Modeling Crown Structure from LiDAR Data with Statistical Distributions

Quang V. Cao and Thomas J. Dean

Abstract: The objective of this study was to evaluate the ability of three statistical distributions to characterize the vertical distribution of foliage mass in canopies of even-aged loblolly pine stands, based on airborne-scanning light detection and ranging (LiDAR) data. The functions were the Weibull and S_B distributions and a mixture of the lognormal and Weibull distributions. Results indicated that the mixture distribution fit the LiDAR data better than the Weibull and S_B distributions, according to three goodness-of-fit statistics. By switching from the lognormal for data near the tree top to the Weibull for data near the base of the crown, the mixture distribution seemed to better resemble the shape of the distribution of LiDAR returns than the Weibull alone. The mixture distribution was particularly effective for plots in which distributions of the returns were too peaked for either the Weibull or S_B. FOR. SCI. 57(5):359–364.

Keywords: canopy, foliage distribution, mixture, Weibull, lognormal, S_B, loblolly pine

THE VERTICAL DISTRIBUTION OF FOLIAGE affects photosynthesis (Wang and Jarvis 1990, Stenberg et al. 1994, Monsi and Saeki 2005), stem taper and size (Larson 1963, Dean and Long 1986, West et al. 1989, Dean et al. 2002), and habitat diversity (MacArthur and MacArthur 1961, Brokaw and Lent 1999). The Weibull distribution has been widely used to describe vertical foliage distributions (Gillespie et al. 1994, Maguire and Bennett 1996, Baldwin et al. 1997, Xu and Harrington 1998). Whereas one end of the Weibull function (the top of the live crown) starts at zero, the other end of the Weibull function (the bottom of the live crown) could, in theory, extend to infinity. To avoid this drawback, Jerez et al. (2005) selected Johnson's S_B function (Johnson 1949), and Maguire and Bennett (1996) used the beta distribution because these distributions possess fixed end points that naturally correspond to the top and the bottom of the live crown. Modeling the vertical distribution of foliage can result in reasonable predictions of the bottom of the crown and the median of the distribution, which represents the center of gravity of foliage. These two attributes may be used with other variables to predict tree taper based on the uniform stress theory (Dean and Long 1986). Finally, yield per unit area can be computed from volume of an "average tree" in a plot, based on predicted values of its total height and dbh.

Light detection and ranging (LiDAR) is an airborne-based, scanning laser technique that can be used to describe the average vertical canopy distribution of a large area. Magnussen and Boudewyn (1998) have shown that the vertical distribution of laser pulses reflected from the canopy matches the upper portion of the vertical distribution of leaf area in Douglas-fir (*Pseudotsuga menziesii*) canopies. Discrete-return LiDAR systems produce three-dimensional data consisting of elevation or "z" values for each horizontally referenced point. Stand and individual tree characteristics have been related to statistical metrics of the vertical distribution of the z values (Andersen et al. 2005, Hall et al. 2005). The spatial pattern of the z values can be used to measure height and tree count (Naesset 2002, Popescu et al. 2002, McCombs et al. 2003, Roberts et al. 2005) and crown diameter (Popescu et al. 2003, Roberts et al. 2005). In a recent study, Dean et al. (2009) used the Weibull distribution to characterize foliage distribution of a loblolly pine (*Pinus taeda* L.) stand based on the LiDAR z values. The objective of this study was to evaluate the performance of three statistical distributions, the Weibull and S_B distributions and a mixture of the lognormal and Weibull distributions, in characterizing the vertical distribution of foliage mass in canopies of even-aged loblolly pine stands, based on airborne-scanning LiDAR data.

Data

LiDAR data were collected from 24 rectangular plots (28 m $\times$ 52 m) in a stand in J.G. Lee Memorial Forest, located in southeastern Louisiana near the city of Pine in Washington Parish. The stand was a 36-year-old, naturally regenerated stand of nearly pure loblolly pine. LiDAR data were collected using an Optech ALTM-1225 laser mapping system operating at 25 kHz and a scanning angle of $\pm 9°$ from an altitude of 610 m and airspeed of 62 m/s. The minimum posting density was 4 laser pulses/m^2. Horizontal position, z value, and intensity were recorded for the first and last return of each outgoing pulse, but intensity values and last returns were not used in this study.

The z values of the vegetation returns were converted to

Quang V. Cao, Louisiana State University Agricultural Center, School of Renewable Natural Resources, Baton Rouge, LA 70803—Phone: (225) 578-4218; Fax: (225) 578-4227; qcao@lsu.edu. Thomas J Dean, Louisiana State University Agricultural Center—fwdean@lsu.edu.

Acknowledgments: We thank Minyi Zhou, Charles K. Beatty, Jennifer Smith, Kevin Stilley, Frank Ueltschi, Joe Nehlig, and Keith Hawkins for their help in the field; Lee Mitchell for processing the LiDAR data; and Drs. D. Evans and S. Roberts for their collaboration. Funding for this project was provided by a cooperative agreement with the US Forest Service, Southern Research Station (SRS 01-CA-11330122-510), Louisiana Department of Agriculture and Forestry's Louisiana Forest Productivity Program (CFMS #577795/Proj #01-FPP1), Mississippi State University's Remote Sensing Technology Center, and McIntire-Stennis funds.

Manuscript received June 3, 2010, accepted December 1, 2010

height above the ground surface by subtracting the ground elevation from the vegetation elevation for each position. The relative distance (x) from the maximum z value (z_{max}) was computed from

$$x = 1 - (z/z_{max}), \quad 0 \le x \le 1. \tag{1}$$

At the base of the crown, the origin of the z values becomes confounded with the stem and midstory vegetation. Furthermore, the relative frequency of LiDAR returns will never approach zero below the canopy because of reflections from the stem and lesser vegetation (Figure 1). The LiDAR data were right-truncated at a truncation point (t), which does not correspond to the base of the canopy but was assumed to represent the lowest return that could be reliably attributed to reflections from the canopy. The value of t was determined by fitting a series of straight lines through successive groups of two data points moving right, starting from the mode of the value of x. These data points consisted of the relative return frequencies within 40 elevation classes for each plot, each elevation class interval equaling 2.5% of the highest z value in the plot. When the slope increased more than 50%, the truncation point was set to the middle of the last two elevation classes.

Statistical Distributions

Three statistical distributions were used in this study to characterize the vertical distribution of foliage based on LiDAR data: the Weibull and S_B probability density functions (PDFs) and a mixture of the lognormal and Weibull distributions.

The Weibull Distribution

A two-parameter Weibull PDF (Bailey and Dell 1973) was fit to the x values for each plot:

$$f_w(x) = (c/b)(x/b)^{c-1} \exp[-(x/b)^c], \quad x > 0, \tag{2}$$

where b and c are scale and shape parameters, respectively. Because the data were truncated at t, maximum likelihood estimates for the Weibull parameters were obtained for each plot by maximizing the log-likelihood function, $\ln(L)$, of the right-truncated Weibull PDF,

$$\ln(L) = \sum \ln[f_w(x_i)] - n \ln\{1 - \exp[-(t/b)^c]\}, \tag{3}$$

where x_i is relative distance of the ith z value return from the maximum value of z in that plot, and $\ln(\cdot)$ is the natural logarithm. The summation sign includes values of i ranging from 1 to n, where n is the number of x_i values that satisfy $x_i < t$.

The S_B Distribution

The S_B function (Johnson 1949) is a transformation from the standard normal distribution and has the following PDF:

$$f_s(x) = \frac{\lambda\delta}{x(\lambda - x)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{x}{\lambda - x}\right)\right)^2\right], \tag{4}$$

where γ and δ are shape parameters and λ is the range or the maximum value of x. Similar to that for the Weibull, the maximum likelihood estimates for the above three parameters were obtained for each plot by maximizing

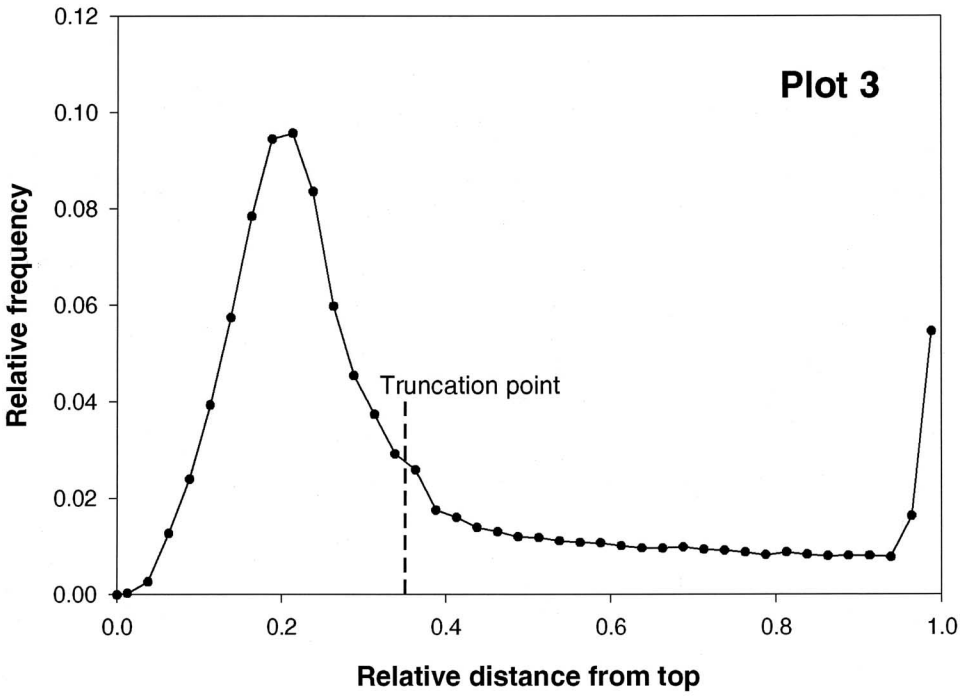

Figure 1. Distribution of the LiDAR relative frequency for an example plot. The truncation point was selected to represent the lowest return that could be reliably attributed to reflections from the canopy.

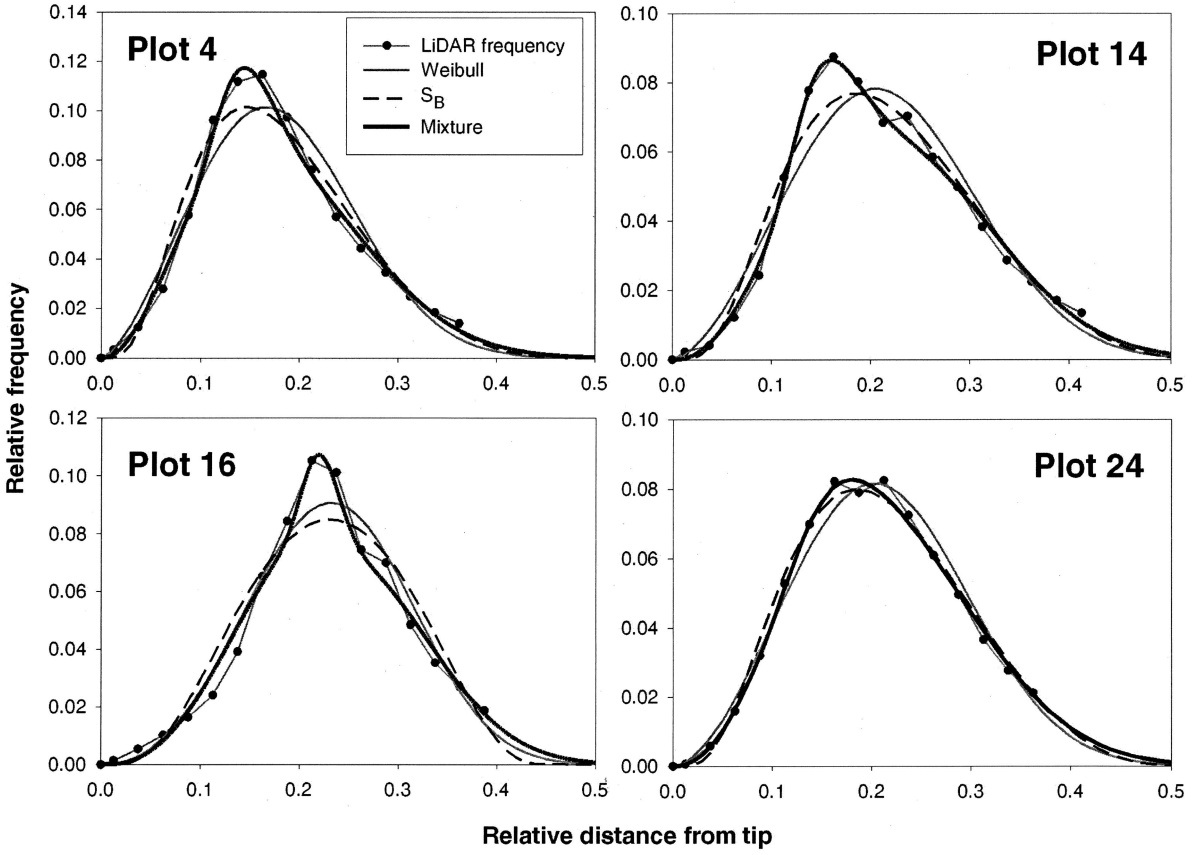

Figure 2. Graphs of the three statistical distributions (Weibull, S_B, and mixture) that fit the frequency of z values from LiDAR in the upper canopies for four example plots.

the log-likelihood function of the S_B PDF, right truncated at t:

$$\ln(L) = \sum \ln[f_s(x_i)] - \Phi\left(\gamma + \delta \ln\left(\frac{t}{\lambda - t}\right)\right), \qquad (5)$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution.

The Mixture Distribution

The lognormal distribution fit the LiDAR data better than the Weibull distribution for small x values but did not do as well for large x values, because it has a heavy right tail that approaches the x-axis very slowly. The result was an overprediction of frequency for large x values. Cao and Wu (2007) faced a similar problem when they modeled wood fiber length and attempted to solve it by combining lognormal and Weibull distributions. The resulting mixture distribution has the following cdf:

$$F_M(x) = [1 - w(x)]F_L(x) + w(x)F_W(x), \quad x > 0, \qquad (6)$$

where $F_W(x) = 1 - \exp[-(x/b)^c]$ = Weibull cdf with parameters b and c, $F_L(x) = \Phi[(\ln(x) - \mu)/\sigma]$ = lognormal cdf with parameters μ and σ, and $w(x) = 1 - \exp(-b_w x)$ = a weight function.

The mixture distribution is a weighted average of the lognormal and Weibull distributions with weight $w(x)$ ranging from 0 to 1. When x is close to zero, $w(x)$ approaches zero, and the mixture distribution resembles the lognormal distribution near the tree top. As x increases, $w(x)$ changes its value from 0 to 1, causing the mixture distribution to gradually change from the lognormal to the Weibull. We arbitrarily set $w(t) = 0.99$ to ensure that the mixture distribution approaches the Weibull distribution at the truncation point (near the base of the canopy), thus avoiding the possibility of a heavy right tail due to considerable weight attributed to the lognormal. This constraint results in

$$b_w = -\ln(0.01)/t. \qquad (7)$$

The PDF of the mixture distribution has the form

$$f_M(x) = [1 - w(x)]f_L(x) + w(x)f_W(x) + w'(x)[F_W(x) - F_L(x)], \quad x > 0, \qquad (8)$$

where $w'(x) = b_w \exp(-b_w x)$, $f_W(x)$ = Weibull PDF as defined in Equation 1,

$$f_L(x) = \left(\frac{1}{x\sigma\sqrt{2\pi}}\right)\exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right]$$

= lognormal pdf.

The mixture distribution has four parameters, μ, σ, b, and c, which were obtained for each plot by maximizing the

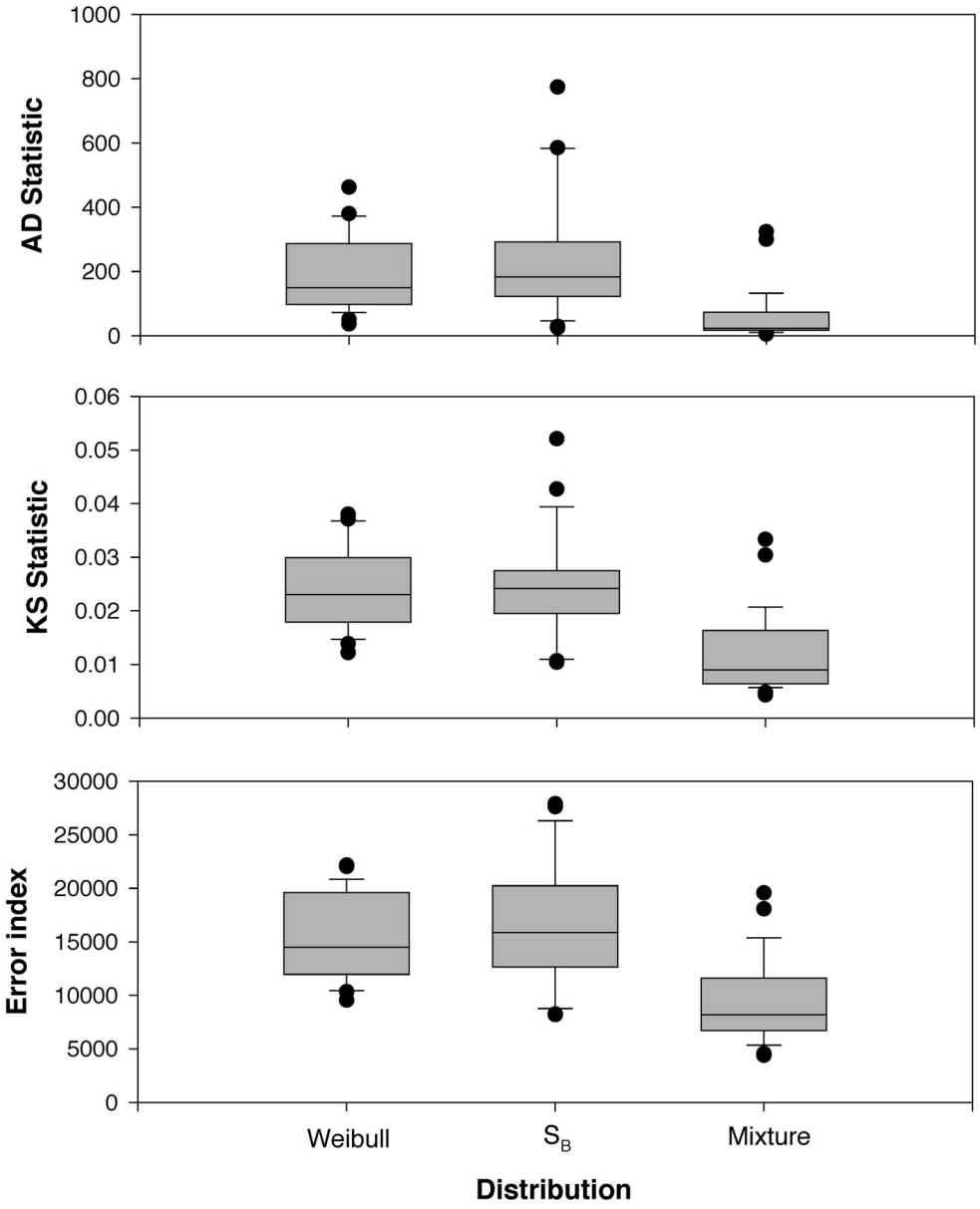

Figure 3. Box plots show the median, 25th and 75th percentiles (lower and upper extent of box), and values outside the 5th and 95th percentiles of the three goodness-of-fit statistics for the Weibull, S_B, and mixture distributions.

log-likelihood function of the right-truncated mixture PDF:

$$\ln(L) = \sum \ln[f_M(x_i)] - n \ln[0.01F_L(t) + 0.99F_W(t)]. \quad (9)$$

An attempt to fit the LiDAR data with a mixture of two S_B distributions produced inconsistent results. The S_B mixture fit the data well for some plots but poorly for others.

Evaluation

Three goodness-of-fit statistics were computed for each distribution and for each plot. The Anderson-Darling (AD) statistic (Anderson and Darling 1954) for the *i*th plot was calculated as

$$\mathrm{AD}_i = -n_i - \sum_{j=1}^{n_i} (2j-1)[\ln(u_j) + \ln(1 - u_{n-j+1})]/n_i, \quad (10)$$

where $u_j = F(x_j)$ = one of the three cdfs evaluated at x_j, x_j are values of x, sorted in ascending order for each plot ($x_1 \leq x_2 \leq \leq x_{n_i}$), and n_i is the number of observations in the *i*th plot.

The one-sample Kolmogorov-Smirnov (KS) statistic for the *i*th plot was computed as

$$\mathrm{KS_i} = \max\{\max_{1 \leq i \leq n_i}[(j/n_i) - u_j], \max_{l \leq i \leq n_i}[u_j - (j-1)/n_i]\}. \quad (11)$$

The error index (EI) of Reynolds et al. (1988) for the *i*th plot was calculated as

$$\mathrm{EI}_i = \sum |n_{ik} - \hat{n}_{ik}|, \quad (12)$$

where n_{ik} and $\hat{n}_{ik}$ are observed and predicted numbers of LiDAR returns in elevation class k for the *i*th plot, and the sum includes all elevation classes in the *i*th plot.

Results and Discussion

Figure 2 shows that the mixture distribution fit the relative frequency of LiDAR data as well as or better than did the Weibull and S_B distributions for these example plots. All three distributions performed reasonably well in representing the relative frequency of LiDAR returns up to the point at which the data were truncated.

Overall, the mixture distribution clearly fit the data better than either the Weibull or S_B distribution, producing consistently lower values for the three goodness-of-fit statistics (Figure 3). Duncan's multiple range tests revealed that the mixture distribution was significantly different ($P < 0.01$) from the Weibull and S_B distributions, based on the mean values of the AD statistic, KS statistic, and error index, respectively. On average, the Weibull and S_B distributions produced an approximately 260% higher AD value, 107% higher KS value, and 69% higher error index value than the mixture distribution.

It should be noted that the increase in number of parameters from two (the Weibull) to three (the S_B) did not result in a better fit. In fact, the Weibull distribution did produce slightly lower mean values of the three goodness-of-fit statistics than the S_B, even if the difference was not significant. The shape of the Weibull distribution therefore might be more suitable for the LiDAR frequency than for that of the S_B.

The mixture distribution has the most parameters (four) among the three distributions evaluated in this study. By switching from the lognormal for small x to the Weibull for large x values, the mixture distribution seemed to better resemble the shape of the distribution of LiDAR returns than the Weibull alone. The mixture distribution was particularly effective for plots in which distributions of the x values were too peaked for either the Weibull or S_B (Figure 2).

Conclusions

This study demonstrates that the vertical distribution of foliage, as represented by laser pulses reflected from the canopy, can be successfully characterized by statistical distributions such as the Weibull, S_B, or a mixture of lognormal and Weibull distributions. All three distributions fit reasonably well the relative frequency of first LiDAR returns up to the point where the data were truncated. The mixture distribution, in particular, did the best job in capturing the shape of the distribution of LiDAR z values.

Literature Cited

ANDERSEN, H.E., R.J. MCGAUGHEY, AND S.E. REUTEBUCH. 2005. Estimating forest canopy fuel parameters using LiDAR data. *Remote Sens. Environ.* 94:441–449.

ANDERSON, T.W., AND D.A. DARLING. 1954. A test of goodness of fit. *J. Am. Statist. Assoc.* 49:765–769.

BAILEY, R.L., AND T.R. DELL. 1973. Quantifying diameter distributions with the Weibull function. *For. Sci.* 19:97–104.

BALDWIN, V.C. JR., K.D. PETERSON, H.E. BURKHART, R.L. AMATEIS, AND P.M. DOUGHERTY. 1997. Equations for estimating loblolly pine branch and foliage weight and surface area distributions. *Can. J. For. Res.* 27:918–927.

BROKAW, N., AND T. LENT. 1999. Vertical structure. P. 373–399 in *Maintaining biodiversity in forest ecosystems*, Hunter, M.J. (ed.). Cambridge University Press, Cambridge, UK.

CAO, Q.V., AND Q. WU. 2007. Characterizing wood fiber and particle length with a mixture distribution and a segmented distribution. *Holzforschung* 61:124–130.

DEAN, T.J., Q.V. CAO, S.D. ROBERTS, AND D.L. EVANS. 2009. Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. *For. Ecol. Manag.* 257:126–133.

DEAN, T.J., AND J.N. LONG. 1986. Validity of constant-stress and elastic instability principles of stem formation in *Pinus contorta* and *Trifolium pratense*. *Ann. Bot.* 58:833–840.

DEAN, T.J., S.D. ROBERTS, D.W. GILMORE, D.A. MAGUIRE, J.N. LONG, K.L. O'HARA, AND R.S. SEYMOUR. 2002. An evaluation of the uniform stress hypothesis based on stem geometry in selected North American conifers. *Trees* 16: 559–568.

GILLESPIE, A.R., H.L. ALLEN, AND J.M. VOSE. 1994. Amount and vertical distribution of foliage of young loblolly pine trees as affected by canopy position and silvicultural treatment. *Can. J. For. Res.* 24:1337–1344.

HALL, S.A., I.C. BURKE, D.O. BOX, M.R. KAUFMANN, AND J.M. STOKER. 2005. Estimating stand structure using discrete-return lidar: An example from low density, fire prone ponderosa pine forests. *For. Ecol. Manag.* 208:189–209.

JEREZ, M., T.J. DEAN, Q.V. CAO, AND S.D. ROBERTS. 2005. Describing leaf area distribution in loblolly pine trees with Johnson's S_B function. *For. Sci.* 51:93–101.

JOHNSON, N.L. 1949. Systems of frequency curves generated methods of translation. *Biometrika* 36:149–176.

LARSON, P.R. 1963. *Stem form development of forest trees.* For. Sci. Monograph 5. Society of American Foresters, Bethesda, MD. 42 p.

MACARTHUR, R.H., AND J.W. MACARTHUR. 1961. On bird species diversity. *Ecology* 42:594–598.

MAGNUSSEN, S., AND P. BOUDEWYN. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Can. J. For. Res.* 28:1016–1031.

MAGUIRE, D.A., AND W.S. BENNETT. 1996. Patterns in vertical distribution of foliage in young coastal Douglas-fir. *Can. J. For. Res.* 26:1991–2005.

MCCOMBS, J.W., S.D. ROBERTS, AND D.L. EVANS. 2003. Influence of fusing LiDAR and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *For. Sci.* 49: 457–466.

MONSI, M., AND T. SAEKI. 2005. On the factor fight in plant communities and its importance for matter production. *Ann. Bot.* 95:549–567.

NAESSET, 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.

POPESCU, S.C., R.H. WYNNE, AND R.F. NELSON. 2002. Estimating plot-level tree heights with LiDAR: Local filtering with a canopy-height based variable window size. *Comput. Electron. Agric.* 37:71–95.

POPESCU, S.C., R.H. WYNNE, AND R.F. NELSON. 2003. Measuring individual tree crown diameter with LiDAR and assessing its influence on estimating forest volume and biomass. *Can. J. Remote Sens.* 29:564–577.

REYNOLDS, M.R., T.E. BURK, AND W.C. HUANG. 1988. Goodness

of-fit tests and model selection procedures for diameter distribution models. *For. Sci.* 34:373–399.

ROBERTS, S.D., T.J. DEAN, D.L. EVANS, J.W. MCCOMBS, R.L. HARRINGTON, AND P.A. GLASS. 2005. Estimating individual tree leaf area in loblolly pine plantations using LiDAR-derivedmeasurements of height and crown dimensions. *For. Ecol. Manag.* 213:54–70.

STENBERG, P., T. KUULUVAINEN, S. KELLOMÄKI, E.J. JOKELA, AND H.L. GHOLZ. 1994. Crown structure, light interception, and productivity of pine trees and stands. P. 20–34 in *Environmental constraints on the structure and productivity of pine forests ecosystems: A comparative analysis*, Gholz, H.L., S. Linder, and R.E. McMurtrie (eds.). Ecol. Bull. 43. Munksgaard International Booksellers and Publishers, Copenhagen, Denmark.

WEST, P.W., D.R. JACKETT, AND S.J. SYKES. 1989. Stresses in, and the shape of, tree stems in forest monoculture. *J. Theor. Biol.* 140:327–343.

WANG, Y.P., AND P.G. JARVIS. 1990. Influence of crown structural properties on PAR absorption, photosynthesis, and transpiration in Sitka spruce-application of a model (MAESTRO). *Tree Physiol.* 7:297–316.

XU, M.G., AND T.B. HARRINGTON. 1998. Foliage biomass distribution of loblolly pine as affected by tree dominance, crown size, and stand characteristics. *Can. J. For. Res.* 28: 887–892.