is a step forward, but further improvement should be possible.

Knowledge of the probable species of individual crown regions identified in the data would enable us to stratify model estimates by species. This will most likely increase precision in any stand-level estimates of interest. With this goal in mind, researchers have continuously tested new forms of RS data seeking improvements in detecting stand- and individual tree-level species information. As RS technology and computer algorithms have improved, so have the classification results achieved.

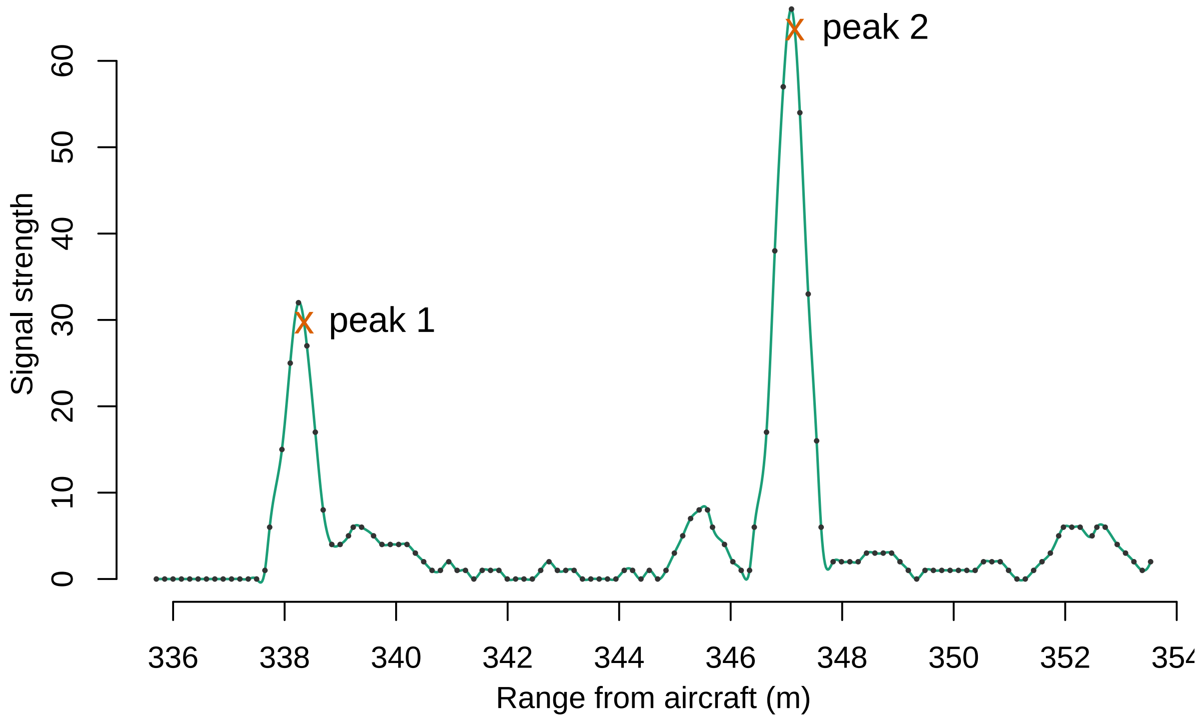
With the advent of scanning Lidar, many aspects of a forest inventory can now be accomplished using the aerial form of this data [6,7]. Even more recently, the abilities of terrestrial Lidar scanners for forest inventory are also being investigated [8,9]. While multi-spectral [10] and hyperspatial [11] imagery products have traditionally enabled the identification of individual tree species, the cost of additional datasets is likely to challenge budget constraints. Ideally, the use of a Lidar dataset alone would be sufficient to achieve the necessary species identification accuracy.

Many authors have investigated the potential of Lidar data for species classification. Sometimes Lidar is mixed with additional data sources, such as color or multi-spectral imagery [12–17]. Because each individual situation presents a unique arrangement of challenges, the overall results have been mixed. While the number of variables one can compute from discrete point Lidar data is infinite, there are only a few concepts that can be represented by these variables. These concepts are: crown density, crown shape, crown surface texture, and received energy from individual peaks. Most authors have incorporated variables from more than one of these concepts.

Crown density describes the leaf and branch size and arrangement and is typically measured using the proportions of the returns hitting vegetation versus those hitting the ground [2,18]. Crown shape information is often compared using parameters of surface models fit over the top of the Lidar point cloud [19–21]. The distribution of return heights, often described using select percentiles of the return heights [1,22], includes information about both crown density and crown shape. Crown surface texture refers to the roughness of a tree crown, and has been measured using a canopy height model [19]. The instantaneous light energy received by the sensor when each peak is detected is typically referred to as intensity. The measured intensity is affected by several physical traits such as leaf size, chemistry, and incident angle, which are all affected by species. While most authors incorporate this intensity information, both Ørka *et al.* [23] and Kim *et al.* [24] found that intensity alone could be a reasonable predictor of species.

In the last half-decade, a newer format of Lidar information, commonly referred to as "waveform" or "fullwave" Lidar, has slowly increased in availability. In contrast to the more common discrete point Lidar systems, this newer Lidar system takes advantage of increased processor speeds and data storage capacity by digitally sampling at a high rate the return signal received at the sensor. The result mimics the appearance of a wave, and an example of such a waveform can be seen in Figure 1. For comparison, if the waveform shown in Figure 1 were to be passed through an onboard peak detector, the result might resemble the two exes immediately following the peak crests.

Figure 1. An example waveform and probable associated discrete return points. The 120 waveform samples are shown as circles and a spline fit to these data appears as a solid line. A peak detector might detect two peaks at about 338 and 347 meters and return the intensity value when the peak is detected as shown with exes. Without knowledge of future sample values, real time peak detection algorithms usually produce a slight lag in peak location.



While a few authors have looked to waveform data for improving classification accuracy the first step has always been to decompose the waveforms into discrete peaks. One advantage of this technique is that information about peak shape can be preserved. The shapes of these peaks have successfully been used in distinguishing vegetation from other surfaces [25] and in distinguishing the roughness of surfaces [26]. Additionally, pulse width or cross section information has been helpful in classifying deciduous from coniferous species [20,27]. Little work has been done, however, to see if patterns within the original waveform data, prior to peak decomposition, provide information for species classification.

In a previous paper [28], we showed that Fourier transformation of the waveforms crossing each crown led to moderate accuracy while classifying three hardwood species. In this work, no two-

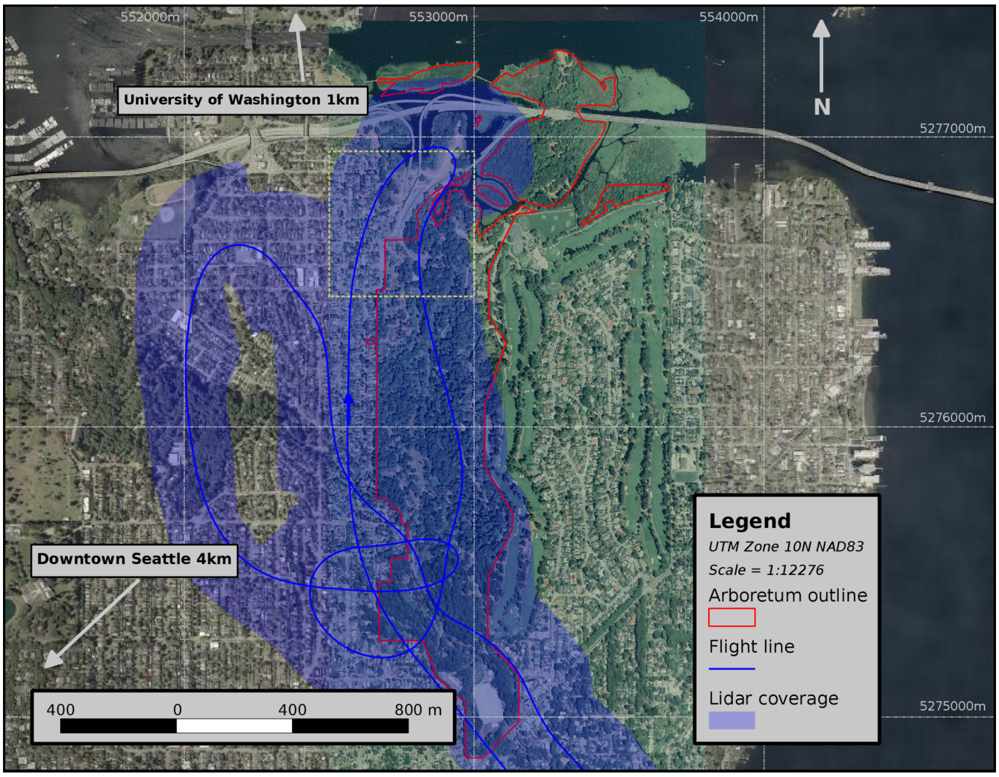
or three-dimensional information about crown structure computed from a discrete point array was included. However, the amplitude of a rather high frequency component of the Fourier transforms played an important role in distinguishing two of the species. This frequency was high enough that even high-density discrete point Lidar data could not contain such information. The purpose of this paper is to validate the results of the original paper as well as further test the importance of waveform Lidar in determining tree species. The latter is done by testing classification performance both before and after the addition of waveform information to a full suite of crown density, crown shape, crown surface texture and intensity metrics.

2. Methods

2.1. Waveform Lidar Data Acquisition

Waveform data were obtained by Terrapoint, USA during the evening of 7 August 2008 over the University of Washington Arboretum in the city of Seattle, Washington. Sensor altitude above canopy surface ranged from 145 to 412 m with mean distance of 310 m. Scan angle varied from −30 to 30 degrees from zenith. Pulse frequency was 133 kHz. The majority of the arboretum was covered in one loop in the North-South direction. An area map of the Arboretum with the flight line plotted is given in Figure 2. As this was a sample flight with little planning prior to flight, significant gaps exist between segments of the flight line. As a result many of the trees with field data are in the margins of the swath area. Overall data density averaged about 10 pulses per square meter near nadir at ground level.

Figure 2. A map of the University of Washington Arboretum in the city of Seattle. The Arboretum boundary is shown as a red line. The helicopter flight path is plotted as a blue line and the associated Lidar coverage area is blue-tinted.



2.2. Crown Segmentation and Field Data Collection

Our field data were collected in a slightly different manner as one would collect information if an inventory was required. We segmented tree crowns from the Lidar data prior to visiting the field so that we could verify that each tree matches its associated data segment. Several trees of five species were collected in this manner to ensure that our data were as clean as possible. This segmentation and field data collection took part in three steps:

1. All waveforms were deconvolved and then decomposed into individual peaks using a simple peak-detection algorithm. This point information was indexed into a voxel array structure.
2. A segmentation algorithm was used on the voxel array data to map the volume of space occupied by individual tree crowns into clusters of voxels.
3. Outlines of these clusters were used to locate the trees on the ground and identify the species.

2.2.1. Waveform Deconvolution

The light pulse emitted from the laser is not instantaneous, but rather it increases to a peak and decreases again over a given period of time. The energy reflected by an object that is received by the sensor is then spread over the same time period. Due to this spread, adjacent targets falling within the path of the laser beam are more likely to appear as one peak in the waveform. A point-spread function $p(t)$ can be used to represent the portion of this pulse energy received at the sensor over time. The recorded waveform, $\phi(t)$, then represents not the direct reflectivity function, $\psi(t)$, of the targets in the path of the pulse, but rather the convolution of this reflectivity function with the point spread function. The Richardson–Lucy algorithm [29,30], commonly used to de-blur images, is one method to deconvolve these functions using an assumed $p(t)$.

If the known point spread function is an odd n_p time units long, then the single waveform sample, ϕ_i, observed at time i within a recorded waveform can be represented under this system as:

$$\phi_i = \sum_{j=1}^{n_p} p_j \psi_{(i-h+j)} \tag{1}$$

where $h = (n_p + 1)/2$. The algorithm incorporates iteration to estimate the direct reflectivity function $\psi(t)$ from the recorded signal, where at iteration 0, $\psi_i^{(0)} = \phi_i$, and at each following iteration l

$$\psi_i^{(l+1)} = \psi_i^{(l)} \sum_{j=1}^{n_p} p_j \frac{\psi_{(i-h+j)}}{c_j} \tag{2}$$

where

$$c_j = \sum_{k=1}^{n_p} p_k \psi_{(j-h+k)}^{(l)} \tag{3}$$

and iteration is continued until a satisfactory convergence criteria is met. We used the absolute value of the mean difference between two iterations

$$M^{(l)} = \left| \sum_{i=1}^{N} \left(\psi_i^{(l+1)} - \psi_i^{(l)} \right) /N \right| \tag{4}$$

as a measure of distance and set the convergence criteria to be such that $M^{(l)}$ must be less than 0.01.

We used a point spread function that spans 9 time steps (about 9 nanoseconds in duration), derived from a binomial distribution, based on the similarity between this distribution and the Gaussian shape typically attributed to Lidar pulses. Given that x has a binomial(n, p) distribution

$$p_i = P(x = i - 1 | n = 8, p = 0.5) \ \ \forall \ \ i \in 1 \dots 9 \tag{5}$$

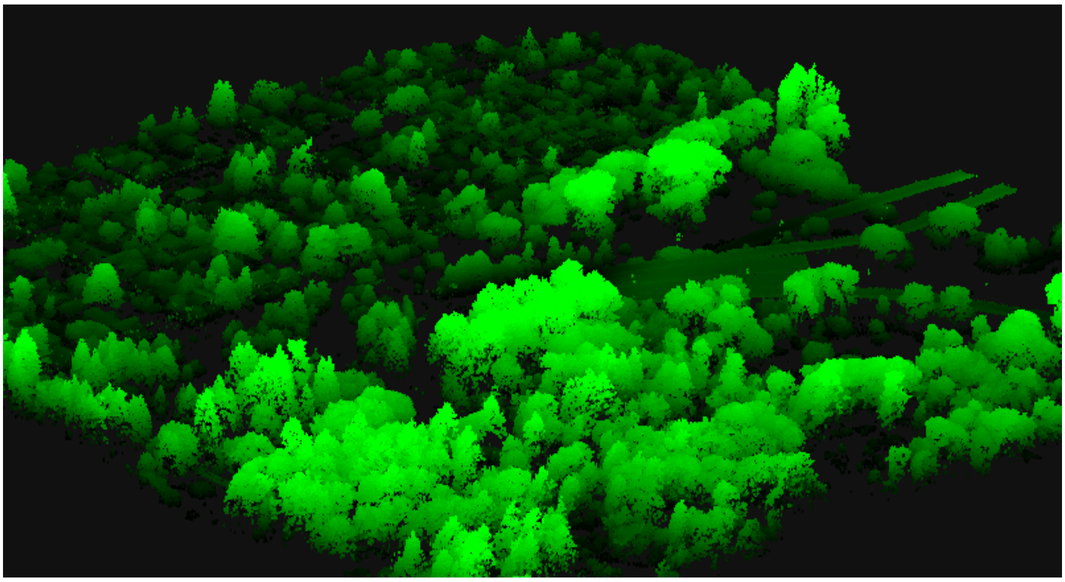
2.2.2. Peak Detection and Creation of the Voxel Array

A simple peak finding algorithm was performed on the waveform data after deconvolution with the Richardson–Lucy algorithm. The range at maximum for each peak found within the waveforms was used to compute an x, y, z position. Two additional pieces of information were kept for each peak. First was the total energy of the peak, or the sum of the intensity values for all waveform samples occurring during the defined peak. Second, we recorded the total range duration of the peak. The peak-finding algorithm had the following steps:

```
1. Decide on $M_{min}$, the minimum size of a peak that will be recorded
2. Set $D_0 = 0$
3. $d = \psi_2 - \psi_1$. If $d^2 > 0.01$ then $D_1 = d$, otherwise $D_1 = 0$.
4. Set $M$ = 0, $P$ = FALSE
5. For $i$ from 2 to 60 do:
    (a) $d = \psi_i - \psi_{i-1}$. If $d^2 > 0.01$ then $D_i = d$, otherwise $D_i = 0$.
    (b) If $\psi_i > 0.1$ and not ($D_i > 0.1$ and $D_{i-1} < 0.1$), then
            • $P$ = TRUE
            • $I$ = $i$
            • $M$ = $M + \psi_i$
        else if $P$ = TRUE and $M > M_{min}$
            • Record $I$, $M$, and $P$ as a new peak
        else
            • $P$ = FALSE
            • $M$ = 0
            • $I$ = 0
```

This process resulted in a fairly heavy discrete point dataset, with an average of 104 points per square meter within tree vegetation. An visual example of this data is presented in Figure 3. We used a three-dimensional (voxel) grid overlaid on the volume of interest with horizontal dimensions set at one meter each and vertical dimension set at 0.75 m. A single file was created to contain the grid location of all peaks occurring within this grid. The file header also stored information about which voxels, referenced by layer, row and column, contain points. The voxels with one or more points were used to segment out the individual tree crowns as well as to compute statistics for species classification.

Figure 3. Three-dimensional plot of discrete point Lidar data extracted from waveform Lidar data. This view is looking northwest at the area outlined with a tan colored dotted line in Figure 2. This image shows the variable density of the trees in the University of Washington Arboretum. A thicker patch of black cottonwood and red alder is visible in front of the elevated freeway ramp. Most of the crown of larger conifers are visible in other parts of the image. Outside the Arboretum, many neighborhood streets and power lines are visible.



2.2.3. Crown Segmentation

In order to obtain three-dimensional crown information about each tree crown, we created a voxel-based segmentation algorithm. Under this three-dimensional region growing algorithm, individual layers of the voxel array are read one at a time starting with the topmost layer. Individual voxels from each layer are added to new or existing voxel-clusters depending on their distance from these existing clusters. The ability of a cluster to incorporate a new voxel depends on current number of member voxels, vertical center of mass of these member voxels, as well as distance from the new voxel. Table 1 lists the default values for parameters that control search window size, cluster competition for new voxels, and cluster size limits. The algorithm, and the meaning of these parameters are given below.

Table 1. Default values for user-defined parameters of the crown segmentation algorithm.

Parameter	Default Value
rad_{min}	2.00
FL_{max}	1.50
FL_{min}	1.00
A	4.6
B_L	3.00
B_O	7.00
W	3.00
N	8.00
$mass_{min}$	1.00

1. Create an empty list of clusters.

2. Create an array, $\vec{C}$, with one element for each row and column combination of the voxel structure. This will represent the current owner cluster of a given vertical column in the voxel structure. Element (r, c) in this array will be set to NULL whenever a voxel (l, r, c) in the current layer l contains no points. Thus, a vertical voxel column above a given row and column combination can change ownership if a vertical gap exists in the voxel structure.

3. Create a second array, $\vec{P}$, which will function in a similar manner as $\vec{C}$. The difference will be that, once ownership of cell (r, c) is established, cell (r, c) will never be set to NULL again. This way, a search for neighboring clusters can be conducted for each (l, r, c) voxel investigated.

4. If any layers of interest remain, read in the next layer, l, of voxel cells from the voxel file. For each voxel cell of interest in layer l do:

   (a) Check if cell (l, r, c) contains points. If not, set element (r, c) in array $\vec{C}$ to NULL and proceed to next voxel.

   (b) Check grid $\vec{C}$ to determine if the current cell (r,c) is already owned. If yes, add the voxel (l, r, c) as a member to the indicated owner cluster and proceed to the next voxel. If no, do the following:

      i. Check for neighboring clusters on grid $\vec{P}$ within a the square region described by the intersection of rows $r - N$ to $r + N$ and columns $c - N$ to $c + N$. For each neighboring cluster found do the following:

      - Compute the cluster's radius as

$$rad_i = max\left(\frac{\sqrt{r_x r_y n_i}}{\sqrt{\pi l_i}}, rad_{min}\right) FL_i \tag{6}$$

which is the minimum of: (1) the radius of a circle of equivalent area as the average number of cells per layer in the given cluster; and (2) a constant rad_{min}, multiplied by a correction factor, FL_i. r_x and r_y are the voxel dimensions in the x and y direction, n_i is the number of member voxels in cluster i, and l_i is the current number of layers in cluster i. The value of F_L is determined by a function of the vertical length of the cluster:

$$FL_i = [(FL_{max} - FL_{min}) \times modlog\,(\delta z, A, B_L) + FL_{min}] \tag{7}$$

Here, FL_{max}, FL_{min}, A, and B_L are user-defined constants specified ahead of time, and δz is the vertical length in map units of the cluster i. This combination makes the search radius for taller trees greater, while rad_{min} makes sure newer clusters with few voxel members can still incorporate more voxels. The function $modlog\,(x, A, B)$ is the two parameter logistic function, re-parametrized so that A defines the change in x necessary for the output to change from 0.01 to 0.99

(assuming a positive slope parameter in the original logistic function), and B defines the value of x for which an output of 0.5 occurs.

$$modlog\,(x, A, B) = \left[1 + exp\left(\frac{AW}{A}(x - B)\right)\right]^{-1} \tag{8}$$

Here, AW is a constant value that is computed as

$$AW = ln(0.99/0.01) - ln(0.01/0.99) = 2ln(99) \approx 9.19 \tag{9}$$

- Using the radius derived with Equation (6) for each neighboring cluster, compute the "mass", M_i, of crown over the given cell using the formula:

$$M_i = O_i * FV_i * FH_i \tag{10}$$

where O_i is the two-dimensional area of overlap between: (1) a circle of radius rad_i centered around the cluster centroid; and (2) a circle around the cell center with radius W, which is a user-defined constant. FH_i is a correction factor for horizontal distance $dist_H$ between the cell center and cluster centroid.

$$FH_i = modlog\,(dist_H, A, B_O) \tag{11}$$

and FV_i is computed as a sum of individual corrections for each layer in the cluster.

$$FV_i = \sum_{j=l_{min}}^{l_{max}} modlog\,(j - l, A, B_O) \tag{12}$$

ii. If any of the M_i for the neighboring clusters exceed the minimum set by $mass_{min}$, then select the neighboring cluster with the largest mass and add the voxel to this cluster. If no neighboring clusters meet the minimum mass requirement, begin a new cluster with the voxel (l, r, c) as the first member. Set element (r, c) in both $\vec{C}$ and $\vec{P}$ to the label of the new cluster.

5. Go through the preliminary clusters and merge a cluster into one of its neighboring clusters if it and the neighbor meet the following conditions:

(a) The cluster is not too big (rad_i for the given cluster is less than 7 m).

(b) The distance from cluster top to cluster centroid is less than 5 m, cluster is not nearly vertical (unit direction vector from top to centroid has z-component less than 0.94), or cluster bottom is below 2 m.

(c) The cluster is smaller than the neighbor (has more total voxels than the neighboring cluster).

(d) The cluster and the neighbor have a large enough interface (at least 60 percent of the horizontal rectangle envelope containing the cluster is shared with the envelope of the neighbor).

(e) If the cluster has more than 3 layers and is larger than 50 voxels, the direction vector passing from cluster top through cluster centroid crosses through at least 4 voxels of neighbor (voxels must be within 2 m of vector).

6. Go through the remaining clusters and delete any cluster that meets any of the following conditions:
   (a) The cluster is too small (contains less than 50 voxels).
   (b) The cluster is too flat (ratio of cluster height to the average of cluster width in rows and cluster width in columns is less than 0.8).
   (c) The cluster is too short (cluster top is less than 5 meters high).

We investigated the feasibility of the alternative voxel-based segmentation algorithm proposed in [31] as well as those investigated in [32]. Our decision to use the algorithms listed above is not based on their merit, but more simply on the speed and simplicity of implementation and refinement. Additionally, it was simpler to perfect an algorithm in one office than with multiple correspondences with another party. This algorithm was just a means to achieve one step towards the final goal, and it was not the focus of the study. Accordingly, only a visual evaluation of this algorithm was performed. All parameter settings were optimized for the given data using trial and error.

2.2.4. Collection of Tree Species

Using the voxel clusters produced in the last step, we created a GIS layer containing the two-dimensional outlines of each crown. The crown outline data were placed on a field computer with a built-in GPS receiver. Current position in the field was used to match voxel cluster outlines to the specimens of individual trees of five native species: Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) (DF), western redcedar (*Thuja plicata* Donn ex D. Don) (RC), black cottonwood (*Populus balsamifera* L. ssp. trichocarpa (Torr. & A. Gray ex Hook.) Brayshaw) (BC), bigleaf maple (*Acer macrophyllum* Pursh) (BM), and red alder (*Alnus rubra* Bong.) (RA). The first two of these are coniferous (CO) species, and the last three are deciduous hardwood (HW) species.

Clusters which contained parts of multiple trees, as well as those that contained only part of a single tree, could be identified in the field. These clusters were split along vertical planes or combined as necessary back at the office. This work was done in March of 2011, during which hardwoods were still in a leaf-off condition. To avoid scan angles too far from nadir, we stayed within 60 m of the flight line. In doing so, we were able to identify 22 to 29 individuals of each species, totaling 130 trees. Most conifers were large enough that crown segmentation was clear. However, we had some difficulty separating crowns of hardwood species growing in close proximity. Because we desired certainty that only a single tree crown is represented by a cluster, we skipped a small number of trees which could not be accurately deciphered. Table 2 lists statistics describing the height distribution by species of the trees in the final training data set.

2.3. Waveform and Point Extraction

Using the voxel cluster representing each of the trees identified in the field data, we were able to extract the waveforms that cross each tree from indexed waveform data. First, we found the set of voxels, $\mathbf{V}_{top}$, that represent the highest layer in each row and column combination. To ensure that the collected waveforms represent only the tree of interest, we first discarded all waveforms that do not start within 3.0 m Euclidean distance from a voxel center in $\mathbf{V}_{top}$. Additionally, we also removed waveforms from this set that do not actually cross through a voxel in $\mathbf{V}_{top}$. Doing so should omit waveforms that cross through other objects prior to hitting the cluster of interest, which could introduce errors into our classification.

Table 2. Tree height statistics by species of the trees contained in the training data.

Species	Count	Min.	25th Pct.	Median	75th Pct.	Max.
		(m)	(m)	(m)	(m)	(m)
BC	24	30.52	36.16	37.67	38.88	42.76
BM	22	25.78	26.90	28.54	31.34	35.84
DF	29	25.03	31.21	35.35	37.70	40.89
RA	28	16.23	22.31	25.23	29.30	35.53
RC	27	24.67	29.30	30.66	32.92	38.87

Additionally, using the voxel-array indexing, all peaks from the decomposed waveforms falling within the voxels of a cluster could be very quickly identified. We retained all information about all peaks contained within the voxel cluster of interest.

2.4. *Classification Variables*

2.4.1. Fourier Transform Characteristics

The discrete Fourier transform decomposes a time series of length N into a sum of sine and cosine wave components of N different frequencies. The coefficient for each component frequency can be computed as

$$X_j \equiv \sum_{k=0}^{N-1} x_k \left[\cos\left(\frac{2jk\pi}{N}\right) - \imath \sin\left(\frac{2jk\pi}{N}\right)\right] = a_j + b_j \imath \;\; \forall \;\; j \in 0 \ldots N-1 \tag{13}$$

where coefficient X_j is complex-valued, and gives an exact specification of the component wave function

$$f_j(t) = \left[\frac{a_j}{N}\cos\left(\frac{2jt\pi}{N}\right) - \frac{b_j}{N}\sin\left(\frac{2jt\pi}{N}\right)\right] + \left[\frac{b_j}{N}\cos\left(\frac{2jt\pi}{N}\right) + \frac{a_j}{N}\sin\left(\frac{2jt\pi}{N}\right)\right]\imath \tag{14}$$

for which we can ignore the imaginary part. The sum of the real parts of these component functions

$$S(t) = \sum_{j=0}^{N-1} \Re\left(f_j(t)\right) \tag{15}$$

gives a wave function that will pass through each sample value in the original time series, as shown in Figure 4. Thus the series of coefficients, X, describes the time series in the “frequency domain”.

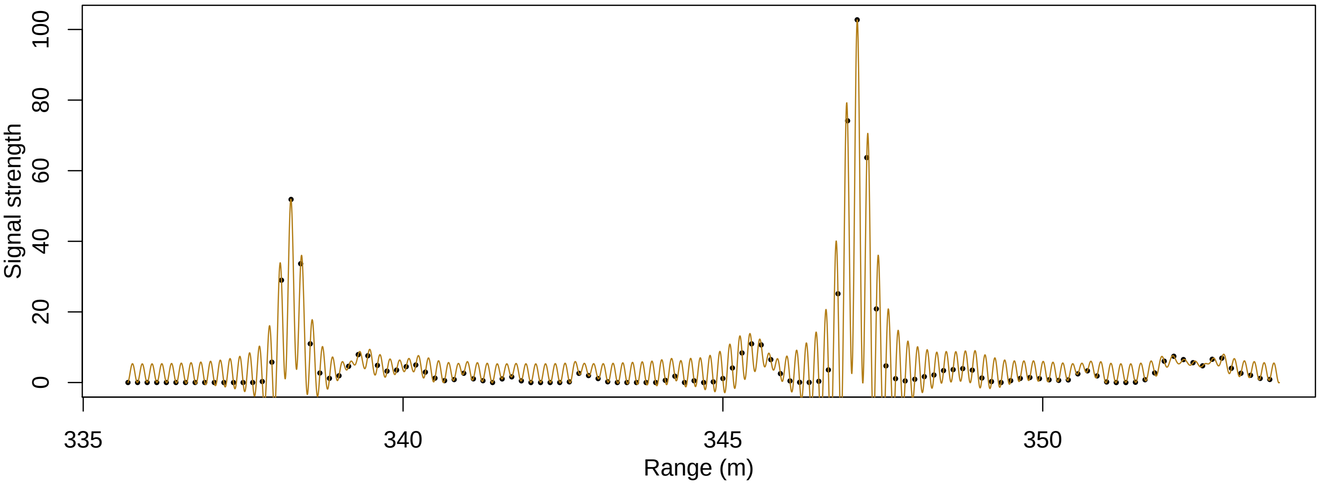
One can compute the phase shift Φ_j and amplitude A_j of each component wave, $f_j(t)$, using Equations (16) and (17), respectively.

$$\Phi_j = \arg(X_j) = atan2(b_j, a_j) \tag{16}$$

$$A_j = \left(\sqrt{a_j^2 + b_j^2}\right)/N \tag{17}$$

A_j is meaningful in this case because it represents the degree of periodicity within the original waveform samples at frequency j/N in cycles per sample. If the inter-sample distance is known, then the frequency can be represented in cycles per meter along the trajectory of the laser pulse. Note that when $j = 0$, the frequency is 0 and this particular component function is a horizontal line crossing the y-axis at A_0. This horizontal line acts as an intercept term in the model, and A_0 will equal the mean sample value in the original waveform.

Figure 4. The combination of all component waves from the discrete Fourier transform, given in Equation (15), passes through each point in the original time series.



The fft function in the R programming language [33] was used to compute the discrete Fourier transform on each waveform collected for all trees. While many waveforms were longer, all waveforms had at least 60 samples. For this reason, and because the number of samples affects the frequencies represented in Fourier transform, we limited our analysis to only the first 60 samples of each waveform.

Additionally, the fft function restricts the coefficients for frequencies higher than $1/2$ cycle per sample to be the complex conjugates of the coefficients of lower frequencies. This is necessary because, without any restriction, two parameters for each component (essentially A_J and Φ_j) would need to be estimated. Such a model would have infinitely many solutions. As a result, only the first 31 components (0 to 30) have any real meaning to our analysis. Accordingly, only the amplitudes for the first 31 frequencies from each transformed waveform were kept.

The median, m_j, and interquartile range (IQR), q_j, of each of these 31 amplitude values were computed across all waveforms hitting each tree. This differs from the mean and variance used in Vaughn *et al.* [28] because it was believed that large outliers within the amplitudes for a given frequency could highly affect results. The median and IQR were chosen instead because they are less influenced by such outliers. In total, we built a dataset of 62 Fourier transformation variables, m_0 to m_{30} and q_0 to q_{30}, for each tree.

2.4.2. Point and Voxel Cluster Characteristics

The peaks extracted from the waveforms are equivalent to three-dimensional points, such as those in a discrete point Lidar dataset. Several variables were computed from the collective properties of these points. In the past, many such variables have been proposed, and these can be classified into one of two categories: point arrangement statistics and intensity statistics. The former will yield information about crown shape, while the latter gives information about the ability of a tree's foliage to reflect near-infrared light.

In order to obtain information about both crown shape and reflective properties, we used both the point representation and the voxel cluster representation of each tree. Most of the variables were chosen to mimic those presented in previous studies. A few were created in the hope of obtaining information similar to that provided by the Fourier transformation of waveform data. All of these point-derived variables can be theoretically computed from a modern discrete point Lidar dataset with at least three returns per pulse as well as recorded intensity data. For each tree in the dataset we computed the following variables for use as predictors in the test classifications:

$h_{25}, h_{50}, h_{75}, h_{90}$: These four variables are estimates of the 25th, 50th, 75th and 90th percentiles of the relative height (point height relative to maximum point height) distribution.

i_1, i_2, i_3 : These three variables are the mean intensities of the first, second, and third peaks recorded for each pulse.

d_{12}, d_{13}, d_{23} : These three variables are the mean Euclidean distances from first to second, first to third, and second to third returns, respectively, across all pulses.

$\hat{\lambda}$: We collected all distances between two consecutive peaks, regardless of position in the waveform. This variable is the estimated rate parameter of an exponential distribution, with form $f(x) = \lambda e^{-\lambda x}$, fit to these distances, across all pulses for each tree.

p_{top} : For each tree, we reserved the set of member voxels which contain the highest recorded peaks in the vertical region projected above their respective row and column. This set, $\mathbf{T}$, was also limited to contain only voxels positioned above 6 m from ground level. Peaks falling within voxels in set $\mathbf{T}$, represent those occurring near the surface of the tree, and the proportion of such peaks within the total number of peaks might represent canopy surface penetrability. The proportion of all peaks falling in the same row and column as a voxel in $\mathbf{T}$, and positioned less than 1.5 m vertical distance this voxel was computed to create this variable.

r_{area} : We also reserved the set, $\mathbf{U}$, of all member voxels positioned within the top eight voxel layers of each tree. The flattened projection of $\mathbf{U}$ gives a two-dimensional representation of the shape of the crown top. A convex hull can be drawn around the row and column centers of this projection. For illustration, both the projection and the convex hull are depicted in Figure 5(a). The area of the convex hull and the area of the projection may differ in cases where a tree is more branchy and less smooth along the outline. To measure this difference, we computed this variable which is the area of the projection divided by the area of the convex hull. Branchy trees should give lower values of this variable.

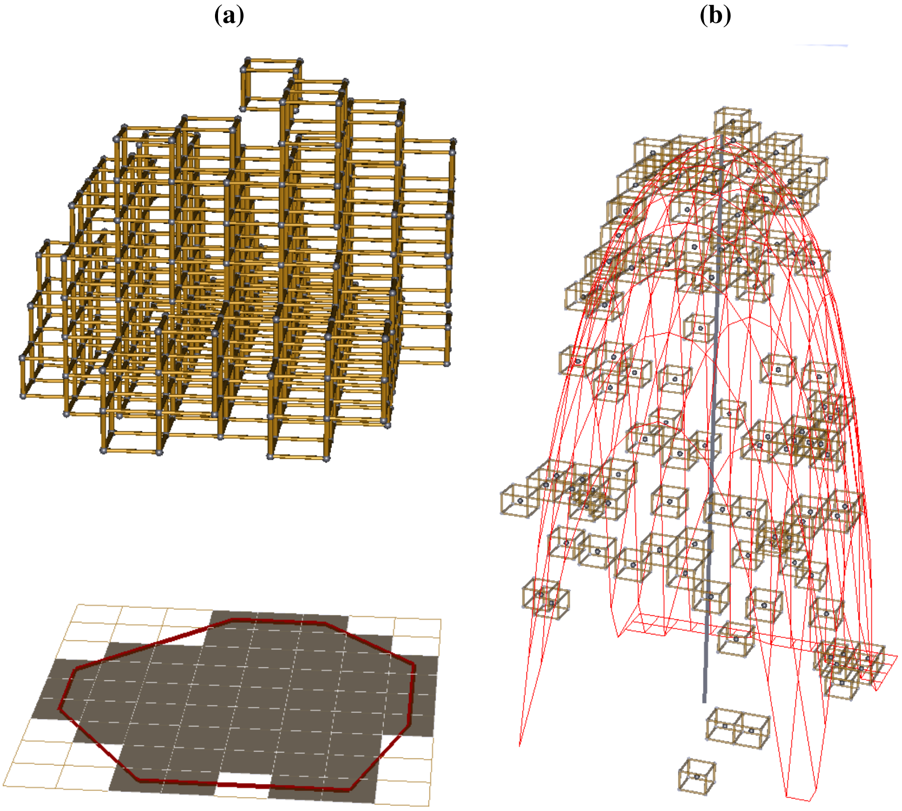
$p_{n1}, p_{nt,}p_{nb}$: Using only the layer, row, and column address of member voxels of a tree, one can easily compute which neighbors a given voxel has. These three variables are the proportions of voxel members of the voxel cluster containing only one neighbor voxel, only a bottom neighbor, and only a top neighbor, respectively. Re-indexing the points to a 0.5 m × 0.5 m × 0.5 m voxel array improved the power of these variables.

s_a, s_b : A function, given in Equation (5), was fit in polar coordinates to only the centers of the voxel cells in set **T** described above

$$dZ = e^{(a+b\ sin(R-c))\Theta} - 1 \tag{18}$$

Here, dZ is the vertical difference between the given cell and the maximum layer within the voxel cluster; R and Θ are the horizontal polar coordinates of the given cell center from the cluster centroid in the horizontal plane; and a, b and c are parameters estimated by the R function `nls`. An example model is plotted in Figure 5(b). Parameters a and b were kept to form these two variables, while parameter c was discarded.

Figure 5. (**a**) A visual depiction of variable r_{area}. The top eight layers of a tree's voxel representation, set **U**, are shown projected to two dimensions. The variable r_{area} is the area of this projection divided by the area of the convex hull encompassing the row and column centers of this projection. (**b**) A visualization of the crown surface model in Equation parametrized by the variables s_a and s_b. Only the voxels in set **T** are shown.



Many of the above variables describe crown shape and should be almost entirely unrelated to the information available from the Fourier transforms of the original waveforms. These include the height percentiles ($h_{25}, h_{50}, h_{75}, h_{90}$), the voxel neighbor statistics ($p_{n1}, p_{nt,}p_{nb}$), the surface model parameters (s_a, s_b), as well as r_{area}.

Conversely, the remaining variables not mentioned above may be correlated with the Fourier transform variables. The Fourier transform statistics indirectly provide quantifications of both the propensity of samples at different distances apart to be part of peaks and the scale differences of these peaks. If variables exist that can alternatively describe these traits, they might act as surrogates for the information available in the Fourier transform variables.

The remainder of the listed point-based variables were intended to provide this surrogate information. A high value of p_{top} may be correlated to waveform shape as a highly reflective crown surface would produce a large first peak and fewer trailing peaks. The ordered intensity statistics (i_1, i_2, i_3) and the distance statistics (d_{12}, d_{13}, d_{23} and $\hat{\lambda}$) describe in several dimensions the average distances between peaks and the average intensities of those peaks. Any further detail about the relationships among the individual peaks in the waveforms would most likely require waveform data to compute, as the only description of each peak in discrete point Lidar datasets is maximum intensity.

2.5. *Correlations and Data Reduction*

2.5.1. Principal Components of Fourier Variables

The Fourier variables computed from the waveforms have very high dimension. Because of the high correlation among the amplitudes for all frequencies, it is very likely that the majority of this information could be described in fewer dimensions. Therefore, we ran separate principal component analyses on both the median and the IQR statistics for frequencies 1 though 31. This was done with the `prcomp` function in the R programming language. Due to the large difference in scale among the 30 frequencies, the options to both center and scale the variables to have mean of 0 and a standard deviation of 1.0 were set prior to computing the singular value decomposition. The first five components of the rotation of the frequency medians (c_{m1} to c_{m5}) as well as the first five components of the IQRs (c_{q1} to c_{q5}) were then used as predictor variables in the classifications. As mentioned previously, the amplitude for frequency 0 on an individual waveform is the mean waveform sample intensity. The median and the IQR of this amplitude value were believed to be especially beneficial, and were therefore kept as predictors and not included in the principal component reductions. This procedure reduced the dimensionality of the Fourier transform variables from 62 to 12.

2.5.2. Canonical Correlations

We used the R function `cancor` to investigate the inherent correlation between two sets of variables. The first set, $\mathbf{X}$, contains those variables computed from the waveforms directly ($c_{m1}, \ldots, c_{m5}$, $c_{q1}, \ldots, c_{q5}$, m_0, and q_0) and the second set, $\mathbf{Y}$, consists of those variables, computed from the extracted discrete point data, that may contain some of the same information provided by the Fourier transforms of the waveforms (i_1, i_2, i_3, d_{12}, d_{13}, d_{23},$\hat{\lambda}$, and p_{top}). The procedure yields eight pairs of canonical variates (U_i, V_i), such that each U_i is a linear combination of the columns in $\mathbf{X}$ and each V_i is a linear combination of the columns of $\mathbf{Y}$. All variates are orthogonal to each other except the U_i and V_i within each pair, which are as highly correlated as possible. Examining these correlations, as well as the influence of the original variables in the canonical variates, gives insight into the nature of the relationships between the two sets of variables.

2.6. Classification

We were not only interested in the overall performance of the variable groups, but also for which species comparisons of the individual variable groups performed best. After comparing the predictions from several routines, including linear discriminant analysis, classification trees, and the neural-network approach, support vector machine (SVM) classification performed the best overall. SVM is typically used as a kernel-based algorithm, in which a linear algorithm is applied to a non-linear mapping $\phi(\mathbf{x})$ of the original data onto a higher-dimensional space. This allows for curvature in the surface dividing two groups in the original space. The SVM algorithm uses only the dot products of vectors in the higher-dimensional space, and a kernel function allows computation of these higher-dimensional dot products directly from the original data vectors. We chose the "radial basis" kernel function, given in Equation (19), because it performed the best during initial testing.

$$< \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) >= e^{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{19}$$

We used the `svm` function, part of the R library `e1071` [34], to perform support vector machine classification. The default setting for the γ parameter, 1 divided by the number of columns in the predictor matrix, was used because any adjustments resulted in reduced accuracy. Unable to find any improvements, we left all other parameters of the svm function at their default settings.

We tested several predictor groups, individually and in combination, for the classifications. These groupings were: (**a**) h_{25}, h_{50}, h_{75}, and h_{90}; (**b**) i_1, i_2, and i_3; (**c**) d_{12}, d_{13}, d_{23}, and $\hat{\lambda}$; (**d**) p_{top} and r_{area}; (**e**) p_{n1}, p_{nt}, and p_{nb}; (**f**) s_a and s_b; (**g**) "Point": all variables in groups **a** to **f**; (**h**) $c_{m1}, \ldots, c_{m5}$; (**i**) $c_{q1}, \ldots, c_{q5}$; (**j**) m_0 and q_0; (**k**) "Fourier": all variables in groups **h** to **j**; and (**l**) "All": all variables combined. This breakdown was performed to understand the utility of each group for species classification as compared to the other groups. In particular we were concerned with the improvement in prediction accuracy that might come with incorporating information from waveform Lidar over using discrete point Lidar alone. Of course, the actual degree of improvement is dependent on our choices of variables from each dataset. However, the large number of variables were particularly chosen in order to span the full range of information available from the discrete point data.

We also applied the SVM on seven different classifications for each predictor group. This was to examine the species differences that were most sensitive to each predictor group. This information can help to better understand how the Fourier information either does or does not improve classification. The seven classifications were: (1) all species; (2) only hardwood species; (3) all species remapped to either CO or HW; (4) BC and BM only; (5) BC and RA only; (6) BM and RA only; and (7) DF and RC only.

For each classification, five-fold cross validation was used to test the performance of each predictor group. The trees in each species (or each growth form) were randomly split into five groups of similar size, and these species groups were combined into five data groupings. The species predictions for each grouping were performed by a decision rule based on the other four groupings combined. In this way no trees fall in a training set and validation set at the same time. Overall accuracy for each predictor group and classification was computed as the number of correctly predicted trees divided by the total number of trees.

For the classification of all five species, we performed the exact test by Liddell [35] to ascertain whether the addition of the Fourier transformation variables significantly improved prediction accuracy in classification of all five species. Liddell's test is designed to compare two proportions, measured on the same subjects. In this case, we used the proportion of correctly predicted trees using the variables in group g (point-derived variables) and the proportion of correctly predicted trees using the variables in group l (all variables).

3. Results

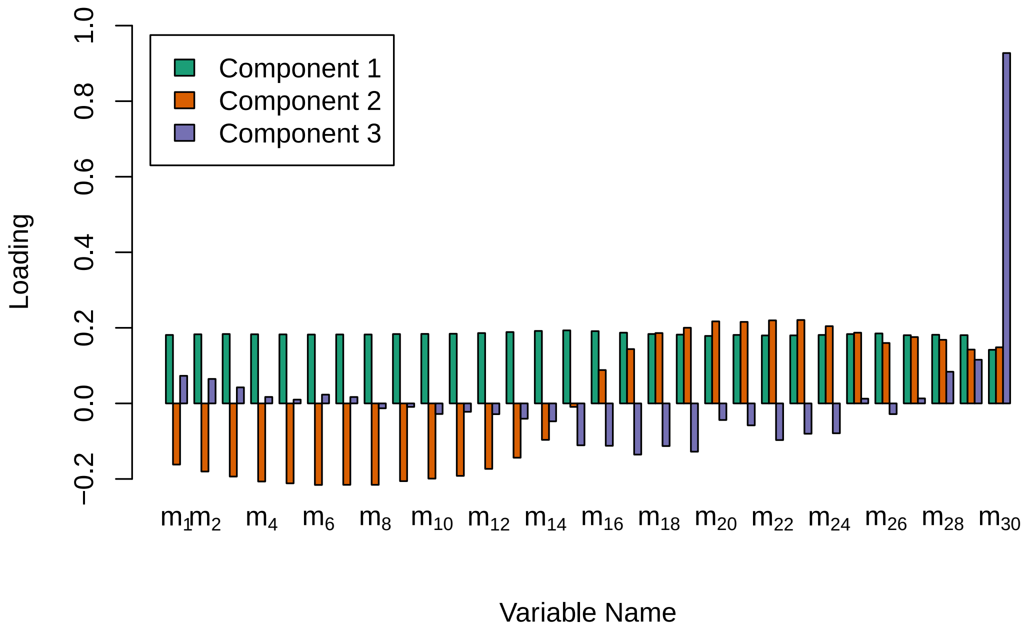
3.1. Correlations and Data Reduction

3.1.1. Principal Components of Fourier Variables

The first five principal components of the Fourier median variables (m_1 to m_{30}) had standard deviations of 5.11, 1.65, 0.68, 0.51, and 0.31 respectively. The loadings of the first three of these components are displayed in graphical form in Figure 6. Based on the factor loadings in this figure, these first three components can be interpreted as:

1. the mean of the amplitudes over all frequencies
2. a comparison of the lower half of the frequencies against the higher half
3. the middle frequencies against the combined low and high frequencies.

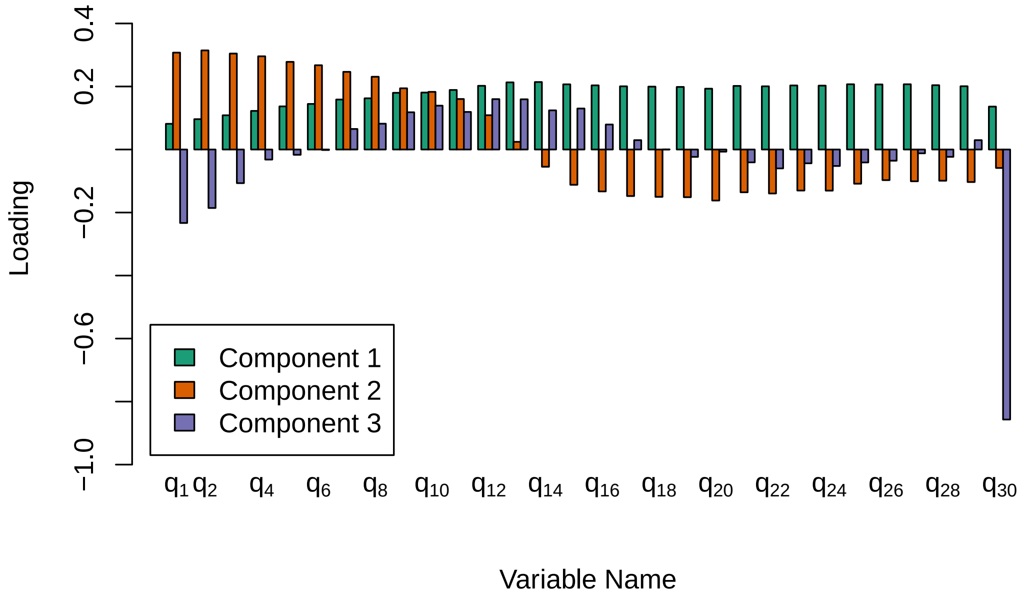
Figure 6. Loadings of the first three principal components of the Fourier median variables (m_1 to m_{30}).



Component 1 is likely a measure of the total pulse energy reflected by the target at the sensor. In general, the lower frequencies in the transformed waveforms have amplitudes several orders of magnitude larger than the amplitudes of the higher frequencies. After centering and scaling the data for each frequency, which removes this imbalance, the influence of each frequency becomes more equal. Components 2 and 3 measure the influence of the different groups of frequencies relative to the other groups.

For the IQRs, the standard deviations were 4.54, 2.72, 0.82, 0.67, and 0.40 respectively. Figure 7 shows the loadings for the first three components for the IQR variables (q_1 to q_{30}). A similar pattern to what appears in Figure 6 is displayed here as well. The first factor measures the mean amplitude, and the second and third factor compare the frequencies in similar groups.

Figure 7. Loadings of the first three principal components of the Fourier interquartile range variables (q_1 to q_{30}).



3.1.2. Canonical Correlations

The first two pairs of canonical variates were highly correlated, with correlations of 0.98 and 0.90 for (U_1, V_1) and (U_2, V_2) respectively. The following six pairs had correlations of 0.50 or lower. The high amount of correlation between the first two pairs demonstrated that there is some overlap of information among the two datasets. In other words, a portion of information from the Fourier transforms of the waveforms can be obtained from patterns from the discrete points extracted from the waveforms.

Coefficients from the first two rotations are shown in Table 3. These first rotations result in the variates U_1, U_2, V_1, and V_2. Given the coefficients and means given in the table, values of U_1 are most influenced by the variables c_{m1}, m_0 and q_0. These variables are all related to the amount of energy received by the sensor in the Lidar instrument. Similarly, the intensity means, i_1 to i_3, show strong influence in both V_1 and V_2. This result is nearly as visible by just looking at the correlation between some of these variables alone. In fact, i_1 shares a correlation of 0.85 with c_{m1} when the two variables are compared directly. Surface point density, p_{top}, played little part in either V_1 and V_2, suggesting that this information may not be obtainable from the waveform Fourier transformations directly. On the other hand, d_{12}, d_{13}, and d_{23} do play a significant part in V_1 and V_2, suggesting that some of this information overlaps between the two variable sets.

Table 3. Coefficients from the canonical correlation procedure for the first two canonical variates of both datasets. Mean and standard deviation of all variables are included for reference.

			Coefficient					Coefficient	
Var.	Mean	S.D.	U_1	U_2	Var.	Mean	S.D.	V_1	V_2
c_{m1}	0.0	5.11	0.013	–0.035	i_1	95.8	17.15	0.005	–0.002
c_{m2}	0.0	1.65	–0.008	0.013	i_2	34.5	6.38	0.005	0.002
c_{m3}	0.0	0.68	0.000	–0.017	i_3	14.5	4.00	0.004	0.017
c_{m4}	0.0	0.51	0.001	–0.090	d_{12}	1.4	0.08	–0.146	0.169
c_{m5}	0.0	0.31	–0.014	0.046	d_{13}	2.6	0.15	0.151	–0.309
c_{q1}	0.0	4.54	–0.007	0.009	d_{23}	1.3	0.09	–0.167	–0.021
c_{q2}	0.0	2.72	–0.001	0.009	$\hat{\lambda}$	0.3	0.03	–0.119	0.039
c_{q3}	0.0	0.82	0.001	–0.000	p_{top}	0.2	0.07	–0.049	–0.044
c_{q4}	0.0	0.62	0.000	–0.005					
c_{q5}	0.0	0.40	0.014	–0.040					
m_0	11.6	1.60	0.036	0.100					
q_0	3.0	0.64	0.042	–0.036					

3.2. *Classification Results*

The combined variables from the discrete point and waveform datasets worked well for the classification of the five species. An overall accuracy of over 85 percent was achieved, with 111 out of 130 trees correctly classified. Table 4 is the confusion matrix for the classification of all species using all variables. Cottonwood (BC), maple (BM) and Douglas-fir (DF) seem to be most easily separated from the rest of the species as well as each other. The largest pairwise confusion occurred between alder (RA) and cedar (RC), a hardwood and a conifer. Similarly, confusion between RA and both conifers was greater than that between the three hardwood species. While there was some confusion between the conifers, it was all in one direction; no DF were predicted to be RC.

Table 4. Confusion matrix for the classification of all five species using all available predictor variables from both the discrete point and waveform data.

			Predicted			Producer
Species	BC	BM	DF	RA	RC	Accuracy
BC	22	0	0	1	1	91.7
BM	1	19	0	1	1	86.4
DF	1	1	26	1	0	89.7
RA	1	0	2	22	3	78.6
RC	1	0	2	2	22	81.5
User Accuracy	84.6	95.0	86.7	81.5	81.5	*85.4

*Overall accuracy, $\hat{\kappa} = 0.817$

Tables 5 and 6 give the results, as overall percent accuracy and kappa statistics respectively, for all classifications using each predictor group. In all but two cases, the addition of the twelve Fourier transformation variables improved the accuracy over the eighteen point variables. In the five species classification, the addition increased the overall accuracy by over six percent (8 of the 130 trees in the dataset). The Liddell test procedure returned a test statistic value of 2.40 and a one-sided p-value of 0.0384. This indicates that there is a very low probability that including the Fourier transformation variables did not actually improve classification accuracy.

Table 5. Overall percent classification accuracy results of the support vector machine applied with a five-fold cross validation to different predictor variable groups and species groups.

	Species Classification Group						
			CO	BC	BC	BM	DF
Pred. Group	All	HW	HW	BM	RA	RA	RC
	(%)	(%)	(%)	(%)	(%)	(%)	(%)
a–$h_{25}, h_{50}, h_{75}, h_{90}$	33.1	54.1	59.4	63.0	80.8	72.0	73.2
b–i_1, i_2, i_3	53.1	66.2	65.4	80.4	65.4	70.0	96.4
c–$d_{12}, d_{13}, d_{23}, \hat{\lambda}$	40.0	60.8	75.9	84.8	65.4	84.0	51.8
d–p_{top}, r_{area}	51.5	63.5	65.4	80.4	63.5	76.0	83.9
e–p_{n1}, p_{nt}, p_{nb}	46.9	52.7	78.9	58.7	75.0	68.0	71.4
f–s_a, s_b	38.5	55.4	77.4	87.0	50.0	70.0	57.1
g–Point	79.2	87.8	85.0	97.8	94.2	88.0	91.1
h–$c_{m1}, \ldots, c_{m5}$	57.7	59.5	67.7	69.6	78.8	82.0	89.3
i–$c_{q1}, \ldots, c_{q5}$	49.2	50.0	67.7	73.9	63.5	72.0	91.1
j–m_0, q_0	46.9	52.7	75.2	78.3	69.2	64.0	80.4
k–Fourier	66.2	71.6	75.9	82.6	88.5	84.0	92.9
l–All	85.4	90.5	86.5	97.8	94.2	92.0	94.6

Of the point-derived variables, no single group seemed to perform best in all situations. In fact, each individual group seemed to have species for which it was highly important. The relative height percentiles in group $\mathbf{a}$ were best for separating BC and RA. The point intensities in group $\mathbf{b}$ did well in BC versus BM, but performed the best on the conifers. Inter-peak distance measures in group $\mathbf{c}$ worked the best in differentiating BM from BC and RA. The crown roughness and permeability variables in group $\mathbf{d}$ performed relatively well for distinguishing BM as well as for splitting the two conifers. Voxel neighbor statistics in group $\mathbf{e}$, a measure of crown surface texture, worked best when distinguishing conifers from hardwoods. Finally, the crown surface shape variables in group $\mathbf{f}$ excelled at distinguishing BC and BM.

Conversely, the Fourier transformation variable groups had a lot less variation in performance across the species groups. The median variables in group $\mathbf{h}$ were the best overall, but the other variables did not follow by far in any of the individual classifications. Each of the groups of Fourier transformation variables did better on an individual basis than most or all of the point variables in the five species classification.

Table 6. Kappa statistics of the support vector machine applied with a five-fold cross validation to different predictor variable groups and species groups.

	Species Classification Group						
			CO	BC	BC	BM	DF
Pred. Group	All	HW	HW	BM	RA	RA	RC
a–$h_{25}, h_{50}, h_{75}, h_{90}$	0.158	0.298	0.163	0.244	0.615	0.426	0.464
b–i_1, i_2, i_3	0.411	0.493	0.309	0.604	0.312	0.400	0.928
c–$d_{12}, d_{13}, d_{23}, \hat{\lambda}$	0.245	0.405	0.514	0.692	0.295	0.678	0.031
d–p_{top}, r_{area}	0.391	0.444	0.287	0.604	0.254	0.508	0.679
e–$p_{n1}, p_{nt,}p_{nb}$	0.332	0.279	0.576	0.183	0.499	0.324	0.434
f–s_a, s_b	0.226	0.328	0.538	0.736	0.000	0.370	0.133
g–Point	0.740	0.817	0.700	0.956	0.884	0.756	0.821
h–$c_{m1}, \ldots, c_{m5}$	0.468	0.388	0.325	0.385	0.578	0.629	0.785
i–$c_{q1}, \ldots, c_{q5}$	0.362	0.246	0.335	0.469	0.271	0.421	0.821
j–m_0, q_0	0.331	0.284	0.493	0.561	0.377	0.255	0.608
k–Fourier	0.575	0.570	0.508	0.650	0.766	0.672	0.857
l–All	0.817	0.857	0.727	0.956	0.884	0.838	0.893

4. Discussion

For five species the overall accuracy of just over 85 percent achieved by the combination of point-derived and Fourier transformation variables compared favorably to similar research on several species [2,12,36,37]. Most Lidar research focuses on two to three species, and as expected, increasing this number is generally associated with a loss in overall accuracy. As shown in Table 5, we consistently achieved over 90 percent accuracy when the number of species is reduced to two or three. While it does little good to compare different study areas directly by this measure, we were comforted that the observed accuracies were in the same neighborhood as some of the higher accuracies in previous publications [1,16,19,22].

Because of the large number of predictor variables, dimensionality was a possible reason for concern. The SVM function is a good choice for this study because it is able to handle a large number of dimensions in the predictor set. However, one negative aspect of the SVM function is that it can require some fine-tuning for maximal performance for a given data set. Performing such a customization for several predictor groups and several species comparisons would have both allowed too much bias and taken far too much time. This lack of tuning likely results in strange predictive behavior. One example of such behavior is the reduction in accuracy from 96.4 to 91.1 percent comparing the classifications of the two conifers using variables in group g versus those of group b alone. While such a difference was disturbing to see, we were not as concerned with individual classification results as we were with general patterns across predictor groups and species combinations.

We reduced the dimensionality of the two sets of Fourier transformation variables, medians and IQR, to six variables each using principal component analysis. Despite the heavy reduction, the 72 percent accuracy achieved by all twelve of the Fourier transformation variables in the classification of the three hardwoods nearly matched the 75 percent accuracy reported previously in Vaughn *et al.* [28]. This is a significant result because many of the amplitudes for individual frequencies are correlated, and we can extract the important information contained in these amplitudes using a much smaller number of orthogonal predictors.

The strongest individual predictor group for the five species classification was group $\mathbf{h}$, which is the combination of principal components for the Fourier transform medians. This group was not strongest for any of the other classifications, but it does perform nearly the best for differentiating cottonwood (BC) and alder (RA) and for differentiating the two conifers, Douglas-fir (DF) and cedar (RC). This indicates that these variables are actually quite strong as predictors. If other structural information was not available due to difficulty in segmenting out individual crowns, these variables might still be useful for at least classifying pixel-sized areas of the canopy.

The group containing all waveform information, $\mathbf{k}$, was better than the group containing all point-derived variables, $\mathbf{g}$, for only one pairing of species, DF and RC. This was a surprise because conifers were left out of the previous work in Vaughn *et al.* [28] due to the poor performance of the Fourier transform variables on the same two species. For this study, we limited the scan angle to less than 12 degrees off nadir. This might explain the observed reversal, as one could imagine scan angle having a large effect on the amount of conifer crown intercepting the laser pulse.

The point-derived variable group, $\mathbf{g}$, or components of this group performed very well on all species comparisons. Several of these variables are based on other publications, as will be described below, and similar results have been found before. The intensity information worked best for the classification of all five species as well as for differentiating the two conifers. Previous results have been mixed on the utility of intensity information. Ørka *et al.* [23] used only intensity, while Holmgren and Persson [21] found only the standard deviation of intensity to be important. Moffiet *et al.* [2] found that intensity varied too highly to be used as a predictor.

The point height distribution statistics in group $\mathbf{a}$, which appear in similar form in many species classification studies [1,19–21], were generally unimpressive as an individual predictor group. The main exception is for the classification of BC and RA, where a classification accuracy of over 80 percent is achieved with just these four height distribution percentiles. As with intensity, results have been mixed in the past with this group of variables. Vauhkonen *et al.* [19] achieved fairly good classification with just the relative height distribution percentiles before improving on these results with other variables. Holmgren and Persson [21] found that the weakest individual variable was the 90th percentile of relative height. Ørka *et al.* [1] also find little value to relative height statistics in species classification.

In Tables 5 and 6, we can see that the inter-peak distance variables in group $\mathbf{c}$ performed very well distinguishing maple (BM) from the two other hardwood species. The means of three of the four variables in this group, i.e., d_{12}, d_{13}, and d_{13}, are smaller for BM than for all other species. As the name suggests, bigleaf maple has very large leaves that may be larger than even the pulse footprint. Each leaf hit is then very likely to record a noticeable peak in the return signal. This might result in more detectable peaks close to the crown surface. The crown surface model parameters in group $\mathbf{f}$ are also quite strong at differentiating BM from cottonwood (BC). More open-grown maples tend to present a more dome-like form, which is represented in our crown surface model by smaller values of the s_a parameter.

The voxel-based neighbor statistics in group $\mathbf{e}$ were strong predictors for distinguishing conifers from hardwoods, but this group of texture variables underperformed as an individual group in all other classifications. The realizations of these variables varied greatly for the BC and BM trees, but were much more stable for the other species. The reason for this difference is unclear, though there seems to be some association between larger heights and larger values of p_{n1} for these two species. Perhaps larger cottonwoods and maples are more prone to lone branches that would lead to a larger number of one-neighbor cells. Including height as a predictor might account for this difference. Ørka *et al.* [1] also found that many of their predictors were dependent on height.

Few authors have investigated Lidar-derived crown texture directly as a predictor variable. Vauhkonen *et al.* [19] looked at textural features of the Lidar-derived canopy height model. While analysis of changes in intensity characteristics of the returns can be seen as texture analysis [18], this neglects the three-dimensional texture that is evident in many tree crowns. In high spatial resolution raster imagery analysis, texture has been considered important for species identification [38,39]. It is interesting that this idea did not transfer over to point cloud analysis. One explanation might be that three-dimensional texture is difficult to quantify. We determined that canopy texture, or "roughness", could be measured more easily in the voxel representation of the data. As with two-dimensional raster data, it was very simple to identify the neighbors of a voxel using the row, column, and layer indices.

Our intentional attempt to create variables from the discrete point data that aliased the information available in the Fourier transformations was not successful. These variables, such as those in groups $\mathbf{b}$ and $\mathbf{c}$, should contain information about intensity relationships among the peaks and inter-peak distances. According to the canonical correlation analysis, only two of the eight canonical pairs had high correlations. This indicates that we did indeed capture some of the same information available from the Fourier transformations. However, the remaining information was influential enough to improve the classification accuracy in most of the species comparisons. We expect that a noticeable part of these results may stem from our choice of variables, and a different choice of variables could lead to a different conclusion. However, much detail is lost in the conversion of wave signal to discrete point data, and we suspect that some information important to species detection will always be part of this loss.

The trees measured in this study, most notably the conifers, are mostly open grown. The results as they stand would not directly translate to a high density commercial forest. With the increased density a smaller portion of each tree's crown would be uniquely identifiable. Because the waveform information used in this study only contains one-dimensional information about position from wave start, the density should not greatly affect the results reported here. However, many of the spatial variables we used, such as those in groups $\mathbf{d}$, $\mathbf{e}$ and $\mathbf{f}$ would likely be affected by this reduced crown visibility. The effects would probably be largest in hardwood stands due to the additional difficulty of differentiating adjacent hardwood trees with intermingling branches and no distinct tops. In such a case, the result would most likely be an increase in the power of waveform information over crown shape information for species classification.

5. Conclusions

The prevalence of aerial Lidar for forest inventory information has increased over the last decade, and discrete point Lidar data has already shown much promise in distinguishing individual tree species [19,21]. In this study, we were able to provide one more example of high tree species classification accuracy from discrete point airborne Lidar data. Using only variables available from very dense discrete point Lidar data, an overall classification accuracy of 79.2 percent (kappa = 0.740) was found for five native commercially viable species. Multiple comparisons of two or three species resulted in even greater accuracy, achieving as high as 97.8 percent (kappa = 0.956).

For classification with discrete point data only, we introduced several new variables. In the five species classification, two groups of these new variables performed second and third best of all the groups. In each of the small classifications of two or three species or species groups, one of the groups of new variables performs at least second best. Note that the computation of each of these variables was very simple using a voxel representation of each tree. To provide such a representation, we introduced a new segmentation algorithm that can be easily adapted to local crown properties.

Perhaps of most interest, we discovered that summary information derived from entire waveforms provided predictive power above and beyond that of the discrete point data alone. This addition raised the overall accuracy to 85.4 percent (kappa = 0.817). The waveform information was important for separating Douglas-fir from western hemlock, increasing accuracy 3.5 percent (kappa increase = 0.072), and bigleaf maple from red alder, increasing accuracy 4.0 percent (kappa increase = 0.082). For other species comparisons, waveform information provided no gain in accuracy.

Other researchers have found that decomposing airborne waveform Lidar data into discrete points is useful for classification purposes [20,27,40], using peak characteristics derived while decomposing the waveforms into discrete peaks. However, to our knowledge, this was the first study to directly compare waveform to discrete Lidar data for species classification performance. One should note that no such test can be entirely conclusive due to the immense number of options for variables and classification techniques. Regardless, our results, along with positive waveform results in the other studies, show that any additional cost of waveform data may very likely be sensible if optimal accuracy is a primary goal. Further investigation into the potential of this data format may lead to a complete species-level remote sensing-based inventory from Lidar data alone that can consistently provide all the information needed.

Acknowledgements

The authors sincerely thank Terrapoint, USA for the provision of waveform Lidar data that is otherwise difficult to obtain. Thanks to the Precision Forestry Cooperative at the University of Washington, and to the Corkery Family Trust for funding portions of this work. David Briggs and Jim Flewelling provided very attentive and helpful feedback on earlier drafts of this paper. Their help is much appreciated.

References

1. Ørka, H.O.; Næsset, E.; Bollandsås, O.M. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* **2009**, *113*, 1163–1174.
2. Moffiet, T.; Mengersen, K.; Witte, C.; King, R.; Denham, R. Airborne laser scanning: exploratory data analysis indicates potential variables for classification of individual trees or forest stands according to species. *ISPRS J. Photogramm.* **2005**, *59*, 289–309.

3. Korpela, I.; Dahlin, B.; Schäfer, H.; Bruun, E.; Haapaniemi, F.; Honkasalo, J.; Ilvesniemi, S.; Kuutti, V.; Linkosalmi, M.; Mustonen, J.; Salo, M.; Suomi, O.; Virtanen, H. Single-tree forest inventory using Lidar and aerial images for 3D treetop positioning, species recognition, height and crown width estimation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2007**, *36*, 227–233.
4. Vauhkonen, J.; Korpela, I.; Maltamo, M.; Tokola, T. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ.* **2010**, *114*, 1263–1276.
5. Breidenbach, J.; Næsset, E. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. *Remote Sens. Environ.* **2010**, *114*, 911–924.
6. Hyyppä, J.; Hyyppä, H.; Litkey, P.; Yu, X.; Haggrén, H.; Rönnholm, P.; Pyysalo, U.; Pitkänen, J.; Maltamo, M. Algorithms and methods of airborne laser-scanning for forest measurements. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2004**, *36*, 1682–1750.
7. Næsset, E.; Gobakken, T.; Holmgren, J.; Hyyppä, H.; Hyyppä, J.; Maltamo, M.; Nilsson, M.; Olsson, H.; Persson, Å.; Söderman, U. Laser scanning of forest resources: the Nordic experience. *Scand. J. Forest Res.* **2004**, *19*, 482–499.
8. Jung, S.-E.; Kwak, D.-A.; Park, T.; Lee, W.-K.; Yoo, S. Estimating crown variables of individual trees using airborne and terrestrial laser scanners. *Remote Sens.* **2011**, *3*, 2346–2363.
9. Moskal, L.M.; Zheng, G. Retrieving forest inventory variables with terrestrial laser scanning (TLS) in urban heterogeneous forest. *Remote Sens.* **2012**, *4*, 1–20.
10. Leckie, D.G.; Gougeon, F.A.; Walsworth, N.; Paradine, D. Stand delineation and composition estimation using semi-automated individual tree crown analysis. *Remote Sens. Environ.* **2003**, *85*, 355–369.
11. Brandtberg, T. Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. *Fuzzy Sets Syst.* **2002**, *132*, 371–387.
12. Jones, T.G.; Coops, N.C.; Sharma, T. Assessing the utility of airborne hyperspectral and Lidar data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sens. Environ.* **2010**, *114*, 2841–2852.
13. Ke, Y.; Quackenbush, L.J.; Im, J. Synergistic use of QuickBird multispectral imagery and Lidar data for object-based forest species classification. *Remote Sens. Environ.* **2010**, *114*, 1141–1154.
14. Puttonen, E.; Litkey, P.; Hyyppä, J. Individual tree species classification by illuminated-shaded area separation. *Remote Sens.* **2009**, *2*, 19–35.
15. Heinzel, J.N.; Weinacker, H.; Koch, B. Full Automatic Detection of Tree Species Based on Delineated Single Tree Crowns—A Data Fusion Approach for Airborne Laser Scanning Data and Aerial Photographs. In *Proceedings of Silvilaser 2008*, Edinburgh, UK, 17–19 September 2008; pp. 76–85.
16. Holmgren, J.; Persson, Å.; Söderman, U. Species identification of individual trees by combining high resolution Lidar data with multi-spectral images. *Int. J. Remote Sens.* **2008**, *29*, 1537–1552.

17. Leckie, D.G.; Gougeon, F.A.; Hill, D.; Quinn, R.; Armstrong, L.; Shreenan, R. Combined high-density Lidar and multispectral imagery for individual tree crown analysis. *Can. J. Remote Sens.* **2003**, *29*, 633–649.
18. Brandtberg, T. Classifying individual tree species under leaf-off and leaf-on conditions using airborne Lidar. *ISPRS J. Photogramm.* **2007**, *61*, 325–340.
19. Vauhkonen, J.; Tokola, T.; Packalén, P.; Maltamo, M. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. *Forest Sci.* **2009**, *55*, 37–47.
20. Reitberger, J.; Krzystek, P.; Stilla, U. Analysis of full waveform Lidar data for the classification of deciduous and coniferous trees. *Int. J. Remote Sens.* **2008**, *29*, 1407–1431.
21. Holmgren, J.; Persson, Å. Identifying species of individual trees using airborne laser scanner. *Remote Sens. Environ.* **2004**, *90*, 415–423.
22. Korpela, I.; Ørka, H.O.; Maltamo, M.; Tokola, T.; Hyyppä, J. Tree species classification using airborne Lidar—Effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica* **2010**, *44*, 319–339.
23. Ørka, H.O.; Næsset, E.; Bollandsås, O.M. Utilizing airborne laser intensity for tree species classification. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2007**, *36*, 300–304.
24. Kim, S.; McGaughey, R.J.; Andersen, H.E.; Schreuder, G. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sens. Environ.* **2009**, *113*, 1575–1586.
25. Wagner, W.; Hollaus, M.; Briese, C.; Ducic, V. 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. *Int. J. Remote Sens.* **2008**, *29*, 1433–1452.
26. Hollaus, M.; Aubrecht, C.; Höfle, B.; Steinnocher, K.; Wagner, W. Roughness mapping on various vertical scales based on full-waveform airborne laser scanning data. *Remote Sens.* **2011**, *3*, 503–523.
27. Hollaus, M.; Mücke, W.; Höfle, B.; Dorigo, W.; Pfeifer, N.; Wagner, W.; Bauerhansl, C.; Regner, B. Tree Species Classification Based on Full-Waveform Airborne Laser Scanning Data. In *Proceedings of Silvilaser 2009*, College Station, TX, USA, 14–16 October 2009.
28. Vaughn, N.R.; Moskal, L.M.; Turnblom, E.C. Fourier transformation of waveform Lidar for species recognition. *Remote Sens. Lett.* **2010**, *2*, 347–356.
29. Lucy, L.B. An iterative technique for the rectification of observed distributions. *Astron. J.* **1974**, *79*, 745–754.
30. Richardson, W.H. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.* **1972**, *62*, 55–59.
31. Reitberger, J.; Schnörr, C.; Krzystek, P.; Stilla, U. 3D segmentation of single trees exploiting full waveform Lidar data. *ISPRS J. Photogramm.* **2009**, *64*, 561–574.
32. Gupta, S.; Weinacker, H.; Koch, B. Comparative analysis of clustering-based approaches for 3-d single tree detection using airborne fullwave lidar data. *Remote Sens.* **2010**, *2*, 968–989.
33. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2009.

34. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A. *e1071: Misc Functions of the Department of Statistics (e1071)*; R Package Version 1.5-25; TU Wien: Vienna, Austria, 2011.
35. Liddell, F.D.K. Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure. *J. Epidemiol. Community Health* **1983**, *37*, 82–84.
36. Suratno, A.; Seielstad, C.; Queen, L. Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS J. Photogramm.* **2009**, *64*, 683–693.
37. Katoh, M. Classifying tree species in a northern mixed forest using high-resolution IKONOS data. *J. Forest Res.* **2004**, *9*, 7–14.
38. Franklin, S.E.; Hall, R.J.; Moskal, L.M.; Maudie, A.J.; Lavigne, M.B. Incorporating texture into classification of forest species composition from airborne multispectral images. *Int. J. Remote Sens.* **2000**, *21*, 61–79.
39. Dikshit, O. Textural classification for ecological research using ATM images. *Int. J. Remote Sens.* **1996**, *17*, 887–915.
40. Heinzel, J.; Koch, B. Exploring full-waveform LiDAR parameters for tree species classification. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 152–160.