**Title**: Estimating stand volume in broad-leaved forest using discrete-return LiDAR: plot-based approach

**Authors**: Keiko Ioki[a*], Junichi Imanishi[b], Takeshi Sasaki[a], Yukihiro Morimoto[b], Katsunori Kitada[c]

**Affiliations**:

[a]Graduate School of Agriculture, Kyoto University
Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

[b]Graduate School of Global Environment Studies, Kyoto University
Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

[c]Nakanihon Air Service Co., Ltd.
2-10-2, Kyobashi, Chuou-ku, Tokyo 104-0031, Japan

[*]**Corresponding author**:
Tel.: +81 75 753 6083; fax: +81 75 753 6082
E-mail address: ioki@kais.kyoto-u.ac.jp (K.Ioki)

**Abstract**

Quantitative assessment of forests is important at a variety of scales for forest planning and management. This study investigated the use of small-footprint discrete return lidar for estimating stand volume in broad-leaved forest at plot level. Field measurements were conducted at 20 sample plots in the study area in western Japan, composed of temperate broad-leaved trees. Five height variables and two density variables were derived from the lidar data: 25th, 50th, 75th, and 100th percentiles, and mean of laser canopy heights as height variables ($h_{25}$, $h_{50}$, $h_{75}$, $h_{100}$, $h_{mean}$); and ground fraction and only-and-vegetation fraction ($d_{GF}$, $d_{OVF}$) as density variables, defined respectively as the proportion of laser returns that reached the ground, and the proportion of only echoes (i.e., single pulse returns for which the first and last pulses returned from the same point) within vegetation points. In addition, the normalized difference vegetation index (NDVI), which is often used as an estimator for leaf area index (LAI) and above-ground biomass, was derived from multispectral digital imagery as an alternative density variable ($d_{NDVI}$). Nonlinear least square regression with cross-validation analysis was performed with single variables and combinations; a total of 23 models were studied. The best prediction was found when $h_{75}$ and $d_{OVF}$ were used as independent variables, resulting in adjusted $R^2$ of 0.755 and root-mean-square error (RMSE) of 41.90 m$^3$ ha$^{-1}$, corresponding to 16.4% of the mean stand volume, better than or comparable to the prediction models of previous studies.

**Keywords**

Airborne laser scanning , Canopy height, Forest inventory, Stand structure

## Introduction

Obtaining quantitative information of forests at multiple scales is necessary for forest planning and management. The tree volume of a stand has been one of the most important characteristics, both economically and environmentally. As it interacts with total stand biomass, estimating stand volume is important as a potential contributive factor for understanding forest carbon dynamics. However, accurate and extensive inventories of forest are labor demanding and time consuming. As the need for amounts and quality of information increases, remote sensing becomes a more powerful technological instrument in forest management.

Conventional two-dimensional remote sensing techniques, such as aerial photography or radar sensors, have been widely applied for acquiring forest distribution and mapping land-cover patterns (Wulder 1998). Moreover, the use of light detection and ranging (lidar), which provides three-dimensional information of forest characteristics, has significantly increased in the last decade. The measurement operates by emitting pulses from the sensors and determining the elapsed time between the return signals from the target surfaces (Lefsky et al. 2002a). Laser sensors can directly measure the vertical distribution of tree canopies and provide highly accurate estimates of vegetation height, cover, and canopy structure.

There are two major categories of lidar system: waveform with large footprint (8–70 m) and discrete return with small footprint (0.1–0.3 m) (Lim et al. 2003b). Both of these sensor types have been successfully used to estimate forest stand volume and above-ground biomass, as well as other biophysical characteristics such as number of individual trees, tree height, and basal area (e.g., Lefsky et al. 1999a; Means et al. 2000; Næsset and Bjerknes 2001; Persson et al. 2002). In this study, we adopted discrete return lidar.

Studies on the estimation of stand volume or aboveground biomass of broad-leaved forests are limited, although many studies have been conducted for coniferous forests since an early stage (Maclean and Krabill 1986; Nelson et al. 1988). For deciduous broad-leaved forests of eastern Maryland, USA, Lefsky et al. (1999b) estimated the above-ground biomass using full-waveform lidar, SLICER. In the tropical forests in Panama and Costa Rica, Drake et al. (2002, 2003) estimated above-ground biomass by waveform lidar, LVIS. More recently, discrete-return lidar has been applied to several broad-leaved forests in North America (Popescu et al. 2003, 2004; Lim et al. 2003a; Lim and Treitz 2004). Although some experiments proved to be successful, the effective predictors vary from case to case. Therefore, more studies are needed to establish a versatile model for estimating stand volume or above-ground biomass of broad-leaved forests.

Two approaches can be adopted with respect to the estimation using discrete-return lidar: (1)

individual tree based (e.g., Persson et al. 2002; Maltamo et al. 2004; Popescu et al. 2003, 2004) and (2) stand/plot-based (e.g., Nelson et al. 1988; Næsset 1997, 2002; Means et al. 2000; Lim et al. 2003a; Lim and Treitz 2004). In the present trial, the latter approach was applied as it seems difficult to distinguish individual trees in the broad-leaved forest where tree canopies are closed.

Therefore, the present study aimed to explore estimating stand volume of broad-leaved forest by plot-based approach using discrete-return lidar.

## Materials and methods

### *Study site*

Expo'70 Commemorative Park is located in Suita City, Osaka, western Japan (34º47´N, 135º31´E) (Fig. 1), which belongs to the warm-temperate zone, where evergreen broad-leaved (laurel) forest is regarded as the climax vegetation. The park is a part of Senri Hill, which is about 50–130 m above the sea level. The topography of the study site is relatively flat. In 1970, the area was used as the site for the world exposition, Expo'70. After the large-scale reclamation, the site was afforested by the forest restoration project from 1972 to 1976 (Morimoto et al. 2006). Thirty years has passed since the reclamation, and most of the forest stands have canopies of more than 10 m in height. The study site is about 64.5 ha, including both evergreen and deciduous broad-leaved forests, ranging from sparse to dense forests. Forty-four evergreen and 13 deciduous broad-leaved tree species are found at this site.

### *Ground reference data*

Field data were collected in 2004–2006, for a total of 20 plots, which include 14 evergreen and 6 deciduous broadleaved stands. The basic size of the plots was 15 × 15 m, although some plots with different sizes such as 20 × 20 m, 10 × 20 m were established, subject to the forest conditions. Within each plot, diameter at breast height (DBH), tree height, and species of all living woody plants more than or equal to 1 cm in diameter were determined. The location of each plot was obtained by compass survey using a sitemap at 1:500 scale and a compass (Tracon LS-25 Surveying Compass, Ushikata Mfg. Co., Ltd., Japan).

### *Stand volume*

Stand volume in each plot was calculated from the ground measurements for each individual tree. The individual tree volume is considered to be a function of DBH, height, and tree form, however the equation that involves only DBH and height was used for practical reasons in this study. For all species, tree volume equations based on the Timber Volume Table, West Japan edition (Japanese Forest Agency 1970) were used, as follows.

Trees with DBH < 12.0 cm:

$$\mathrm{Log}_{10}V = 1.856641 \log_{10} D + 0.819044 \log_{10} H - 4.070481 \quad (1)$$

Trees with $12.0 \leq \mathrm{DBH} < 22.0$ cm:

$$\mathrm{Log}_{10}V = 1.864235 \log_{10} D + 0.973986 \log_{10} H - 4.232323 \quad (2)$$

Trees with $\mathrm{DBH} \geq 22.0$ cm:

$$\mathrm{Log}_{10}V = 1.752091 \log_{10} D + 1.131128 \log_{10} H - 4.272709 \quad (3)$$

where:

$V$ = tree volume ($m^3$)

$D$ = DBH (cm)

$H$ = tree height (m)

Total plot volume was computed as the sum of the individual tree volumes. The characteristics of the sample plots are presented in Table 1.

Lidar data and multispectral imagery

Lidar data were acquired on October 4, 2004, using an Optech Airborne Laser Terrain Mapper 2050 (Optech Inc., Canada). The aircraft's position was calculated from a GPS receiver at fixed intervals. Each laser-point position was derived from the amplitude peak of the first or last returned pulse and transformed to x, y, z-coordinates in the local coordinate system based on the world geodetic system. A laser beam divergence of 0.19 mrad resulted in a footprint on the ground of approximately 19 cm. The study site was measured from an altitude of 1,000 m above ground level and flight speed was 130 knots; scan mirror frequency of 67.2 Hz, pulsing frequency of 50 kHz, and scan range of ±5.3° gave a scan width of 173 m. The mission was designed with up to 40% sidelap to fix the interpoint distance less than 0.5 m. In addition to the lidar data, high-resolution multispectral images were acquired simultaneously by a digital camera with near-infrared mode. The images consisted of green (510–600 nm), red (600–720 nm), and near-infrared (720–800 nm) bands. At each pixel of the image, the result was recorded as a luminance within the range from 0 to 255. The images were rectified and converted into an ortho-image with pixel array dimensions of 7,124 × 7,246, at a resolution of ca. 18 cm on the ground.

*Extracting variables*

According to the previous studies (Næsset 1997, 2002; Lim et al. 2003a), we considered stand volume estimated from the following model:

$$v = \beta_0 h^{\beta 1} d^{\beta 2} \quad (4)$$

where $v$ is the stand volume, $h$ is a height variable, $d$ is adensity variable, and $ß_0$, $ß_1$, and $ß_2$ are regression coefficients.

In order to extract the variables for the model, the lidar data and digital aerial photograph data were processed by the following procedures. Firstly, ground points were classified by creating digital terrain model (DTM) in TerraScan software (Terrasolid Inc., Finland). Classification of the ground is based on iterative building of a triangulated surface model (Soininen 2003). It starts by selecting local low points and controlling initial point selection with the maximum building size parameter. The triangles in the initial model are mostly below the ground, with only the vertices touching the ground. Then classification starts molding the model upwards by iteratively adding new laser points with terrain angle, iteration distance, and iteration angle. Using these classified ground points, the triangulated surface model was created and exported onto a grid with 1- m spacing. In the triangulated model, the maximum triangle size was 50 m. After classification of the ground points, the rest of the points were classified as vegetation points. The height of each point was calculated as the difference between the altitude of the vegetation points and the altitude of DTM.

Each laser point was also classified into three echo types: (1) first echo: first pulse returns of multiple returns, (2) last echo: last pulse returns of multiple returns, and (3) only echo: single pulse returns when the first and last pulses returned from the same altitude. For each sample plot, five laser height variables were derived from the vegetation points: 25th percentile ($h_{25}$), 50th percentile ($h_{50}$), 75th percentile ($h_{75}$), 100th percentile ($h_{100}$), and mean ($h_{mean}$). Two canopy density variables were derived from the ratio of the number of laser return points. The numbers of the four different types of points used were: (1) number of first echoes ($n_f$), (2) number of only echoes ($n_o$), (3) number of points classified as ground ($n_g$), and (4) number of points classified as vegetation ($n_v$).

Ground fraction ($d_{GF}$) was computed as

$$d_{GF} = n_g / (n_f + n_o) \qquad (5)$$

where $d_{GF}$ represents the ratio of the pulse that reached the ground to the projected lasers. $d_{GF}$ is revealed to be a significant variable for estimating leaf area index (LAI) in previous work (Sasaki et al. 2008). Only-and-vegetation fraction ($d_{OVF}$) was computed as

$$d_{OVF} = n_{o \text{ and } v} / n_v \qquad (6)$$

where $n_{o \text{ and } v}$ is the number of only echoes in vegetation points. $d_{OVF}$ is a newly proposed density-related variable that assumes that the proportion of only echoes increases in forest stand with dense canopies as last echoes hardly penetrate and return from the forest floor. Furthermore, using the red and near-infrared bands of the acquired ortho-image, the normalized difference vegetation index (NDVI) was computed as another density variable ($d_{NDVI}$), because NDVI has been widely applied as an estimator of LAI and vegetation biomass (e.g., Tucker 1979; Gamon et al. 1995; Carlson and Ripley 1997).

$$d_{NDVI} = (NIR - R) / (NIR + R) \quad (7)$$

where NIR and $R$ indicate reflectance in the near-infrared and red wavebands, respectively. The above processing was done in ERmapper 7.1 (Earth Resource Mapping, Australia) and exported into ArcGIS 9.1 (ESRI Japan, Japan). GIS zonal statistical analysis was carried out in order to obtain the mean value of the raster cells within each plot.

*Statistical analysis*

Stand volume was regressed against one of the height variables and/or one of the density variables using the above-described model (Eq. 4). Nonlinear least-square regressions were performed to develop models using the lidar-derived and NDVI parameters. First, each parameter was tested as a single independent variable. Then all the combinations of the height variables and the density variables were examined. Consequently, a total of 23 models were studied.
Leave-one-out cross-validation was then performed. For each cross-validation split, one plot was tested on the predictor model derived from the n - 1 remaining plots. The cross-validated coefficient of determination ($R^2$) and the root-mean-square error (RMSE) were calculated for comparison of the models. All $R^2$ values reported are adjusted for the effects of multiple independent variables. The analyses were done with R 2.7.2 (R Development Core Team, Austria).

**Results**

Stand volume of the 20 sample plots was regressed against the predictor variables. The results from each model are summarized in Table 2. The regression analysis for the models with the height variables alone resulted in adjusted $R^2$ values between 0.377 and 0.730, with RMSE in the range 50.93–90.57 m$^3$ ha$^{-1}$, corresponding to 19.9–35.4% of the mean stand volume. The highest adjusted $R^2$ value was found when $h_{mean}$ was used (adjusted $R^2$ = 0.730), followed by $h_{50}$ (adjusted $R^2$ = 0.714). When estimating by the density variables, none of these performed better than the height variables.

Using combinations of height and density variables, the models basically performed better than the models using the same height or density variable, obtaining adjusted $R^2$ values in the range 0.376–0.755 (RMSE = 41.90–80.05 m$^3$ ha$^{-1}$, corresponding to 16.4–31.2% of the mean stand volume). Among the density variables, $d_{OVF}$ was revealed to be the most effective when combined with the height variables. The best prediction was found when $h_{75}$ and $d_{OVF}$ were used (adjusted $R^2$ = 0.755). The RMSE was 41.90 m$^3$ ha$^{-1}$, corresponding to 16.4% of the mean stand volume. Figure 2 shows a plot of the relationship between the observed and predicted stand volume by the best model.

*Discussion*

This study examined the performance of various variables derived from discrete-return lidar data for stand volume estimation. Height and density variables derived from lidar data similar to in previous studies (e.g., Næsset 1997, 2002; Means et al. 2000; Holmgren 2004) were examined, except for $d_{OVF}$, which was newly proposed in the present study. When those variables were examined separately, while the models using one of the height variables obtained relatively good fit, the results from the models with one of the density variables were not significant. This suggests that use of a height variable is more important than use of a density variable for estimating stand volume.

The highest adjusted $R^2$ value was found when $h_{75}$ and $d_{OVF}$ were used as independent variables. $h_{75}$ is likely to be related to the height of upper canopy layer in each stand (Fig. 3). This is concurrent with other studies that also included upper height percentile in their prediction models, such as 80th or 90th percentile (Means et al. 2000; Næsset 2002; Holmgren 2004). $d_{OVF}$ was found to be more effective than $d_{NDVI}$, which is known as an estimator of LAI or green leaf biomass (e.g., Tucker 1979; Gamon et al. 1995; Carlson and Ripley 1997). Although $d_{GF}$ was effective when combined with $h_{25}$, the mechanism was unclear. As they are new variables, the performance of $d_{OVF}$ and $d_{GF}$ should be examined further.

The next best model in this study was that of $h_{mean}$. Compared with the other height variables considered in this study, $h_{mean}$ seems to have the advantage that it reflects canopy density more than $h_{50}$ or $h_{75}$ (Fig. 3). In closed canopy forests, which usually have high stand volume, lasers tend to hit the top part of the crowns intensively, resulting in higher $h_{mean}$. In sparse-canopy forests, which

usually have less stand volume, laser pulses tend to return from lower vegetation, leading to lower hmean (Fig. 3). The mean height and the derivative variables from mean height were also found to be useful in previous studies of hardwood/deciduous forests (Lefsky et al. 2002b; Popescu et al. 2003; Nelson et al. 2004), although they were not compared with height percentile variables.

The best prediction model ($h_{75}$ and $d_{OVF}$) from this study was better than or comparable to the results in the other studies for broad-leaved forests. Lefsky et al. (1999b) estimated above-ground biomass in deciduous forest of Eastern Maryland, USA, obtaining RMSE of 45.8 Mg ha$^{-1}$, corresponding to 19.2% of the mean aboveground biomass. Popescu et al. (2004) estimated tree volume by individual tree-based approach in deciduous forest in the southeastern USA, having $R^2$ of 0.39 with RMSE value of 52.84 m$^3$ ha$^{-1}$, corresponding to 32.3% of the mean stand volume.

The results in this study demonstrated that the presented model for estimating stand volume from discrete-return lidar data achieved better or comparable prediction in broad-leaved forest compared with previous studies. Further work to refine and validate this approach should be done with different datasets, e.g., on a slope, or for a mixture of coniferous and deciduous trees, as results could be site specific and dependent on each forest condition (Means et al. 2000; Næsset 2004). Clarifying the property of each derived variable is also required to fully understand behavior of estimation models.

**Acknowledgement**

We would like to thank the staff of the Commemorative Organization for the Japan World Exposition '70 for their support. This research was partly financed by the Organization for Landscape and Urban Green Technology Development, Japan.

## References

Carlson TN and Ripley DA (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens Environ 62:241-251

Drake JB, Dubayah RO, Knox RG, Clark DB, Blair JB (2002) Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. Remote Sens Environ 81:378-392

Drake JB, Knox RG, Dubayah RO, Clark DB, Condit R, Blair JB, Hofton M (2003) Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships. Global Ecology and Biogeography 12:147-159

Gamon JA, Field CB, Goulden ML, Griffin KL, Hartley AE, Joel G, Peñuelas J, Valentini R (1995) Relationships between NDVI, canopy structure, and photosynthesis in three California vegetation types. Ecological Applications 5:28-41

Holmgren J (2004) Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. Scand J For Res 19:543-553

Japanese Forest Agency (1970) Timber volume table, West Japan edition. J-FIC, Tokyo, pp 63 (in Japanese)

Lefsky MA, Cohen WB, Acker SA, Parker GG, Spies TA, Harding D (1999a) Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. Remote Sens Environ 70:339-361

Lefsky MA, Harding D, Cohen WB, Parker G, Shugart HH (1999b) Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. Remote Sens Environ 67:83-98

Lefsky MA, Cohen WB, Parker GG, Harding DJ (2002a) Lidar remote sensing for ecosystem studies. BioScience 52:19-30

Lefsky MA, Cohen WB, Harding DJ, Parker GG, Acker SA, Gower ST (2002b) Lidar remote sensing of above-ground biomass in three biomes. Global Ecology and Biogeography 11:393-399

Lim K, Treitz P, Baldwin K, Morrison I, Green J (2003a) Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. Can J Remote Sensing 29:658-678

Lim K, Treitz P, Wulder M, St-Onge B, Flood M (2003b) LiDAR remote sensing of forest structure. Progress in Physical Geography 27:88-106

Lim K and Treitz PM (2004) Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand J For Res 19:558-570

Maclean GA and Krabill WB (1986) Gross-merchantable timber volume estimation using an airborne lidar system. Can J For Res 12:7-18

Maltamo M, Eerikäinen K, Pitkänen J, Hyyppä J, Vehmas M (2004) Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distributions functions. Remote Sens Environ 90:319-330

Means JF, Acker SA, Brandon JF, Renslow M, Emerson L, and Hendrix CJ (2000). Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering & Remote Sensing 66:1367-1371

Morimoto Y, Njoroge JB, Nakamura A, Sasaki T, Chihara Y (2006) Role of the EXPO '70 forest project in forest restoration in urban areas. Landscape Ecol Eng 2:119-127

Nelson R, Krabill W, Tonelli J (1988) Estimating forest biomass and volume using airborne laser data. Remote Sens Environ 24:247-267

Nelson R, Short A, Valenti M (2004) Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. Scand J For Res 19:500-511

Næsset E (1997) Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens Environ 61:246-253

Næsset E & Bjerknes KO (2001) Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sens Environ 78:328-340

Næsset E (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens Environ 80:88-99

Persson Å, Holmgren J, Sôderman U (2002) Detecting and measuring individual trees using airborne laser scanner. Photogrammetric Engineering & Remote Sensing 68:925-932

Popescu SC, Wynne RH, Nelson RF (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can J Remote Sensing 29:564-577

Popescu SC, Wynne RH, Scrivani JA (2004) Fusion of small-footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia, USA. Forest Science 50:551-565

Satake Y, Hara H, Watari S, Tominari T (1989) Wild flowers of Japan: Woody plants I, II. Heibunsya, Tokyo (in Japanese)

Sasaki T, Imanishi J, Ioki K, Morimoto Y, Kitada K (2008) Estimation of leaf area index and canopy openness in broad-leaved forest using an airborne laser scanner in comparison with high-resolution near-infrared digital photography. Landscape Ecol Eng 4:47-55

Soininen A (2003) TerraScan User’s Guide

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation.

Remote Sens Environ 8:127-150

Wulder M (1998) Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Progress in Physical Geography 22:449-476