

University of New Hampshire

University of New Hampshire Scholars' Repository

Master's Theses and Capstones

Student Scholarship

Fall 2012

An investigation of two methods for assessing the vertical structure of forest stands

Daniel S. Maynard

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/thesis>

Recommended Citation

Maynard, Daniel S., "An investigation of two methods for assessing the vertical structure of forest stands" (2012). *Master's Theses and Capstones*. 740.

<https://scholars.unh.edu/thesis/740>

This Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Master's Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact Scholarly.Communication@unh.edu.

**AN INVESTIGATION OF TWO METHODS FOR ASSESSING THE
VERTICAL STRUCTURE OF FOREST STANDS**

BY

DANIEL S. MAYNARD

B.S., University of New Hampshire, 2003

M.A., Harvard University, 2006

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Master of Science

in

Natural Resources

September, 2012

UMI Number: 1521563

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1521563

Published by ProQuest LLC 2012. Copyright in the Dissertation held by the Author.

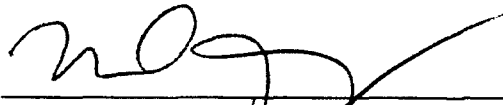
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.



Thesis Director, Mark J. Ducey
Professor of Forest Biometrics and Management



Co - Thesis Director, Russell G. Congalton
Professor of Remote Sensing and Geographic
Information Systems



Joel N. Hartter
Assistant Professor of Geography

July 27, 2012
Date

Acknowledgements

I would like to sincerely thank my committee members: Dr. Mark Ducey, for all of his guidance and for providing me with two exciting research ideas to pursue; Dr. Russell Congalton, for his invaluable assistance in all aspects of graduate life; and Dr. Joel Hartter, for his advice and encouragement.

I would also like to extend my gratitude to my friends, family, and fellow graduate students. In particular, I would like to thank Mickey Campbell for his help and friendship throughout the past two years, and my wife, Rachel, for her unending support.

This research was funded by the Department of Natural Resources and the Environment at the University of New Hampshire through the teaching assistantship program, and by the Disaster Resilience for Rural Communities Program through the National Institute of Food and Agriculture within the U.S. Department of Agriculture (Award #2010-67023-21705).

Contents

Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Abstract	viii

Chapter	Page
Introduction	1
1 Vertical Point Sampling with a Digital Camera: Slope Correction and Field Evaluation	4
1.1 Vertical Point Sampling	5
1.1.1 The Height Squared Factor with a Circular Inclusion Area	9
1.1.2 Derivation of the Slope Expansion Factor	11
1.1.3 Adjusting for Slope	16
1.2 Field Study: Bias Assessment	18
1.2.1 Study Design	19
1.2.2 Statistical Analysis	21
1.2.3 Results	24
1.2.4 Discussion	28
1.2.5 Conclusions	31

2	The Relationship between Survival Analysis and Point Quadrat Sampling: Theory and Applications	32
2.1	Point Quadrat Sampling	34
2.1.1	MacArthur and Horn Estimator	35
2.1.2	Laser Point Sampling - Radtke and Bolstad Estimator	37
2.2	Survival Analysis	38
2.2.1	Equivalence of Point Quadrat Sampling and Survival Analysis	39
2.3	Regression Analysis of Point Quadrat Data	41
2.3.1	Study Design	42
2.3.2	Statistical Analysis	45
2.3.3	Results	50
2.3.4	Discussion	57
2.3.5	Conclusions	62
	Conclusions	63
	Works Cited	65

List of Tables

1.1	The inclusion area expansion factor	16
1.2	Stand-level attributes across the thirty-three sample points	24
1.3	Correlation between individuals' photo counts	25
1.4	Average tree count per photo for slope-adjusted and unadjusted data	26
1.5	Results of the mixed-effects regression models	27
1.6	Results of the mixed-effects ANOVA model	28
2.1	Summary statistics of the 55 forest plots	51
2.2	Results from the Weibull regression models	56

List of Figures

1.1	Tallying trees in vertical point sampling	6
1.2	The geometry of a photograph	7
1.3	The inclusion area for a tree on flat terrain	8
1.4	Identifying tree tops in vertical photographs	9
1.5	The inclusion area for a tree on sloped terrain	12
1.6	The geometry of the sloped and projected inclusion area	13
1.7	Top-down view of the projected inclusion ellipse	14
1.8	Pawtuckaway State Park	19
1.9	Histogram of the distribution of slopes	25
2.1	Oregon Study Area	43
2.2	Sampling Design Layout	44
2.3	Density functions for the Weibull distribution	45
2.4	The estimated canopy profiles using the Weibull regression model	53
2.5	Parametric and non-parametric canopy profiles for six sites	54
2.6	Parametric and non-parametric LAI estimates	55

ABSTRACT

AN INVESTIGATION OF TWO METHODS FOR ASSESSING THE VERTICAL STRUCTURE OF FOREST STANDS

by

Daniel S. Maynard

University of New Hampshire, September, 2012

In this study we investigate the limitations of two methods for assessing forest structure: vertical point sampling with a camera and laser point quadrat sampling. Vertical point sampling with a camera is a method by which the height squared per unit area of a forest can be quickly estimated. First, we derive the bias incurred for failing to adjust for slope when implementing this sampling method, and we show that slope can generally be ignored as long as the majority of sample points occur on slopes less than 35 degrees. In the second part of this study we outline the equivalence between survival analysis methods and laser point quadrat analysis methods. We use a survival-based parametric regression model to analyze laser point quadrat data and estimate canopy structure and density. The results show that survival analysis techniques can yield improved results over traditional non-parametric point quadrat analysis methods.

Introduction

The vertical structure of a forest encompasses the vertical distribution of all above-ground foliage and woody material within the stand. Typically, the term “vertical structure” is used to refer solely to specific canopy characteristics, such as canopy density, or to the distribution of tree heights across the stand. In addition, some measurements of forest structure encompass both a horizontal and vertical dimension (e.g., percent crown closure or crown radius), and so the vertical structure of a forest naturally includes both horizontal and vertical components.

Vertical structure is an important indicator of forest productivity and function. As photosynthesis is carried out in the leaves of a tree, canopy characteristics are strong indicators of individual tree growth, vigor, and overall stand productivity (Waring et al., 1981; Mitchell et al., 1983; Waring, 1983; Maguire and Kanaskie, 2002). Canopies are responsible for energy production, carbon sequestration, and water transfer (Ford and Deans, 1978; Hollinger, 1989; Baldocchi et al., 2002; Baldocchi and Harley, 2006). They regulate the quantity of light and water that reach the forest floor, thereby influencing temperature, humidity, and moisture (Seidel et al., 2011). These effects can alter understory vegetation patterns and corresponding wildlife populations (Hayes et al., 1997; DeMaynadier and Hunter Jr., 1999; Jennings et al., 1999; Martens et al., 2000; Wilson et al., 2007).

Due to these direct and indirect effects on forest structure and function, a thorough understanding of the the vertical structure of forests is integral for making responsible and effective stand-level management decisions. For example, knowledge of differing aspects of vertical structure is necessary for preventing wildfires, maintaining wildlife habitat, managing timber, mitigating invasion by insects or exotic plants, and for aesthetics or recreation. Overall, the accurate assessment of vertical structure is needed to quantify the health,

productivity, habitat suitability, growth, and ecosystem processes of forested ecosystems.

There are many different methods for assessing vertical structure, including (but not limited to) ground-based measurements (e.g., the use of a clinometer to measure tree heights), photographic techniques (e.g., hemispherical photography to measure gap fraction), stratified clipping and scaffolding to directly measure canopy density and structure, and optical methods such as terrestrial LIDAR. While each of these methods is useful in specific settings, most methods of assessing vertical structure are subject to one or more limitations: they can be expensive, time-consuming, difficult to implement, or they can yield data that are imprecise or difficult to interpret.

The purpose of this study is to address some of these difficulties by expanding upon two existing methods of assessing vertical structure: (1) vertical point sampling, and (2) point quadrat sampling. Each of these methods has been used with varying success in the past, though recent improvements in methodology by Ducey and Kershaw (2011) and Radtke and Bolstad (2001) have resulted in simpler and more efficient techniques for implementing these sampling procedures. By addressing the current limitations in both of these methods, these sampling procedures have the potential to be quick and efficient methods of assessing different aspects of forest structure with minimal additional investment in time or resources.

Vertical point sampling is a method by which trees are sampled with probability proportional to their squared height, resulting in an estimate of total squared-height per unit area (Hirata, 1955; Grosenbaugh, 1958). Recently, Ducey and Kershaw (2011) proposed a technique of conducting vertical point sampling using a digital camera. This new method eliminates many of the problems associated with the traditional field-implementation of vertical point sampling. However, a question that remains with this new sampling method is the degree to which sloping terrain biases the resulting estimate. The first chapter of this paper addresses this issue by deriving the relationship between the inclusion area for a tree on flat terrain and the corresponding inclusion area for that tree on sloping terrain. A method to adjust for slope is provided, along with the results of a field study conducted to investigate the bias incurred for ignoring sloping terrain.

Point quadrat sampling is a method by which canopy structure and density can

be measured from ground-based measurements. This method, first employed by Levy and Madden (1933) and MacArthur and Horn (1969), was modified by Radtke and Bolstad (2001) to use a laser rangefinder to obtain canopy-height measurements. The second chapter of this study addresses the current limitations of laser point quadrat sampling by outlining how point quadrat sampling is analogous to the field of survival analysis as commonly used in biomedical studies. The equivalence between these two fields allows for a wide range of preexisting survival analysis methods to be used in the analysis of point quadrat data. The correspondence between these two fields is outlined, and a field example is provided to illustrate the power of using existing survival analysis techniques to analyze laser point quadrat data. This relationship has the potential to drastically reduce the number of samples needed to obtain accurate estimates of canopy density and structure.

Chapter 1

Vertical Point Sampling with a Digital Camera: Slope Correction and Field Evaluation

Horizontal point sampling (HPS) is a form of probability-proportional-to-size sampling, where the probability that each tree is tallied is directly proportional to the squared-diameter of that tree (Bitterlich, 1948; Grosenbaugh, 1958). This method, also known as variable-radius plot sampling or prism cruising, is a common tool used to estimate various stand-level attributes due to its simplicity and efficiency. An analogous method, termed vertical point sampling, is a technique by which trees are sampled with probability proportional to their squared height (Hirata, 1955; Grosenbaugh, 1958). In contrast to HPS, vertical point sampling (VPS) has rarely been used in practice, primarily due to the difficulty implementing this procedure and to the perceived lack of usefulness of the resulting estimate of “height squared per unit area.”

Recently, Ducey and Kershaw (2011) developed a method by which vertical point sampling can be quickly conducted using a digital camera. This new method alleviates many of the difficulties that arise when carrying out VPS using the traditional methods. Additionally, the authors show that the resulting estimates of squared-height correlate strongly with various stand level attributes that are often time-intensive and difficult to obtain, including biomass, cubic volume, and stand-density-index. The relationship between squared-height and these additional variables makes VPS a potentially useful tool in ratio estimation or as part of a double-sampling scheme.

A remaining issue when conducting vertical point sampling with a camera (VPSC) is the effect of sloping terrain on the estimate of height squared per unit area. In VPSC, a

vertical photo of the canopy is taken at a randomly selected sample point and the number of tree tops appearing in the photo is tallied. On sloping terrain, the trees located uphill from the sample point appear “taller” in the resulting image, while downhill trees appear “shorter.” This distortion leads to an expanded inclusion area for trees on sloping terrain that increases the probability that these trees are sampled.

In this study, we investigate two methods of accounting for sloping terrain when conducting VPSC. First, we present a method by which the sampling procedure can be altered to directly adjust for slope at each sample point, thus yielding unbiased estimates. Secondly, we conducted a field study to assess the bias incurred for ignoring slope altogether. By combining these results, we are able to provide general recommendations as to when the slope-adjustment procedure should be followed, and when sloping terrain can be ignored with minimal impact on the resulting estimates.

1.1 Vertical Point Sampling

In vertical point sampling, the radius of the inclusion area of each tree is selected to be a fixed proportion of its height, rather than a fixed proportion of its diameter (Hirata, 1955). This situation results in the inclusion area of each tree being proportional to the height squared of the tree (Grosenbaugh, 1958). A “Height Squared Factor” (HSF) is selected to control the number of trees counted per sample point (analogous to the basal area factor, or BAF, when conducting horizontal point sampling). This HSF determines the size of the sampling area for each tree: a larger HSF means that trees have smaller inclusion circles and are sampled less often; a smaller HSF means that trees have larger inclusion areas and are sampled more often.

To conduct vertical point sampling, first a random sample point must be selected within the stand (typically by way of a map or a GIS). Once at the point, all of the trees with an inclusion area that “overlaps” that sample point are tallied (Figure 1.1). The total number of trees tallied at that sample point is then multiplied by the pre-determined HSF, resulting in an estimate of “height-squared per unit area.” Similar to horizontal prism sampling, this procedure can be conducted at a series of points throughout the stand to

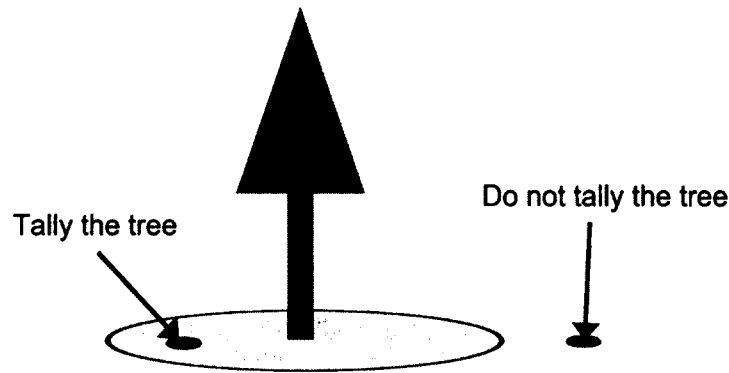


Figure 1.1: Each tree has a circular inclusion area surrounding it that is a function of the height of the tree and the pre-determined HSF. If the randomly selected sample point falls within this inclusion area, that tree is tallied. If the sample point falls outside of the tree's inclusion area, that tree is not tallied.

obtain an average estimate.

In practice, VPS has seen very little use. This is largely due to the fact that the traditional method of determining whether or not each tree should be tallied is time consuming relative to the usefulness of the information (Ducey and Kershaw, 2011). Either a clinometer or a relascope can be employed in a similar manner to the horizontal angle gauge (Bitterlich, 1984), but estimating the location of the top of a tree can be exceedingly difficult in the field, particularly in dense stands or in the presence of foliage. Additionally, squared-height is generally of little interest to foresters as a primary measurement.

To address the difficulty in implementing this technique in the field, Ducey and Kershaw (2011) recently suggested using a digital camera to conduct vertical point sampling. Their proposed method involves:

1. Taking a single vertical photograph of the canopy at a randomly selected sample point.
2. Counting the number of trees tops that appear in the photo.
3. Multiplying this count by the HSF to obtain an estimate of squared-height per unit area.

As the HSF is dictated by the geometry of the image, overlaying a smaller rectangle on the image and counting the tree tops that appear only within this smaller rectangle corresponds to using a larger HSF. In order to simplify the geometry of this situation, the

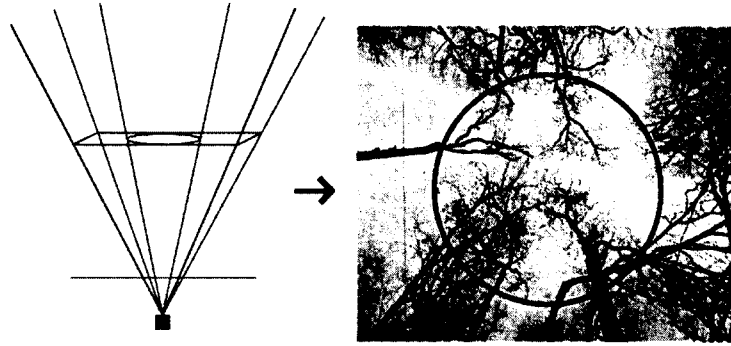


Figure 1.2: The geometry of a photograph: drawing a circle on the image is equivalent to projecting a cone upwards from the camera.

method used throughout this study alters this procedure by digitally overlaying a circle on the photograph, and by counting the number of tree tops located within this circle. In this setting, a smaller superimposed circle corresponds to a larger HSF, and a larger circle corresponds to a smaller HSF. By changing the size of this circle that is being superimposed on the image, the user can adjust the expected number of trees counter per point. In very dense stands there may be dozens of tree tops visible in the image, so a small circle (large HSF) would be used to avoid spending large amounts of time counting each image. Conversely, in very sparse stands only a few tree might be visible in each photograph, and so a large circle (small HSF) would be used to ensure that at least a few trees are counted at each sample point.

To understand how this photo-counting methods succeeds in implementing VPS, note that the vertical projection of a rectangular photograph can be visualized as a 4-sided inverted pyramid. By drawing a circle on the resulting image, we are projecting a cone up through the image (Figure 1.2). If the top of a tree is visible within the superimposed circle, then this is an indicator that the sample point falls within the circular inclusion area of that tree, and so the tree should be tallied. Under this method, the outside boundary of a given tree's inclusion area is the collection of points where the top of a tree is at the outer edge of the cone (Figure 1.3). For all such points, the distance from the camera to the tree is equal to $h \cdot \tan \theta$, where θ is half of the opening angle of the cone and h is the height of the tree. Thus, the inclusion area is a circle with radius proportional to the height of the tree, and with area proportional to the squared height.

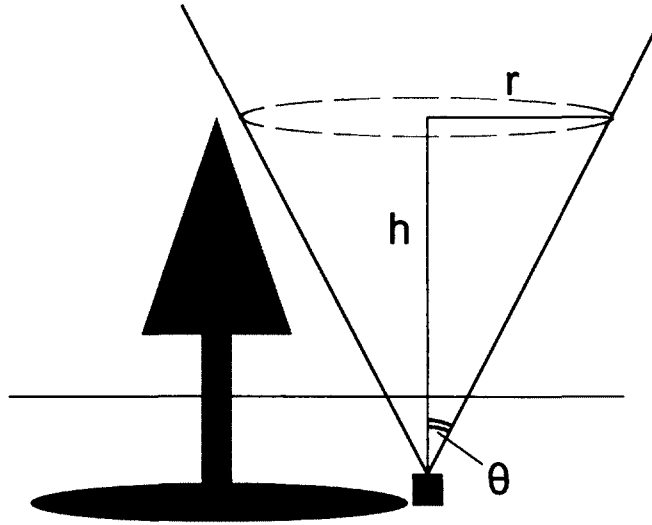


Figure 1.3: On flat terrain, the inclusion area for a tree of height h (indicated by the shaded circle) is equal to the set of points through which the top of the tree passes through the cone projected upwards by the camera. The radius of this inclusion circle can be defined in terms of the height of the tree and the angle θ , equal to half the opening angle of the cone.

The overriding benefit of VPSC is that it is simple and quick to take a photograph at each sample point, and the counting of tree tops generally takes less than 30 seconds per photograph. This makes it an attractive tool in double sampling, ratio, and regression estimation, where the primary variable of interest is time consuming or expensive to accurately estimate. In such a setting, photographs can be taken at every sample point, with detailed measurements taken at a subset of these points.

An important assumption with all vertical point sampling methods is that each tree has a single well-defined top that can be readily identified. In certain stands and for certain tree species this assumption is justified, though in practice it can often be difficult to identify the exact location of the “top” of the tree due to the presence of foliage or to decurrent growth forms (Figure 1.4). The benefit of VPSC compared to other VPS methods is that the photos can be scored by multiple individuals and stored for later verification, thus potentially reducing the bias incurred by inaccurate estimation from any one individual.

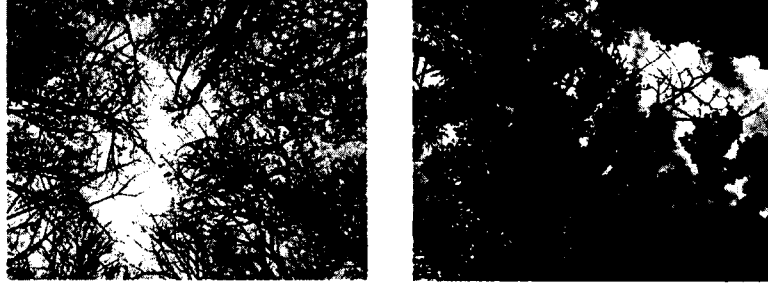


Figure 1.4: Identifying the tops of individual trees in vertical photographs can be difficult in dense stands, in the presence of foliage, or if the trees have decurrent growth forms.

1.1.1 The Height Squared Factor with a Circular Inclusion Area

Before addressing the effect of sloping terrain on a tree's inclusion area, it is necessary to illustrate how this inclusion area is defined for flat terrain. When a tree is located on a slope of zero degrees, the inclusion area for this tree is a circle, equal in area to the cross-section of the cone at the point where the top of the tree intersects the cone (Figure 1.3). The outline of this circular inclusion area can be visualized by moving the camera in a circle around the tree such that the top of the tree remains at the exact boundary of the cone. By rewriting the radius (r) of this inclusion area in terms of the height (h) of the tree and the angle (θ) that defines the cone, the inclusion area is:

$$A_{flat} = \pi(h \tan \theta)^2 \quad . \quad (1.1)$$

If a single sample point is selected in a sampling region of area \tilde{A} , the probability that this tree is included (counted) is equal to this above inclusion area divided by the total area. To estimate the total squared height per unit area (τ), we can use a Horvitz-Thompson estimator (Horvitz and Thompson, 1952), where the summation is over the v trees counted in the photo at that point:

$$\begin{aligned} \tau &= \sum_i^v \frac{h_i^2}{\pi(h_i \tan \theta)^2 / \tilde{A}} \\ &= \frac{\tilde{A}}{\pi \tan^2 \theta} \cdot v \quad . \end{aligned} \quad (1.2)$$

Thus, we have that the HSF depends only upon the angle that defines the cone,

and on the total sampling area. For practical purposes, it is usually easiest to divide by the total area in order to obtain the estimate in terms of squared-height per unit area. This avoids unnecessarily large numbers and having to know the actual stand area. This results in a circular Height Squared Factor of:

$$HSF = \frac{1}{\pi \tan^2 \theta} \quad . \quad (1.3)$$

Ducey and Kershaw (2011) show that the minimum HSF of a rectangular inclusion area for a given camera can be easily reconstructed by placing the camera some known distance (D) from a wall, attaching a tape measure to the wall at the same height of the camera, and using this to measure the length (L) of the image. The width (W) of the image is then equal to this length divided by the aspect ratio of the image (assuming that the aspect ratio is written in terms of L/W).

A similar process can be used to obtain the minimum HSF for a given camera when using a circular inclusion zone. First, follow the above prescription given by Ducey and Kershaw (2011) to obtain estimates of W and D . If a circle is drawn on the photo such that the diameter equals the full width of the image, then we have that:

$$\tan \theta = \frac{W/2}{D} = \frac{W}{2 \cdot D} \quad , \quad (1.4)$$

which, when combined with Equation 1.3, yields the minimum HSF for that particular camera (i.e., the HSF that is achieved by overlaying a circle with diameter exactly equal to the width of the image):

$$HSF = \frac{4 \cdot D^2}{\pi W^2} \quad . \quad (1.5)$$

If a larger HSF is required (i.e., a smaller circle on the photograph), we can solve Equation 1.5 for a new width, W_{new} . By taking the fraction of this new width to the original width we get:

$$F = \frac{W_{new}}{W_{original}} = \sqrt{\frac{HSF_{original}}{HSF_{new}}} \quad . \quad (1.6)$$

The diameter of the new superimposed circle that corresponds to the desired HSF is simply

equal to the full width of the image times this fraction F . That is:

$$\text{New Circle Diameter} = F \cdot (\text{Original Circle Diameter}) \quad . \quad (1.7)$$

For example, suppose the original image width is 6 inches with a corresponding HSF of 1. If a HSF of 4 is desired, then we get that $F = \sqrt{1/4} = 1/2$. Therefore, a superimposed circle with a diameter of $F \times 6 \text{ in.} = 1/2 \times 6 \text{ in.} = 3 \text{ in.}$ will yield the desired HSF.

1.1.2 Derivation of the Slope Expansion Factor

When a tree is on a slope, the point at which the top of the tree passes through the camera-cone remains parallel to the slope and is tilted from the horizontal. The total area along the slope that the tree is visible within the cone is therefore equal to the intersection of a tilted plane and a cone. This results in an ellipse that is stretched both vertically along the slope and horizontally along the contour (Figure 1.5). Since sample points are generally selected using a flat projection of the study area, the inclusion area for any given tree is the projection of this sloped ellipse onto the 2-dimensional surface. The following derivations and methodology reflects this projected sampling scheme, though the results are easily adjusted for the case when the sampling method involves the complete 3-dimensional surface (see equation 1.17 and the following discussion on page 16 for a further explanation).

Here, we derive the ratio of the projected inclusion area to the original inclusion area on flat terrain. In this section, the term “unprojected” refers to any characteristic of the inclusion area that lies along the slope, while the term “projected” refers to any characteristic of the inclusion area that has been projected back down into the 2-dimensional surface.

Consider a tree of height h on a slope of A degrees (Figure 1.6). Suppose a HSF is used such that θ is half of the opening angle of the cone (see equation 1.3), and the radius of the inclusion area is given by L . Let $C = 90 - \theta$, $B = C - A$, $D = 180 - C$, and $E = 180 - C - A$. Let X_1 and X_2 be the two segments that define the length of the sloped inclusion area. It follows that the interior angles of the two triangles bounded by the cone, the original inclusion radius, and the new inclusion length can be written in terms of A , B ,

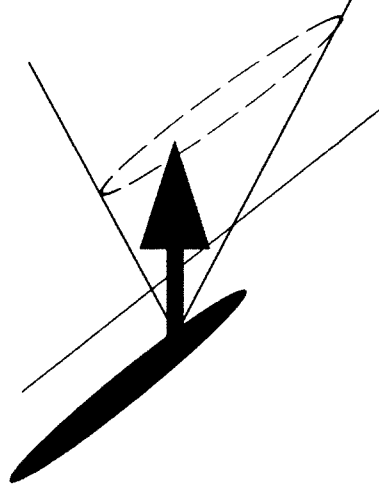


Figure 1.5: On sloped terrain, the set of all points along the slope for which the top of the tree passes through the cone forms an ellipse that is stretched both along the slope and along the contour (indicated by the shaded circle). The projection of this sloped ellipse onto the 2-dimensional surface likewise results in an ellipse.

C , D , and E (Figure 1.6). Using the law of sines, we have that:

$$\frac{\sin E}{L/2} = \frac{\sin C}{X_1} \Rightarrow X_1 = \frac{\sin C}{\sin E} \cdot \frac{L}{2} \quad , \quad (1.8)$$

and:

$$\frac{\sin B}{L/2} = \frac{\sin D}{X_2} \Rightarrow X_2 = \frac{\sin D}{\sin B} \cdot \frac{L}{2} \quad . \quad (1.9)$$

Let P_1 and P_2 be the segments that correspond to the horizontal projections of X_1 and X_2 , respectively.

Then from equations 1.8 and 1.9 it follows that:

$$\begin{aligned} P_1 &= X_1 \cdot \cos(A) \\ &= \frac{\sin C}{\sin E} \cdot \cos A \cdot \frac{L}{2} \\ &= S_1 \cdot \frac{L}{2} \quad , \end{aligned} \quad (1.10)$$

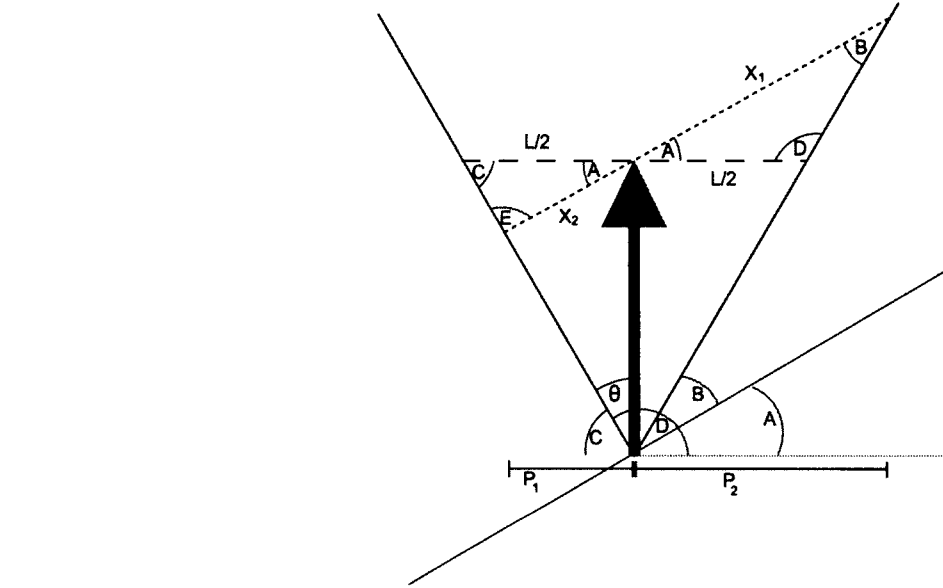


Figure 1.6: The geometry of the major radius of the projected inclusion ellipse for a tree on a slope a A degrees. The sum of lengths X_1 and X_2 gives the length of the major radius of the sloped elliptical inclusion area. The horizontal projection of these lengths is given by P_1 and P_2 , the sum of which equals the length of the major radius of the projected elliptical inclusion area. These lengths can be written in terms of the angles θ , B , C , D , E , and the original inclusion radius, L .

and:

$$\begin{aligned}
 P_2 &= X_2 \cdot \cos(A) \\
 &= \frac{\sin D}{\sin B} \cdot \cos A \cdot \frac{L}{2} \\
 &= S_2 \cdot \frac{L}{2} \quad , \quad (1.11)
 \end{aligned}$$

where:

$$S_1 = \frac{\sin C}{\sin E} \cdot \cos A \quad (1.12)$$

$$S_2 = \frac{\sin D}{\sin B} \cdot \cos A \quad . \quad (1.13)$$

As the intersection of a cone and a plane forms an ellipse, and the projection of an ellipse onto a single axis remains an ellipse, we have that the 2-dimensional projected inclusion area is likewise an ellipse, with the length along the major axis being $P_1 + P_2$. Note that P_1 corresponds to the distance from the tree to the projected lower edge of the

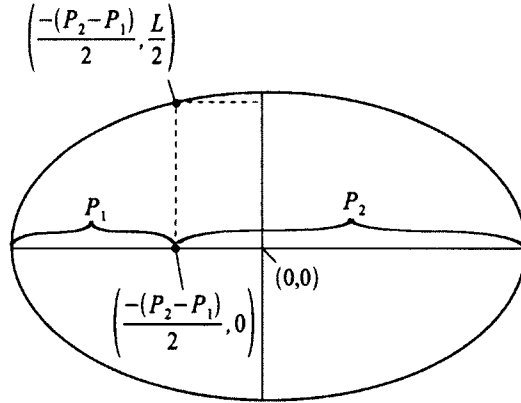


Figure 1.7: The top-down view of the projected inclusion ellipse with major radius equal to $P_1 + P_2$. The tree is projected downward to the point $(-\frac{P_2 - P_1}{2}, 0)$, and the width of the ellipse at this point is equal to the original radius of the circle, L . This gives a point along the outer edge of the ellipse as $(-\frac{P_2 - P_1}{2}, \frac{L}{2})$.

inclusion zone, and P_2 corresponds to the distance from the tree to the projected upper edge of the inclusion zone. If we change coordinates so that the center of the projected inclusion area is at the center of the X-Y coordinate plane, then it follows that the tree is located at the point $(x, y) = (-\frac{P_2 - P_1}{2}, 0)$ (Figure 1.7).

Also note that the the original inclusion circle had a diameter of L . The sloped inclusion area can be viewed as the intersection of the cone and of a plane that has been horizontally rotated about the base of the tree. Thus, the width of the sloped inclusion area at the tree remains L . The projection of this sloped ellipse into 2-dimensions does not not affect the width of the inclusion area (only the length along the slope), and thus the width of the projected ellipse at the tree is likewise L .

For an intuitive understanding of this, suppose you are standing next to a tree exactly along the contour and you take a vertical photo. As the camera and the tree are located at the same vertical position along the slope, the height of that tree in the resulting image does not appear “taller” or “shorter” than it otherwise would on flat terrain. The tree appears in the camera frame exactly as it would if it were on level ground. Thus the width of the inclusion area of the tree along the contour does not change, and remains L .

Combining these two results, we know that at the outer edge of the inclusion area

must pass through the point (Figure 1.7):

$$(x, y) = \left(\frac{-(P_2 - P_1)}{2}, \frac{L}{2} \right) = \left(\frac{-(S_2 - S_1)}{2} \cdot \frac{L}{2}, \frac{L}{2} \right) \quad (1.14)$$

Noting that the major radius, a , of the projected ellipse is equal to $\frac{S_1 + S_2}{2} \cdot \frac{L}{2}$, we can substitute the above point into the equation of an ellipse, and solve for the minor radius, b :

$$\begin{aligned} 1 &= \frac{x^2}{a^2} + \frac{y^2}{b^2} \\ \Rightarrow 1 &= \frac{\left(\frac{S_2 - S_1}{2} \right)^2 \left(\frac{L}{2} \right)^2}{\left(\frac{S_1 + S_2}{2} \right)^2 \cdot \left(\frac{L}{2} \right)^2} + \frac{\left(\frac{L}{2} \right)^2}{b^2} \\ \Rightarrow b &= \frac{L}{2} \cdot \left(\frac{S_1 + S_2}{2\sqrt{S_1 \cdot S_2}} \right) \quad (1.15) \end{aligned}$$

From this, we can calculate the projected elliptical inclusion area for this tree:

$$A_{slope} = \pi ab = \pi \left[\frac{L}{2} \cdot \left(\frac{S_1 + S_2}{2} \right) \right] \left[\frac{L}{2} \cdot \left(\frac{S_1 + S_2}{2\sqrt{S_1 \cdot S_2}} \right) \right] = \pi \left(\frac{L}{2} \right)^2 \cdot \frac{(S_1 + S_2)^2}{4\sqrt{S_1 \cdot S_2}} \quad (1.16)$$

Since the inclusion area along flat terrain is $A_{flat} = \pi(L/2)^2$, we have that the slope expansion factor, α , is:

$$\alpha = \frac{A_{slope}}{A_{flat}} = \frac{(S_1 + S_2)^2}{4\sqrt{S_1 \cdot S_2}} = \frac{(\tilde{S}_1 + \tilde{S}_2)^2}{4\sqrt{\tilde{S}_1 \cdot \tilde{S}_2}} \cdot \cos A \quad , \quad (1.17)$$

where:

$$\tilde{S}_1 = \frac{\sin C}{\sin E} \quad \text{and} \quad \tilde{S}_2 = \frac{\sin D}{\sin B} \quad (1.18)$$

An important property of this expansion factor is that it does not depend on the height of the tree, h , or on the height-dependent inclusion radius, L . Therefore, every tree

Table 1.1: The inclusion area expansion factor (expressed as a percent), for various height squared factors and slopes.

HSF	1	2	3	4	5	6	7
Field of view [†]	59	43	36	32	28	26	24
Slope (degrees)							
0	0	0	0	0	0	0	0
10	2	1	0	0	0	0	0
20	7	3	2	2	1	1	1
30	18	9	6	4	3	3	2
40	46	20	12	9	7	6	5
50	146	47	28	20	15	12	11

[†] Refers to the angular field of view (i.e. 2θ) of the cone in degrees.

located on the same uniform slope has an inclusion area that is expanded by this same factor. Another convenient property of this function is the isolation of the term $\cos(A)$. If one is interested in the unprojected inclusion area along the slope, rather than the projected inclusion area, then the unprojected expansion factor is equal to Equation 1.17 with the $\cos(A)$ term removed.

The behavior of the expansion percentage (the expansion factor α expressed as a percent) is displayed in Table 1.1 for various HSFs and slopes. In this table, the “Field of View” corresponds to the angular field of view of the cone, i.e., 2θ . We see that for relatively moderate HSFs above 3, the increase in the inclusion area on sloping terrain is less than 10% until slopes of 30-40 degrees are reached. For lower HSFs, slopes of 20 degrees or less likewise result in an increase in the inclusion area of approximately 10% or less.

1.1.3 Adjusting for Slope

There will likely be some instances where a low HSF is required, such as in very open stands or where a large proportion of sample points fall on steep terrain. In these settings, it will be beneficial to adjust for slope in order to maintain accurate estimates. The most straightforward way to adjust for this expanded inclusion area is to employ a variable-radius circle technique, where the size of the superimposed circle depends on the severity of the slope. In this setting, photos taken on steeper slopes will have a smaller circle superimposed

on them, so that, for any given tree, the stretched and projected inclusion area corresponding to this smaller circle is equal in area to its original circular inclusion area on flat terrain. While the shapes of these inclusion areas will be different, all that is required to maintain design unbiased estimates is that their inclusion areas are identical.

More specifically, if θ is the angle that corresponds to the desired HSF on flat terrain, we need to select a smaller angle, θ' , such that the projected inclusion area on sloping terrain defined by this narrower angle is equal to the original circular inclusion area on flat terrain. Specifically, for a given tree of height h , we need to solve the following equality for θ' (the use of the prime denotes a variable related to the narrower cone and smaller superimposed circle):

$$\begin{aligned}
 A'_{slope} &= A_{flat} \\
 \Rightarrow \alpha' \pi h^2 \tan^2 \theta' &= \pi h^2 \tan^2 \theta \\
 \Rightarrow \alpha' \cdot \tan^2(\theta') - \tan^2(\theta) &= 0 \quad . \quad (1.19)
 \end{aligned}$$

Since α' in Equation 1.19 is also a function of θ' , this equation cannot be solved in closed form. It can, however, be numerically solved without difficulty in most statistical or mathematical software packages.

An important aspect of equation 1.19 is that it is independent of tree height, h . Thus, overlaying a circle that corresponds to a single angle θ' results in the correct projected inclusion area for every tree on that slope, regardless of tree height. This new value of θ' can be substituted into Equations 1.6 and 1.3 to obtain the diameter of a new superimposed circle that adjusts for slope and results in all trees at that point having the correct inclusion area.

Given the above results, the variable-radius circle procedure for correctly adjusting for slope is as follows:

1. Calculate the minimum HSF of the camera using Equation 1.4.
2. Select a desired HSF and calculate the angle θ using equation 1.3.
3. For those sample points on flat terrain, superimpose a circle on the vertical pho-

tographs with diameter calculated using equation 1.7. Count the tree tops in each photo.

4. For each sample point on sloping terrain, solve Equation 1.19 for the new angle θ' , and calculate the new HSF and corresponding image circle radius using the same method.
5. Superimpose this smaller circle on the vertical photograph, and count the number of tree tops in this smaller superimposed circle.

For two trees of the same height, one on flat terrain and one on a slope, this procedure results in the same inclusion area for both. Therefore it maintains the requirements of sampling with probability proportional to squared-height, and will result in design-unbiased estimates.

1.2 Field Study: Bias Assessment

Since vertical point sampling with a camera is a quick and simple procedure, it will likely be most useful in those settings where an individual or a group is interested in rapidly assessing different elements of forest structure with minimal investment in resources or equipment. For this reason, the variable-radius procedure outlined above might effectively undermine the usefulness of this technique, as it involves access to and knowledge of statistical or mathematical computer software.

To address this concern, a field study was conducted to quantify the bias incurred for ignoring slope altogether. Table 1.1 indicates that most HSF/slope combinations will have some non-zero bias, but that in many settings this bias should be less than 2 or 3%. For example, on slopes less than 20 degrees with a HSF of 5 or greater, bias should be negligible. Conversely, on slopes greater than 40 degrees with a HSF of 4 or less, bias will likely be substantial.

The goals of this field study were threefold: (1) quantify the bias in the photo counts for different values of the slope expansion factor; (2) identify a slope expansion factor threshold, below which bias will be negligible in real-world settings; and (3) compare how different sources of variability impact this bias. These results will indicate those HSF/slope

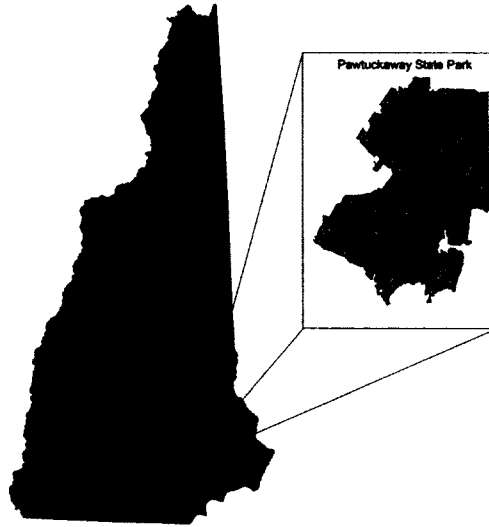


Figure 1.8: The location of Pawtuckaway State Park in southeastern New Hampshire.

combinations where one should ensure they directly adjust for slope, and those HSF/slope combinations where one can ignore slope and still maintain relatively unbiased estimates of height-squared.

1.2.1 Study Design

The study took place at Pawtuckaway State Park in Nottingham, NH during November of 2010. This park is located in southeastern New Hampshire (Figure 1.8), and it encompasses approximately 2200 ha of land, with the Pawtuckaway Mountains forming the center of the park. These mountains are the remnants of a volcanic ring-dike formed approximately 275 millions years ago. The resulting igneous rocks weathered more slowly than the rock in the valleys, leading to relatively drastic terrain differences across a small spatial extent. Glacier activity approximately 13,000 years ago further altered the shape and structure of these mountains, resulting in a large number of rocky outcroppings, boulder fields, and vertical cliff-faces. Across the park, elevation ranges from approximately 70m in the valleys to 335m at the mountain peaks, with slopes from zero to almost 90 degrees. Due to this park having a large number of hiking trails and access roads, these slopes are all readily accessible.

The majority of the park was cleared for farming in the late 1700s, but much of area was abandoned in the late 1800s and early 1900s due to the rocky terrain and poor soil.

In 1961, the state acquired the land, and since 1967 then it has been open to the public as a multi-use state park. Portions of the park are managed by the New Hampshire Fish and Game Department, the New Hampshire Division of Forests and Lands, and the New Hampshire Division of Parks and Recreation.

This study area is part of the northeast coastal forest ecoregion, and is primarily comprised of mixed-oak forests dominated by red oak (*Quercus rubra*), white pine (*Pinus strobus*), American beech (*Fagus grandifolia*), and eastern hemlock (*Tsuga canadensis*). The majority of the land within the park has been largely unmanaged since abandonment in the early 20th century, except when necessary to provide specific habitat or wildlife requirements. The NH Division of Forests and Lands conducts additional period timber sales across small portions of the park.

Within the study area, thirty-three sample points were subjectively selected in order to obtain a full range of slopes. The areas surrounding these sample points showed no visible indications of recent timber harvests or forest management. At each sample point, a vertical photograph was taken, the slope was measured, and a prism sweep was conducted using a metric BAF of 4.6 m²/ha (20 ft²/acre). The camera was held by hand 1m above the ground and it was equipped with a bubble level to ensure that it was held vertically. Diameter at breast height (DBH) was measured for each tree that was tallied using the prism sweep, and the height of every fourth one of these trees was determined with a TruPulse 200 Laser Rangefinder.

The resulting 33 digital photographs were scored by a total of ten participants. These participants were all graduate students in the Department of Natural Resources and the Environment at the University of New Hampshire. Of these ten individuals, four had a background in forestry, five had a background in wildlife ecology, and one had a background in wetland biology. None of these individuals had previously participated in any ground-based photographic methods for estimating forest structure.

Each individual was shown the same image twice: once with the image overlaid with the unadjusted circle corresponding to the nominal HSF, and once with the image overlaid with a variable-radius circle (using the above slope adjustment) that correctly adjusted for the slope at that sample point. The individuals were given no training or practice

beforehand, and they were not made aware of the study objectives or of any element of the study design. They were instructed simply to count the number of tree tops within each circle on each photograph, and to subjectively differentiate between tree branches and tree tops to the best of their ability. The photos were displayed in a random order to each participant so as to avoid any trend-bias in the counting process. This process resulted in 66 photos scored by each participant, and 660 photos scored in total.

The overlaying of the circles on the photos and the numerical analysis for equation 1.19 was carried out in Matlab (R2010a, 2010). The width of the lines of these overlaid circles was set equal to one pixel. This small line width minimized the number of border-line trees (i.e., those tree tops that fell exactly on the line and could not be visually distinguished between being “in” or “out”). The digital photos were displayed in color, with on-screen dimensions roughly equivalent to 4”x6”. The participants were not able to zoom in or move the image on the screen, and they were prompted to enter the number of trees in a text box on the screen.

For this study, a HSF of 1 was selected to allow for adequate assessment of the relative impact of difference expansion percentages. As the bias in this sampling method is a function of the slope expansion percentage, selecting a HSF with a wide range of slope expansion percentages will enable us to investigate the expected bias across a broad range of slope and HSF combinations. Table 1.1 shows that the expansion percentages for a HSF of 1 attain a relatively uniform distribution of values from 2% to 46% as the slope increases from ten to forty degrees, whereas HSFs larger than 1 have a more limited range of expansion percentages. Thus, this relatively small HSF allowed us to investigate the effect of a wide range of expansion factors on the photo counts. Furthermore, a HSF larger than 1 should almost always be used in practice, and so these results also serve as a conservative estimate of the upper limit of the effect of slope.

1.2.2 Statistical Analysis

The approach taken here to analyze the effect of various sources of variability in the photo counts is motivated by the analytical methods employed by Ringvall and Stahl (1999). Here, a mixed-effects model was utilized to investigate the overall effect of slope on the count in

each photo. The outcome of interest was the count on the i^{th} photo (of 66 total photos) by the k^{th} participant on the j^{th} plot. The specific model was:

$$\begin{aligned} \Psi(Count_{ijk}) = & \rho_0 + \rho_1 \cdot \overline{\text{tree height}}_j + \rho_2 \cdot \text{slope}_i \\ & + \rho_3 \cdot \text{plot}_j + \rho_4 \cdot \text{person}_k + \rho_5 \cdot \text{plot}_j \times \text{person}_k + \epsilon_{ijk} \quad , \end{aligned} \quad (1.20)$$

where $\overline{\text{tree height}}_j$ is the average tree height for the j^{th} plot (calculated using a ratio estimator along with DBH), and slope_i is a categorical variable with:

$$\text{slope}_i = \begin{cases} 0 & \text{all slope-adjusted photos} \\ 1 & \text{unadjusted photos with } 0^\circ \leq \text{slope} < 10^\circ \\ 2 & \text{unadjusted photos with } 10^\circ \leq \text{slope} < 20^\circ \\ 3 & \text{unadjusted photos with } 20^\circ \leq \text{slope} < 30^\circ \\ 4 & \text{unadjusted photos with } 30^\circ \leq \text{slope} < 40^\circ \\ 5 & \text{unadjusted photos with } 40^\circ \leq \text{slope} \quad . \end{cases} \quad (1.21)$$

In this model, the factor $\exp(\rho_2)$ corresponds to the multiplicative increase in the count incurred for failing to adjust for each 10 degree increment in slope.

While this model results in an intuitive interpretation of the effect of slope, it assumes a linear effect of slope between 10 degree slope categories. This can lead to underestimates or overestimates of the effect of slope in any single 10 degree category if the true relationship between slope and the photo count is nonlinear. In order to allow for a more flexible effect of slope, a second model was investigated with the slope variable included as a series of indicator variables:

$$\begin{aligned} \Psi(Count_{ijk}) = & \beta_0 + \beta_1 \cdot \overline{\text{tree height}}_j + \beta_2 \cdot I_i(0, 10) + \beta_3 \cdot I_i(10, 20) \\ & + \beta_5 \cdot I_i(20, 30) + \beta_6 \cdot I_i(30, 40) + \beta_7 \cdot I_i(40, \infty) \\ & + \beta_8 \cdot \text{plot}_j + \beta_9 \cdot \text{person}_k + \beta_{10} \cdot \text{plot}_j \times \text{person}_k + \epsilon_{ijk} \quad , \end{aligned} \quad (1.22)$$

where $I_s(X, Y) = 1$ when $X < \text{slope}_i \leq Y$ for the unadjusted photos, and equals zero

otherwise. For the adjusted photos, all indicator variables were set equal to zero. In this model, each of the factors $\exp(\beta_2), \dots, \exp(\beta_7)$ corresponds to the relative increase in the photo count incurred for failing to account for slopes of 0-10, 10-20, 20-30, 30-40, and >40 degrees, respectively.

In both models, slope and average tree height per plot were included as fixed effects. Plot, person, and the interaction of person \times plot were included as random effects. The rationale for including plot and person as random effects is that the sample points and the participants were selected non-randomly, so there is no interest in making population-level inferences about their effect on the count; rather, they are included as random effects to account for the variability they add to the count. The variable corresponding to average tree height was normalized to give a mean of zero and standard deviation of one. The reported coefficients for average tree height therefore correspond to an increase of one standard deviation from the mean.

As the count data were roughly Poisson distributed, $\Psi(\cdot)$ was modeled using the quasi-Poisson log-link function. The quasi-Poisson model differs from a Poisson model only in that it does not assume the variance exactly equals the mean, and so it allows for greater flexibility in the model-fitting process (i.e., it allows for overdispersion). Approximate *P*-values and 95% confidence intervals are reported for each covariate, though it should be noted that these statistics are not exact for mixed-effects models.

A mixed-effects analysis of variance (ANOVA) model was fit to the data in order to compare the relative contribution of each component to the overall variability in the log of the photo counts. Average tree height, slope, person, and plot were included as main sources of variation, and the random effects of plot, person, and plot \times person were modeled via the error term. As the counts were Poisson distributed, the square root of the counts were approximately normally distributed. To test the appropriateness of this assumption, residual plots and Q-Q plots were examined, indicating no violations of independent, normally distributed, and heteroscedastic errors. For the ANOVA model, tests of significance were conducted using *F*-tests, and corresponding *P*-values are reported.

Estimates of squared-height per unit area were calculated for each person for both the unadjusted and slope-adjusted photos, and a paired Student's *t*-test was used to compare

Table 1.2: Stand-level attributes across the thirty-three sample points.

	Mean	S.D.	Range
Quadratic Mean Diameter (m)	0.31	0.11	(0.11, 0.51)
Height [†] (m)	20.2	8.2	(10.0, 46.8)
Basal Area (m ² /ha)	30.1	9.9	(18.4, 55.1)
Trees Per Hectare	658.5	762.0	(120, 3767)

[†] Calculated using ratio estimators along with DBH.

the difference in these estimates. Since the camera was held 1m above the ground, these reported values are actually estimates “squared-height above breast height, per unit area.” Pearson’s correlation coefficient was used for all reported correlations.

All statistical analyses were carried out in R (R Foundation for Statistical Computing, 2010), with the mixed effects model and ANOVA model being implemented via *glmer* and *aov* functions in the “lme4” and “stats” packages, respectively.

1.2.3 Results

Across the 33 plots, a total of 216 trees were identified using the prism sweep, and a height was obtained for 54 of these trees. The quadratic mean diameter was $0.31 \pm 0.11\text{m}$ (range=[0.11, 0.51]), with an average height of $20.2 \pm 8.2\text{m}$ (range=[10.0, 46.8]). The estimates of basal area and stocking density were $30.1 \pm 9.9 \text{ m}^2/\text{ha}$ (range=[18.4, 55.1]) and $659 \pm 762.0 \text{ t.p.h}$ (range=[120, 3767]), respectively (Table 1.2). Out of the 33 sample points, the average slope was 21.4 degrees, with 6 samples points taken on slopes less than 10 degrees, and 6 sample points taken on slopes greater than 30 degrees (Figure 1.9). The plots were primarily composed of red oak and white pine, with a significant number of sugar maple (*Acer saccharum*).

Overall, the correlation between different individuals’ photo counts was relatively high (Table 1.3). The average correlation was 0.7 across all subject-pairs, with a minimum value of 0.3 and a maximum of 0.9. The average photo counts for each individual for the unadjusted and slope-adjusted photos are given in Table 1.4. Since a HSF of 1 was used, these photo counts are also equal to the stand-level estimates of m^2/m^2 for each individual.

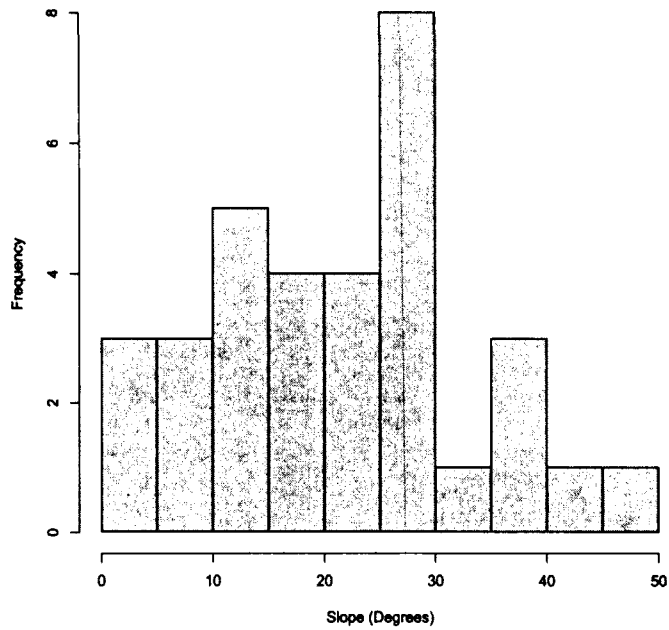


Figure 1.9: Histogram of the distribution of slopes for the thirty-three sample plots.

Table 1.3: Correlation between individuals' photo counts.

Individual	1	2	3	4	5	6	7	8	9	10
1	1	0.8	0.6	0.7	0.7	0.8	0.5	0.7	0.9	0.8
2		1	0.8	0.8	0.8	0.9	0.6	0.8	0.8	0.7
3			1	0.6	0.7	0.7	0.6	0.6	0.6	0.6
4				1	0.8	0.8	0.5	0.8	0.7	0.7
5					1	0.9	0.5	0.9	0.8	0.7
6						1	0.5	0.9	0.8	0.8
7							1	0.4	0.5	0.3
8								1	0.8	0.7
9									1	0.7
10										1

Table 1.4: Average tree count per photo for slope-adjusted and unadjusted data.

Individual	Adjusted	Unadjusted	Difference	(% change)
1	8.5	9.4	0.9	(+10.3%)
2	6.9	7.5	0.6	(+8.8%)
3	4.5	5.2	0.7	(+15.7%)
4	5.3	5.9	0.6	(+12.0%)
5	8.6	9.7	1.1	(+13.0%)
6	6.6	6.8	0.2	(+3.7%)
7	5.6	5.8	0.2	(+4.3%)
8	7.3	7.7	0.4	(+5.4%)
9	8.9	9.9	1.0	(+10.9%)
10	6.0	6.2	0.2	(+2.5%)
Mean \pm S.E.	6.8 \pm 0.5	7.4 \pm 0.5	0.6 \pm 0.1	(+8.8%) [†]

[†] $p < 0.001$ for a paired t-test that the difference in means is equal to zero.

Out of the 10 participants, all individuals counted more trees in the unadjusted photos. Individual #5 had the largest discrepancy, with 1.1 more trees counted in the unadjusted photos, on average, while Individuals #6 and #10 had the smallest discrepancies with 0.2 more photos counted in the unadjusted photos.

Across all individuals, failing to account for slope increased the number of trees counted per photograph by an average of 0.6 trees (6.8 \pm 0.5 vs. 7.4 \pm 0.5; Table 1.4), or an 8.8% increase. Ignoring the random effect of plot, a simple paired t-test evaluating the null hypothesis that the difference in means is equal to zero showed a significant effect of slope (95% CI for the difference = [0.4, 0.8], $p < 0.001$).

When included as a categorical variable in the mixed effects model, failing to adjust for slope increased the count by a multiplicative factor of 1.05 (95% CI=[1.03, 1.28], $p < 0.001$) for each 10 degree increase in slope (Table 1.5). This corresponds to relative increase in the count by factors of 1.05, 1.10, 1.16, 1.22, and 1.28, for slopes of 0-10, 10-20, 20-30, 30-40, and >40, respectively.

When modeling slope via a series of indicator variables representing 10 degree increments of slope, the only statistically significant effect of failing to adjust for slope was observed on those plots on 30-40 degree slopes. These plots also yielded the largest increase

Table 1.5: Results of the mixed-effects regression models (only fixed effects are reported).

Fixed Effect	Factor [†]	95% C.I.	P-Value
<i>Slope as a categorical variable:</i>			
Intercept [‡]	6.08	(5.06, 7.29)	<0.001
Avg. Tree Height	1.15 [§]	(1.03, 1.28)	0.01
Slope	1.05	(1.01, 1.08)	0.004
<i>Slope by 10 degree increments:</i>			
Intercept [‡]	6.33	(5.32, 7.54)	<0.001
Avg. Tree Height	1.14 [§]	(1.02, 1.27)	0.02
Slope:			
0-10	1.06	(0.93, 1.20)	0.37
10-20	1.05	(0.94, 1.17)	0.37
20-30	1.09	(0.99, 1.20)	0.07
30-40	1.24	(1.05, 1.47)	0.01
>40	1.02	(0.78, 1.35)	0.87

[†] Equal to $\exp(\text{coefficient})$.

[‡] Equivalent to the estimated average tree count for the slope-adjusted photo.

[§] Corresponds to an increase of one standard deviation from the mean.

in the count, with a relative increase of 24% (95% CI=[1.05, 1.47], $p < 0.01$). Failing to adjust for slope resulted in an increase in the count of less than 10% for plots on 0-10, 10-12, and 20-30 degree slopes, with reduction factors of 1.06, 1.05, and 1.09, respectively.

Adjusting for slope on sites >40 degrees had a seemingly smaller impact than adjusting for slope on sites of 30-40 degrees (point estimates of 1.24 and 1.02 for 30-40 and 40 degree slopes, respectively). This is likely a function of the relatively small sample size on these steeper slopes (four points on 30-40 degree slopes, and two points on slopes >40 degrees), as evidenced by the wide confidence intervals for both of these point estimates.

The effect of average tree height on the count did not differ appreciably between the two models, with increases of 15% (95% CI=[1.03, 1.28], $p = 0.01$) and 14% (95% CI=[1.02, 1.27], $p = 0.02$) for each standard deviation increase in average tree height. After adjusting for both the random and fixed effects, the average estimated tree count per photo on the adjusted photos (as indicated by the intercept term) likewise did not differ substantially between the two models, with point estimates of 6.08 and 6.33 trees per photo.

Table 1.6: Results of the Mixed-effects ANOVA model.

Source	D.f.	Sum of Squares	Mean Square (variance)	P-Value [†]
Avg. Tree Height	1	17.5	17.5	<0.001
Plot	30	101.7	3.4	<0.001
Individual	8	52.6	6.6	<0.001
Slope Class	5	3.1	0.6	<0.001
Plot x Individual	287	34.5	0.1	<0.001
Error	325	20.5	0.1	

[†] Testing the null hypothesis H_0 : No difference between source components.

In the mixed-effects ANOVA model, each source of variation yielded a statistically significant effect (Table 1.6). The largest source of variation was contributed by average tree height per plot (mean square = 17.5). Individual and plot accounted for the next two largest sources of variation in photo counts, with mean squares of 3.4 and 6.6, respectively. After accounting for tree height, plot, and individual, the relative contribution of slope to the overall variability in the counts was minimal (mean square = 0.6).

1.2.4 Discussion

The variable-radius circle technique presented in this study is a simple and intuitive way to adjust for slope to maintain unbiased estimates, regardless of the selected HSF or of the severity of the sloping terrain. If this slope-adjustment method is not used, then the impact of slope on VPSC should be negligible as long as an appropriate HSF is selected and the majority of sample points occur on slopes of less than approximately 35 degrees. Failing to adjust for slope can significantly affect the photo counts, but bias will only be observed when a majority of the sample points fall on very steep terrain and a small HSF is selected.

As discussed, this study used a HSF of 1 in order to quantify the slope-related bias across a wide range of slope expansion percentages. This HSF of 1 corresponded to expansion percentages ranging from 2% on 10 degree slopes to 46% on 40 degree slopes. To interpolate these results to those situations that employ a HSF other than 1, we first identify the slope expansion percentage that corresponds to the desired slope/HSF combination. We then find the slope that yields this same slope expansion percentage value when a HSF of

1 is selected. The bias in these two slope/HSF combinations is equivalent. For example, with a HSF of 4 on 20 degree slopes, the slope expansion percentage is about 2%. With a HSF of 1, a 2% slope expansion percentage occurs on slopes of 10 degrees. Therefore, the bias incurred on 20 degree slopes with a HSF of 4 is equivalent to the bias incurred on 10 degree slopes with a HSF of 1.

The only significant effect of slope was observed on 30-40 degree slopes (when modeled by 10-degree increments), which corresponds to slope expansion percentages of 19-46%. As mentioned, Ducey and Kershaw (2011) recommend a HSF between 3 and 6, where slope expansion percentages of 19-46% are not encountered until slopes of >40 degrees are reached. With a HSF of 3 or greater, slope expansion percentage will be approximately 6% on slopes of 30-40 degrees; roughly equivalent to the expansion percentages on slopes of <20 degrees as seen in this study using a HSF of 1. The bias incurred on these slopes was negligible and could not be statistically differentiated from other sources of variation in the photo counts. Therefore, the bias incurred on slopes greater than approximately 30 degrees using a HSF of 3 or greater will likewise be negligible.

More conservatively, if the majority of sample points occur on slopes of 30 degrees or less, then the average expansion percentage will be less than 5%, and the corresponding bias will likewise be less than 5%. For an adequately selected HSF that averages approximately 6 tree counts per photo, this will yield a maximum average inflation in the count of 0.3 trees per point in the extreme case that that all of the sample points occur on slopes of 30 degrees or greater. If only a few sample points fall on slopes exceeding 30 degrees, then the expected bias will be substantially less than 5%.

If the variable-radius circle technique is not employed, then some degree of bias will theoretically be unavoidable. Despite this, bias in this study could not statistically be differentiated from random variability on moderate slopes, even with 660 individual photo counts. Approximately 97% of the observed variability between photo counts was due to a combination of the natural variation in tree heights between plots, plot-level differences (e.g., species composition), and counting patterns between individuals. Conversely, less than 3% of the variability in the photo counts was due to slope differences. While these results indicate that slope can significantly bias the estimates of squared-height on steeper slopes,

the impact of slope is negligible compared to the unavoidable randomness contributed by the photo counting process and by the between-plot differences. These sources of variation effectively overwhelm the slope-incurred bias in all but the most extreme cases.

This study was subject to several limitations. The study area was dominated by a mixed-deciduous forest with a relatively closed canopy. In other forest types (e.g., open ponderosa pine stands) where individual tree tops are more easily identified and are less obstructed by foliage, the impact of slope might differ from what was observed here. In these settings, between-subject variability in the photo counts may be less, thus leading to more a prominent influence of slope. However, initial exploratory analysis of data collected in northeastern Oregon coniferous forests indicates that slope-related bias will be near zero in these forest types as well. With a HSF of 5, the sizes of the variable-radius circles were not visually distinguishable on slopes ranging from 0 degrees to 30 degrees, indicating that the only trees that will be incorrectly counted by failing to adjust for slope are those that fall almost exactly on the circle boundary. Despite these preliminary observations, a thorough bias assessment in coniferous forest types is warranted.

Furthermore, the slope correction method outlined above assumes a smooth, uniform slope that is rarely encountered in practice. Measurements of slope at each sample point are often imprecise estimates of the average slope within the immediate vicinity on the plot. In horizontal point sampling with a prism or angle gauge, the vertical displacement of each individual tree relative to plot-center can be adjusted for in order to maintain accurate inclusion probabilities for each tree, regardless of the heterogeneity of the slope (Beers, 1969). VPSC is not conducted on a tree-by-tree basis, so there is no obvious extension of this method to account for irregular slopes. In such settings, the randomness of terrain irregularities likely leads to minimal systematic bias in the overall estimate, but more research into this problem is warranted. Lastly, boundary correction methods for those sample points that fall close to the stand edge have yet to be investigated. Grosenbaugh (1958) outlines one such technique when conducting VPS on a tree-by-tree basis using traditional techniques, but these methods likewise have no immediate extension to VPSC.

1.2.5 Conclusions

The results of the bias assessment presented here suggest an upper bound on the expected bias incurred for failing to adjust for slope. Two methods of accounting for sloping terrain were investigated in this study: (1) maintaining unbiased estimates by digitally superimposing variable-radius circles to adjust for the slope at each sample point, and (2) ignoring slope altogether, provided that the HSF is appropriately selected and the majority of sample points are located on moderate terrain. The decision about which of these two options to use will depend heavily on the intended use of the information. If one aims to correlate the individual sample point estimates of height-squared to other measurements taken at that sample point, (e.g., in a double-sampling scheme to estimate board foot volume for a timber harvest), then directly adjusting for slope would likely be beneficial. Conversely, if an individual or a group is simply interested in quickly assessing different element of vertical structure in order to make rough qualitative inferences about stand structure (e.g., a high school classroom looking to assess the structure of a nearby forest) then there will be minimal need or desire to directly adjust for slope.

As indicated in (Ducey and Kershaw, 2011), vertical point sampling with a camera will likely be useful in those settings where an individual or a group is interested in assessing elements of forest structure without additional investment of time, money, or resources. It will also likely be appealing when different aspects of forest structure or health need to be rapidly assessed over a broad spatial extent. Since it requires no formal training and requires no investment in additional equipment (provided the user has access to a digital camera), VPSC will be a useful forest sampling method to a broad array of individuals, including regional foresters, private land owners, and citizen scientists. The research presented here indicates that vertical point sampling with a camera can be employed regardless of sloping terrain, either by directly adjusting for slope or by ignoring slope altogether.

Chapter 2

The Relationship between Survival Analysis and Point Quadrat Sampling: Theory and Applications

Numerous ground-based methods exist for assessing canopy structure and canopy density. These methods can be broadly separated into direct and indirect techniques (Seidel et al., 2011). Direct methods function by utilizing an instrument that comes into contact with plant tissue, either in a destructive or non-destructive manner. Conversely, indirect methods are those where the canopy parameters of interest are obtained via a mathematical or statistical relationship between the observed measurements and the desired parameters. While both direct and indirect methods have been used with varying degrees of success, each method is at least partially limited by one or more factors: the accuracy and usefulness of the resulting information; the time required to implement the method; the complexity of the sampling protocol; the cost of field equipment; or sizeable post-processing demands (Seidel et al., 2011; Jonckheere et al., 2004).

The two most prominent direct ground-based methods for estimating canopy structure include stratified clipping and scaffolding, where foliage density is measured within intervals in the canopy and related to allometric equations (Fujimori, 1971; Fukushima et al., 1998). These methods allow for estimation of the overall foliage density and the vertical distribution of foliage density within different canopy strata. A drawback is that they are time consuming and difficult to implement (Seidel et al., 2011). An additional direct method for estimating stand-level foliage density is through the use of litter traps, where foliage is accumulated in traps constructed on the forest floor (Ovington, 1963; Aber, 1979). While this is a simple and inexpensive measure of foliage density, it provides minimal

information about the vertical distribution of foliage and it is limited to those stands that have a well-defined leaf-drop period (Jonckheere et al., 2004).

Ground-based indirect contact methods include point quadrat methods, inclined point quadrat methods, and optical or photographic methods (Levy and Madden, 1933; Wilson, 1960; MacArthur and Horn, 1969; Aber, 1979; Rich, 1990). More recently, ground-based optical point quadrat methods such as LIDAR and laser-point sampling have been used with increasing frequency (Vanderbilt et al., 1979; Radtke and Bolstad, 2001; Lovell et al., 2003; Parker et al., 2004; Coops et al., 2007). While these methods can obtain detailed estimates of canopy profiles and foliage density, techniques for better analyzing and interpreting the resulting data are continuously being developed and still warrant further investigation (see: Côté et al., 2011; Zhao et al., 2011; Seidel et al., 2012). More importantly, the technology and data analysis procedures required to implement these techniques can be prohibitively expensive and time-consuming to operate (Radtke and Bolstad, 2001; Seidel et al., 2011).

Lastly, various remote sensing methods, such as airborne LIDAR or MODIS satellite imagery, provide alternate platforms for estimating leaf area index, canopy structure, and vertical structure. While these tools provide a powerful set of methods for estimating different forest parameters (see: Garrigues et al., 2008; Houborg et al., 2007; Zhao et al., 2009; Cao et al., 2012), the focus of this study is to address the limitations of ground-based methods of assessing canopy structure. Ground-based methods have the potential to be more widely used by a broad range of interest groups (e.g., private land-owners, citizen scientists, regional foresters, or conservation organizations), as they require relatively less access to specialized equipment, resources, or expert-knowledge.

In this paper, we build upon the ground-based point quadrat and laser point quadrat sampling methods of Levy and Madden (1933), MacArthur and Horn (1969), and Radtke and Bolstad (2001) by illustrating how these methods are analogous to the field of survival analysis as used in biomedical and technology research. By utilizing this relationship, we can increase the overall efficiency, accuracy, and useability of this technique with minimal additional cost or resources. This will enable laser point quadrat sampling to overcome many of the drawbacks that currently limit existing ground-based sampling methods, and

increase its use as a rapid, simple, and inexpensive method of assessing canopy structure.

As a starting point for incorporating survival analysis methods into point quadrat analysis, this study nests point quadrat sampling within a regression framework. This setting allows for direct quantification of the impact of stand-level attributes on leaf area index and canopy profile. Importantly, by combining information across plots, regression analysis should result in fewer sample points needed to obtain comparable canopy profile estimates. Lastly, the methods outlined in this paper are quick, inexpensive, and simple, and they result in robust comparisons of canopy density between stands.

First, a brief review of existing point quadrat sampling methods is given, followed by an overview of survival analysis methods. The relationship between survival analysis and point quadrat sampling is outlined, and a brief explanation of the regression framework for survival analysis is provided. Lastly, a field study was conducted in northeastern Oregon to investigate the feasibility and efficacy of applying regression-based survival analysis techniques to point laser quadrat data. The results and conclusions from this study are reported, along with limitations and suggestions for further research on the use of survival analysis for point quadrat data.

2.1 Point Quadrat Sampling

The primary variables of interest in point quadrat sampling are leaf area index (LAI), canopy density, and canopy profile. LAI is a unitless metric that approximates the photosynthetic capacity and functionality of forest canopies by reflecting the amount of leaf area per surface area. As leaves are responsible for respiration, gas-exchange, and energy production, total leaf area is indicative of differing functional capacities of forest canopies (Ford and Deans, 1978; Hollinger, 1989; Baldocchi and Harley, 2006). Not only is point quadrat sampling a tool for estimating stand-level LAI, but it can be used to estimate LAI within different canopy strata at various heights above the forest floor. This allows for quantification and visualization of the canopy profile (MacArthur and Horn, 1969; Radtke and Bolstad, 2001).

LAI can be defined numerous ways: as the horizontally projected leaf area (Ross, 1981; Bolstad and Gower, 1990); as the total one-sided leaf area per unit ground area

(Watson, 1947); or as one-half of the total leaf area per unit ground area (Chen and Black, 1992a). Following in the tradition of Levy and Madden (1933), MacArthur and Horn (1969), and Radtke and Bolstad (2001), the proposed methods do not account for the angular distribution of foliage. The term “LAI” throughout this paper therefore refers to the horizontally projected leaf area per unit ground area.

The proposed methodology in this paper makes several additional assumptions. As discussed in Radtke and Bolstad (2001), laser point quadrat sampling measures plant area index (projected plant area per unit ground area), rather than leaf area index, as the laser makes no distinction between foliage and woody material. Despite this distinction, the term “leaf area index” is used throughout this paper to coincide with traditional point quadrat sampling terminology. Chen and Black (1992a) further point out that the horizontal projection for non-flat leaves can be problematic and has limited biological significance. Lastly, the clumping of foliage due to the non-randomness of shoot positions and the interdependence of leaf positions in the canopy can lead to underestimates of leaf area index (Chen and Black, 1992b; Chen and Cihlar, 1995; Radtke and Bolstad, 2001).

While these limitations can affect the interpretability of the results, the purpose of the proposed methods is to further the applicability of point quadrat sampling by highlighting the similarities between survival analysis and existing point quadrat methods. Techniques for addressing these limitations have been discussed elsewhere (see: Wilson, 1960, 1963, 1965; Chen and Black, 1992a; Chen and Cihlar, 1995; Denison, 1997; Welles and Cohen, 1996; Groeneveld, 1997; Radtke and Bolstad, 2001). Additional suggestions for incorporating these techniques into the survival analysis framework are discussed below.

2.1.1 MacArthur and Horn Estimator

The original methods for conducting point quadrat sampling involved passing a needle (i.e., a sewing needle) vertically through the canopy and recording the number of contacts between the needle and foliage (Levy and Madden, 1933). MacArthur and Horn (1969) expanded upon this method by using a camera to measure height to first leaf. Under this scheme, a random point was selected, a tripod was centered over that location, and a camera was fixed to the tripod and sighted vertically up into the forest canopy. Using the

focusing scale on the telephoto lens, distance to the first leaf at that point was measured. By repeating this procedure across a random series of sample points they obtained a set measurements of “height to first leaf.”

Using these measurements, MacArthur and Horn (1969) derived a method for estimating leaf area index. Let $D(h)$ represent the instantaneous density (in # of leaves per vertical meter) of the canopy at height h above the forest floor. This density is equivalent to the instantaneous LAI per meter of height at that point. Let $\varphi(h)$ be the probability that there are no leaves over the first h meters. MacArthur and Horn (1969) derived the following relationship between $D(h)$ and $\varphi(h)$:

$$D(h) = -\frac{d[\log \varphi(h)]}{dh} . \quad (2.1)$$

The total density (or equivalently, LAI) between any two heights h_1 and h_2 is found by integrating the density function between these endpoints. Using the above equality, we have that:

$$\int_{h_1}^{h_2} D(h)dh = \log \frac{\varphi(h_1)}{\varphi(h_2)} . \quad (2.2)$$

As $\varphi(h)$ is not known, MacArthur and Horn (1969) suggest estimating it via the proportion of measurements that exceed h . Under this estimation procedure, equation 2.2 reduces to:

$$LAI(h_1, h_2) = \log \frac{n_{h_1}}{n_{h_2}}, \quad (2.3)$$

where $LAI(h_1, h_2)$ is the leaf area index between heights h_1 and h_2 , and n_{h_1} and n_{h_2} are the number of measurements that exceeded these two heights, respectively. The MacArthur-Horn estimate of the leaf area index of the site is therefore equal to:

$$LAI_{MH} = \log \frac{N}{n_0} = \log(g^{-1}), \quad (2.4)$$

where N is the total number of points, n_0 is the number of points that did not intercept any foliage, and g is the gap fraction (i.e., the proportion of sample points that did not intercept any foliage). Additionally, MacArthur and Horn (1969) note that the canopy

profile is achieved by plotting h vs. $D(h)$.

2.1.2 Laser Point Sampling - Radtke and Bolstad Estimator

Radtke and Bolstad (2001) modified the sampling procedure of MacArthur and Horn (1969) by utilizing a laser rangefinder instead of a camera to measure the distance to first leaf. The laser rangefinder (LRF) was attached to a tripod and a measurement was taken every 10-13 cm in a grid pattern throughout a 13 x 13 m plot within each stand. While telephoto lenses have been used for decades for measuring different aspects of forest structure, LRFs are increasingly being used in many forestry and environmental applications for several reasons: they can be more accurate than telephoto lenses at measuring distances; the distance measurements can be obtained within seconds at each point; and the data are electronically stored for later analysis (Radtke and Bolstad, 2001). In addition, some LRFs come equipped with tilt sensors that are useful for obtaining quick measurements of additional canopy parameters, such as height to live crown, live crown ratio, and crown radius.

Radtke and Bolstad (2001) provide an alternative estimator to that proposed by MacArthur and Horn (1969). They divided the canopy into a series of non-overlapping intervals of defined width. To ensure that there are no overlapping leaves within the i^{th} layer, the authors suggest making the intervals so small that only a single laser shot is intercepted within that interval. Under this assumption, the Radtke-Bolstad estimate of leaf area in interval i is equal to:

$$q_i = \frac{1}{n_i} \quad , \quad (2.5)$$

where n_i is the number of measurements that were intercepted in or above interval i .

To estimate the leaf area between two heights, h_L and h_U , the leaf area estimates for the each interval between these two heights are summed:

$$LAI(h_1, h_2) = \sum_{j: h_L \leq h_j < h_U} \frac{1}{n_j} \quad . \quad (2.6)$$

Additionally, Radtke and Bolstad (2001) show that this estimator is asymptotically equiv-

alent to Equation 2.3.

2.2 Survival Analysis

In biostatistics, whenever a subject is followed until a pre-defined event occurs the resulting data are termed “failure time data” or “time-to-event data” (Kalbfleisch and Prentice, 2002). The group of techniques and methods used to analyze failure time data is known as “survival analysis”. Failure time data occur most frequently in biomedical studies where patients or persons of interest are followed for a period of time until a specific event occurs (e.g., death, adverse reaction, clinical diagnosis, etc.). It is also commonly used in industrial applications, where failure rates and life-expectancies of specific components are of interest.

In these studies, each subject or component is followed until either an event occurs or until they are censored from the study for unrelated reasons. Censoring occurs whenever the subject is removed from the study before the event of interest can be observed. Examples of censoring in biomedical and health studies include death from unrelated causes, loss of contact, self-removal from the study, or the termination of the study before all events have been observed. While integral in most survival analysis applications, censoring is not addressed further in this paper. For a complete discussion of censoring, see Kalbfleisch and Prentice (2002).

Survival analysis methods generally focus on making inference about two functions: the survivor function and the hazard function. The survivor function can be written in terms of the cumulative distribution function, $F(t) = P(T \leq t)$ (i.e., the probability that failure occurs before time t). The probability that a subject survives up to and including time t is given by the survivor function:

$$S(t) = 1 - F(t) = P(T > t), \quad 0 < t < \infty \quad . \quad (2.7)$$

This equation is sometimes also referred to as the “reliability function” (Kalbfleisch and Prentice, 2002; Tableman et al., 2004). This is a monotone decreasing function that equals one at time zero, and approaches zero as $t \rightarrow \infty$.

In survival analysis, the hazard function, denoted by $\lambda(\cdot)$, specifies the instantaneous

rate of failure at a given time t . It can be defined in terms of the survivor function:

$$\begin{aligned}\lambda(t) &= \lim_{h \rightarrow 0^+} \frac{P(t \leq T < t+h \mid T \geq t)}{h} \\ &= -\frac{d \log S(t)}{dt}.\end{aligned}\tag{2.8}$$

See Ibrahim et al. (2010) for a complete derivation of this relationship. The hazard function is often referred to as a hazard rate function, as it can be conceptualized as representing the failure rate at any given time. It follows that the the cumulative hazard between any two time points t_1 and t_2 is given by:

$$\int_{t_1}^{t_2} \lambda(t) dt = \log \frac{S(t_1)}{S(t_2)}.\tag{2.9}$$

Survival analysis is an increasingly advanced and complex field due to its immediate relevancy to clinical trials and medical testing. In traditional survival analysis, most inference involves comparing survivorship curves between populations. Additionally, both the survivor function and the hazard function are useful for identifying those time periods that yield higher risks of failure. An enormous body of literature exists that provides well-established techniques for analyzing failure data, including parametric and non-parametric models, Bayesian methods, sequential analysis methods, competing risk models, correlated failure time models, multidimensional data, sample size considerations, model selection and comparison, and complex censoring mechanisms (Oakes, 2001; Kalbfleisch and Prentice, 2002; Ibrahim et al., 2010).

2.2.1 Equivalence of Point Quadrat Sampling and Survival Analysis

In point quadrat sampling, we can imagine that each laser shot is “enrolled” at height zero and followed upwards in height until the pre-defined event, leaf interception, occurs. Just as in the traditional survival setting, this results in time-to-event data. The difference is simply the reference scale: in traditional survival analysis the subjects are followed across time until the first event occurs; in point quadrat sampling the subjects (laser shots) are followed across distance until the first leaf intercept occurs. As evidenced by the similarities

between eqs. 2.1 and 2.8, and between eqs. 2.2 and 2.9, this change from time to distance does not affect the underlying mathematics.

Under point quadrat sampling, $\varphi(h)$ is defined as the probability that there are no leaves over the first h meters. This is equivalent to the probability that a laser beam “survives” to height h without intercepting a leaf, and it is identical to the survivor function. The MacArthur and Horn density function $D(h)$ is also equivalent to the hazard function $\lambda(t)$, as both are describing an instantaneous risk of failure; the former in terms of leaves per meter, the latter in terms of events per time. Therefore the following congruencies hold:

$$D(\cdot) \cong \lambda(\cdot) \tag{2.10}$$

$$\varphi(\cdot) \cong S(\cdot) \tag{2.11}$$

Furthermore, MacArthur and Horn (1969) suggested estimating $\varphi(h)$ via the proportion of sample points that exceeded h . In survival analysis, this is known as the empirical survival function (Tableman et al., 2004). The overall estimate of leaf area as suggested by MacArthur and Horn (1969) in equation 2.4 is equivalent to the Kaplan-Meier product limit estimate of the cumulative hazard (Kaplan and Meier, 1958).

Using the results established by Kaplan and Meier (1958), we can obtain a variance estimate for the MacArthur-Horn LAI estimate. Let (h_1, \dots, h_n) be the ordered heights at which a leaf was intercepted, d_i denote total number of leaf intercepts at h_i , and n_i denote the total number of points that were intercepted at or above h_i . It follows that the variance is:

$$\widehat{var}(LAI_{MH}) = \widehat{var}\left(\log\left(\frac{1}{g}\right)\right) = \sum_{i=1}^n \frac{d_i}{n_i(n_i - d_i)} \tag{2.12}$$

In addition, the estimate by Radtke and Bolstad (2001) of LAI between two height h_1 and h_2 as shown in eqn. 2.6 is equivalent to the Nelson-Aalen estimate of the cumulative hazard function, provided there are no two identical laser intercept heights (Tableman et al., 2004). The variance of eqn. 2.6 is equal to (Nelson, 1972; Aalen, 1978):

$$\widehat{var}(LAI(h_1, h_2)) = \sum_{j: h_1 \leq h_j < h_2} \frac{1}{n_j^2} \tag{2.13}$$

Not only does survival analysis allow for variance estimation of point quadrat estimators, but it importantly enables point quadrat sampling to be embedded within a regression framework. The methods provided by MacArthur and Horn (1969) and Radtke and Bolstad (2001) can be used to estimate LAI and to obtain a graph of the canopy profile, but they are limited in their ability to incorporate stand-level covariates to improve canopy profile and LAI estimation.

2.3 Regression Analysis of Point Quadrat Data

Radtke and Bolstad (2001) note that the number of sample points needed to obtain accurate estimates using laser point quadrat sampling can be quite high. They observed minimal improvement in LAI estimation over existing optical point quadrat methods with a point-density of approximately 1000 laser points per 13x13m plot (60,000 sample points per hectare). Fukushima et al. (1998) suggest that adequate foliage profile estimates can be obtained with a sample size of 64 points per 1.8x1.8m (approximately 200,000 sample points per hectare) when conducting standard (non-laser) point quadrat sampling. These large sample sizes make direct inference of a single stand difficult, and make comparison between many stands impractical using traditional point quadrat analysis methods.

Regression analysis of point quadrat data can address this situation by incorporating stand-specific covariates into the LAI estimation process. This enhanced methodology can lead to improved estimates of LAI and canopy profile, and allow for quantitative identification of those stand-level variables that affect canopy structure and LAI. Survival-based regression analysis can potentially result in fewer sample points by combining information between plots, depending on the suitability of the selected model.

Regression models for time-to-event data are usually defined via the hazard function. Due to the equivalence of the hazard function and the point quadrat density function, the hazard function in the point quadrat scenario will be denoted by $D(\cdot)$. Regression models can be implemented using a parametric form for the hazard function, or semi-parametrically by way of a Cox proportional hazards (or relative risk) model (Cox, 1972). For a complete discussion of different regression models for survival analysis, see Kalbfleisch and Prentice

(2002) and Tableman et al. (2004).

2.3.1 Study Design

As a simple illustration of the usefulness of the survival-based regression framework for analyzing point quadrat sampling, a field study was conducted in the counties of Union and Baker in northeast Oregon (Figure 2.1). This study is part of a larger project investigating how differences in land ownership and forest management affect differing metrics of forest health, forest structure, and land cover change. The full study area also includes Wallowa county, though data from this county is not included in this analysis.

Combined, Baker and Union cover approximately 13,279 km². Annual precipitation in this region is generally less than 50 cm per year (with some areas receiving as little as 25 cm annually), due in large part to a rain-shadow effect from the Cascade Mountain Range to the west. This lack of precipitation leads to dry, desert-like condition throughout much of the two counties, with large areas of land comprised solely of low-lying shrubs and bushes.

These counties are bounded by the Wallowa-Whitman National Forest to the northeast, the Elkhorn Mountain to the east (a subrange of the Blue Mountains), and the Umatilla National Forest to the north (likewise located within the Blue Mountains). Elevation ranges from 512 m at the lowest point, to nearly 3000 m at the mountain peaks. The valleys between these mountains are largely non-forested, and used primarily for agriculture and farming. The slopes along the mountain ranges are mostly forested, and are dominated by dry upland forest communities comprised of ponderosa pine (*Pinus ponderosa*), Douglas fir (*Pseudotsuga menziesii*), grand fir (*Abies grandis*), and western larch (*Larix occidentalis*); with lesser percentages of lodgepole pine (*Pinus contorta*), engelmann spruce (*Picea engelmannii*), and subalpine fir (*Abies lasiocarpa*). The lower slopes between the valleys and the mountains are comprised primarily of ponderosa pine, due to the warmer temperatures and limited moisture. Subalpine fir forest types were encountered at the higher elevations (exclusively in US Forest Service land), with mixed conifer forest types located at intermediate elevations in between the ponderosa pine and subalpine fir forests.

Across Union and Baker counties, sixty plots were subjectively selected in order to obtain a wide range of forest types and structures. Twenty-one plots were located on private

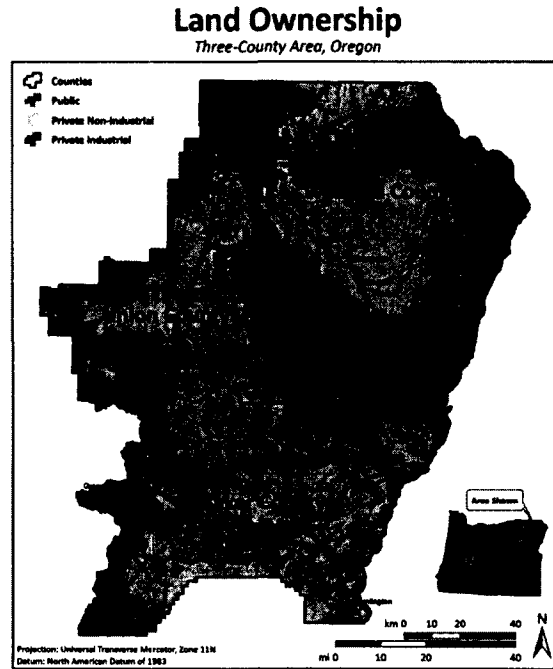


Figure 2.1: The field study took place in the counties of Union and Baker in northeast Oregon. Wallowa county is included within the larger project, though no data is reported from this county for this study.

industrial forests, sixteen were located on non-industrial private property, and twenty-three were located on US Forest Service land. Historically, this region had a resource-based economy that relied heavily on timber harvesting. Recent changes in local and federal policy have lead to drastic decreases in the harvesting rates across the area, with strong implications for forest management throughout the two counties (Adams and Latta, 2003; Youngblood, 2005). Due to this decline in the timber industry, different land owners likely have different management objectives for their land, so the forests located across these three ownership classes likely have disparate canopy structures.

Within each of the sixty stands, a plot center was randomly selected. At each plot, a series of three sample points was established, each spaced 30m apart in a north-south orientation (Figure 2.2). At each point, trees were tallied using a Spiegel Relaskop with a metric BAF of $4m^2/ha$ ($17.4ft^2/ac$). The diameters of all tallied trees were measured, and tree species were recorded. A subset of these trees was identified using a metric BAF of either $4m^2/ha$ ($17.4ft^2/ac$), $9m^2/ha$ ($39.2ft^2/ac$), or $16m^2/ha$ ($69.7ft^2/ac$), and height measurements were obtained for these trees using a clinometer. The specific tree-height BAF

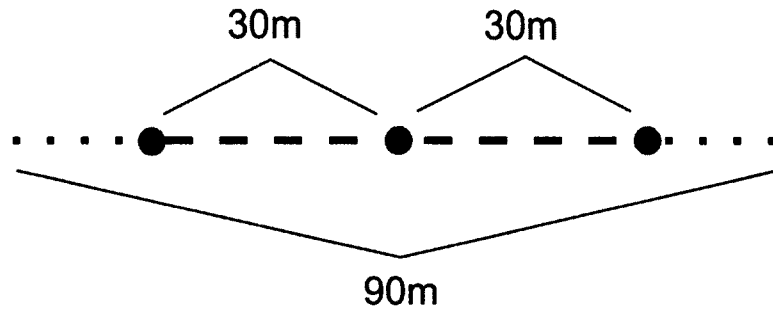


Figure 2.2: Three sample points were established 30m apart in a north-south orientation within each stand. Tree diameter and height measurements were taken on a subset of trees at each point using horizontal point sampling. Laser point quadrat measurements were taken every 1m along a 90m transect that extended 15m beyond each of the outer two sample points.

for each plot was selected in order to obtain approximately 10-12 tree height measurements per plot. Based on the dominant and co-dominant tree species at each plot, the sites were categorized as: (a) ponderosa pine, (b) ponderosa mix, (c) mixed conifer, (d) lodgepole pine, or (e) subalpine fir forest types.

A 90m transect was established in each stand, centered at the middle sample point and extending 15m beyond the northern and southern sample points (Figure 2.2). A Leica DISTO D8 Laser Rangefinder with tilt sensor was used to measure height-to-first-leaf along this transect. This laser rangefinder has a range of 0.05m to 200m with an accuracy of approximately 1mm. It includes a “long range mode” which was used in bright sunlight or for very high canopies.

A field crew member walked along the transect and took a height-to-canopy measurement every 1m, with no measurement taken at the 90m mark. The rangefinder was held approximately 1m from the ground, and the tilt sensor was used to ensure the unit was being held vertically. Using the built-in wireless capability of the DISTO D8, the height measurement at each location along the transect was immediately transmitted to a hand-held tablet computer being operated by another field crew member. As the rangefinder was held approximately 1m above the ground, a value of 1m was added to each laser intercept height.

This procedure resulted in ninety foliage height measurements for each stand. Sky-hits (those points that passed through the entire canopy without intercepting any foliage) returned an error and could not be transmitted to the tablet computer. The number of sky

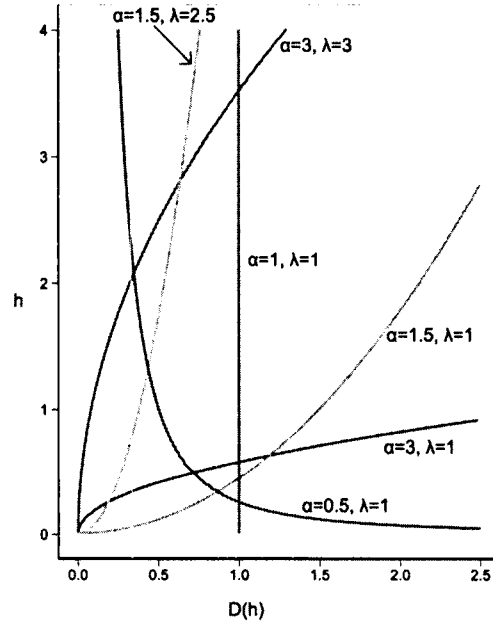


Figure 2.3: Various density functions given by the 2-parameter Weibull distribution. While limited to monotonic functions, the Weibull hazard function provides a wide range of canopy profile shapes.

hits for each stand was therefore indicated by the number of missing measurements for that stand.

2.3.2 Statistical Analysis

The two parameter Weibull distribution was used to model laser intercept heights. Under this distribution, the density function is given by:

$$D(h) = \frac{\alpha}{\lambda} \left(\frac{h}{\lambda} \right)^{\alpha-1} \quad \alpha, \lambda > 0. \quad (2.14)$$

This distribution provides a range of monotonic increasing, decreasing, and constant canopy LAI profiles (Figure 2.3). While the possible forms for $D(h)$ given by the Weibull distribution are undoubtedly oversimplifications of real-world canopy profiles due to the lack of unimodal or bimodal hazard functions, the flexibility of these potential LAI profiles makes it ideal for illustrative purposes and for modeling different forest types with disparate canopy structures. This distribution has previously been used for modeling canopy structure in various forms (Schreuder and Swank, 1974; Moril and Hagihara, 1991; Yang et al., 1993; Coops

et al., 2007), though not via the hazard function. Alternative parametric functions that are commonly used in survival analysis but were excluded from this analysis due to various limitations are the exponential, log-normal, log-logistic, Cauchy, and gamma distributions.

Estimates of basal area per hectare (BA), quadratic mean diameter (QMD), trees per hectare (TPH), and stand density index (SDI) were calculated for each point, and then averaged across the three points to obtain stand-level estimates for each plot. SDI was calculated using an additive version of Reineke's SDI given by Long and Daniel (1990). In addition, as specified in Gove (2003), a two-parameter size-biased Weibull model was fit to the DBH measurements for each plot to estimate the complete diameter distribution of trees in that stand.

Incorporating Sky Hits into the Regression Framework

In the non-parametric methods proposed by MacArthur and Horn (1969) and Radtke and Bolstad (2001), sky-hits do not present any issues. In the parametric framework, however, sky-hits propose a unique challenge, as their survival times are infinite. These points cannot be trivially excluded from the analysis, as they are informative with respect to total leaf area. They also cannot be included as censored observations, as censored observations are still assumed to have finite failure times.

The most straight forward solution is to partition the data (and equivalently, partition the stand) into the set of points that fall beneath foliage (B) and the set of points not beneath foliage (\tilde{B}). Ignoring the difference between foliage and woody canopy material, this latter set is equal to all of the locations in the stand that would result in sky hits. Implicit in this partitioning is the assumption that foliage movement is minimal and that the the gap fraction for each stand is a fixed value that is roughly constant over short periods of time.

Let $D_B(h)$ denote the density of foliage at a given height among all points beneath foliage, and let $D_{\tilde{B}}(h)$ denote the density of foliage at a given height among all points not beneath foliage. It is clear that this latter term is equal to zero for all values of h . The overall density (or leaf area) of the stand is equivalent to an area-averaged density across the stand. If the total area A of a stand is partitioned into the area of the stand beneath

foliage, A_B , and the area not beneath foliage, $A_{\tilde{B}}$, then the overall density function is:

$$\begin{aligned}
D(h) &= \frac{A_B \cdot D_B(h) + A_{\tilde{B}} \cdot D_{\tilde{B}}(h)}{A} \\
&= \frac{A_B}{A} \cdot D_B(h) + \frac{A_{\tilde{B}}}{A} \cdot D_{\tilde{B}}(h) \\
&= (1 - g) \cdot D_B(h) + g \cdot D_{\tilde{B}}(h) = (1 - g) \cdot D_B(h) + g \cdot 0 \\
&= (1 - g) \cdot D_B(h) \quad , \tag{2.15}
\end{aligned}$$

where g is the gap fraction of the stand. This is equal to an area-weighted density function with weights defined by the gap fraction. More rigorously, this same result can be derived via the definition of the hazard function, as given in equation 2.8, using basic rules of probability theory:

$$\begin{aligned}
D(h) &= \lambda(h) \\
&= \lim_{\Delta \rightarrow 0^+} \frac{P(h \leq H < h + \Delta \mid H \geq h)}{\Delta} \\
&= \lim_{\Delta \rightarrow 0^+} \frac{P(h \leq H < h + \Delta \mid H \geq h, B)P(B)}{\Delta} \\
&\quad + \lim_{\Delta \rightarrow 0^+} \frac{P(h \leq H < h + \Delta \mid H \geq h, \tilde{B})P(\tilde{B})}{\Delta} \\
&= \lim_{\Delta \rightarrow 0^+} \frac{P(h \leq H < h + \Delta \mid H \geq h, B)P(B)}{\Delta} + 0 \\
&= P(B) \cdot \lim_{\Delta \rightarrow 0^+} \frac{P(h \leq H < h + \Delta \mid H \geq h, B)}{\Delta} \\
&= P(B) \cdot D_B(h) \\
&= (1 - g) \cdot D_B(h) \quad , \tag{2.16}
\end{aligned}$$

where the fourth equality follows because $P(h \leq H < h + \Delta \mid H \geq h, \tilde{B}) = 0$, as there is no risk of intercepting a leaf when not under the canopy.

A further problem with this model involves the truncated nature of forest canopies. All parametric hazard functions satisfy the condition that their integral on $(0, \infty)$ is equal to infinity. Thus, for any parametric form of $D_B(h)$, it must be that:

$$\int_0^{\infty} D(h)dh = (1 - g) \int_0^{\infty} D_B(h)dh = \infty \quad . \tag{2.17}$$

To alleviate this problem, we can calculate the total leaf area of the stand by integrating the density function only up to the upper limit of the canopy:

$$LAI = \int_0^{h_m} D(h)dh = (1 - g) \int_0^{h_m} D_B(h)dh < \infty \quad , \quad (2.18)$$

which is equivalent to truncating the distribution of canopy foliage at the max tree height h_m .

An important implication of this result is that estimation of the survival regression parameters can be conducted using only the intercepted laser hits, thus eliminating the difficulty in incorporating infinite survival times into the model. Under this framework, estimation of the the overall parametric density equation is a four step process: (1) estimate $D_B(h)$ using only the laser hits that intercepted foliage, (2) estimate g using the proportion of sky hits, (3) estimate $D(h)$ by taking the product of $D_B(h)$ and $(1 - g)$, and (4) truncate $D(h)$ at the maximum canopy height. It should be noted that this above method applies only to the situation where $D_B(h)$ is assumed to take a parametric form, as the non-parametric empirical estimate of $D_B(h)$ will necessarily be equal to infinity at the height of the highest point intercept (MacArthur and Horn, 1969).

Canopy Profiles and LAI Estimates

With the Weibull regression model, covariates act multiplicatively on the density function. Let $\vec{X}_j = (X_{1j}, X_{2j}, \dots, X_{sj})$ be the covariate vector for the j^{th} plot. If the i^{th} intercepted laser shot on the j^{th} plot, denoted by h_{ij} , is assumed to follow a Weibull distribution, then the density function for the j^{th} plot is given by:

$$D_B(h_{ij} | \vec{X}_j) = D_{B,0}(h_{ij}) \cdot \exp(\beta_1 \cdot X_{1j} + \beta_2 \cdot X_{2j} + \dots + \beta_s \cdot X_{sj}) \quad , \quad (2.19)$$

where $D_{B,0}(\cdot)$ corresponds to a baseline density function.

The covariates included in the Weibull regression model were: BA, TPH, SDI, the shape and scale parameters from the size-biased Weibull diameter models, and indicator variables for private and industrial land (the default case being USFS land). QMD was

originally included, but was removed due to high collinearity with the size-biased Weibull diameter parameters. To allow for greater flexibility in the density function, the Weibull regression model was fit separately for ponderosa pine, ponderosa pine mix, and mixed conifer forest types. Using the above framework, the coefficients were estimated using only those laser shots that intercepted foliage.

For each site, both parametric and non-parametric canopy profiles were estimated. The coefficients from the fitted regression models were used to estimate a separate density function for each site. These site-specific density functions were each multiplied by the corresponding empirical gap fraction for that site (the proportion of sky hits at that site), resulting in a smooth estimate of the canopy profile as given in equation 2.15. Canopy profiles were obtained via a graph of h against $D(h)$, and the maximum canopy height for each site was estimated as the maximum of the recorded tree heights or laser intercept heights for that site.

The non-parametric estimates of the canopy profiles were obtained by dividing the canopy height into 1m intervals, and using equation 2.6 to calculate the Radtke-Bolstad LAI estimate within each of these intervals for each site. These interval-specific LAI estimates were then plotted using a step-wise linear function.

Total estimates of LAI were calculated using two different methods: the Kaplan-Meier product-limit estimator (equivalent to the MacArthur-Horm estimator), and parametrically using equations 2.18 and 2.19 via the fitted regression models. As with the parametric canopy profiles, the maximum canopy height used to calculate the parametric LAI was estimated as the maximum of the recorded tree heights or laser intercept heights for each plot. The average of the recorded tree heights was also considered, but the results did not differ appreciably from using the maximum recorded tree height.

Quantifying the Effect of Stand-Level Covariates on LAI

An attractive aspect of the Weibull distribution in the survival analysis setting is the intuitive interpretation of the regression coefficients. Each coefficient $\exp(\beta_k)$ in equation 2.19 is equal to relative difference in density between two stands for a one unit increase in the variable X_k , holding the gap fraction and the remaining covariates constant. From

equation 2.18, it follows that $\exp(\beta_k)$ is equal to the relative difference in total leaf area between two stands for a one unit different in the k^{th} covariate, holding the gap fraction and the remaining covariates constant. This factor, $\exp(\beta_k)$, is referred to as the LAI ratio for the k^{th} covariate.

Confidence intervals for all reported values were calculated using Student's t-distribution. All statistical analyses were conducted in R (R Foundation for Statistical Computing, 2010) using the *survival* package and the *muha* package for parametric and nonparametric estimation, respectively.

2.3.3 Results

Out of the sixty sites, 22 were classified as mixed conifer, 19 were classified as ponderosa pine, 14 were classified as ponderosa mix, 4 were classified as Subalpine fir, and 1 was classified as lodgepole pine. Due to the low number of subalpine fir and lodgepole pine plots, these sites were omitted, resulting in 55 sites used in the analysis.

The sites were primarily open, with gap fractions ranging from approximately 60 to 70% (as estimated by the fraction of sky hits). The sites in each of the three remaining forest types were generally comparable in QMD, height, and live crown ratio (Table 2.1). The mixed conifer stands showed less signs of active management and/or timber harvests. Accordingly, these stands had greater crown closure and denser foliage throughout the canopy strata. These mixed conifer sites had slightly more basal area per hectare than ponderosa and ponderosa mix stands, with 29.5 m²/ha compared to 21.8 and 21.0 m²/ha, respectively (Table 2.1). Trees per hectare and SDI were also markedly larger in mixed conifer stands.

Ponderosa mix stands differed from ponderosa pine stands primarily due to the presence of tall Douglas fir and larch trees. This resulted in a slightly larger mean height for ponderosa mix site (21.8m vs. 17.7m) and slightly larger mean diameter for ponderosa mix sites (0.32m vs 0.30m). The gap fractions and live crown ratios were nearly identical for these two forest types.

Table 2.1: Summary statistics of the 55 forest plots[†].

Variable	Mixed Conifer (n=22) Mean ± S.D.	Ponderosa Pine (n=19) Mean ± S.D.	Ponderosa Mix (n=14) Mean ± S.D.
BA (m ² / ha)	29.5 ± 12.6	21.8 ± 12.6	21.0 ± 9.1
QMD (m)	0.26 ± 0.10	0.30 ± 0.09	0.32 ± 0.07
Trees per Hectare	1036.9 ± 1610.2	353.2 ± 251.1	313.6 ± 214.7
Stand Density Index	52.6 ± 58.4	30.6 ± 20.6	41.1 ± 21.7
Height (m)	19.8 ± 6.4	17.7 ± 5.1	21.8 ± 4.9
Live Crown Ratio (%)	28 ± 13	30 ± 11	28 ± 8
Gap Fraction [‡] (%)	60 ± 19	68 ± 13	68 ± 12
USFS Land (n)	7	3	8
Private Non-Industrial Land (n)	4	10	2
Private Industrial Land (n)	11	6	4

[†] Lodgepole pine and subalpine-fir sites (n=5) are omitted.

[‡] Estimated using the proportion of laser sky hits.

Regression-based canopy profiles for all sites are given in Figure 2.4. The ponderosa pine sites were generally the shortest of the three forest types, with minimal foliage density below 10m that increases relatively quickly near the upper portion of the canopy. In contrast, the ponderosa mix sites showed more uniform foliage density throughout all strata. The mixed conifer stands showed the most even distribution of foliage throughout the canopy, with the majority of stands having very little change in foliage density from 10m in height to the top of the canopy.

To illustrate how these regression-based canopy profiles compare to the traditional non-parametric profiles, both the non-parametric and parametric canopy profiles for six different stands are shown in Figure 2.5. Due to the small number of laser intercepts for each stands (approximately 30 non-sky hits for each site), the parametric profiles underestimate the canopy density, with no foliage or woody material estimated above 25m for any of the stands.

The regression-based profiles greatly improve upon this situation by assuming a smooth curve up to the height of the canopy. For stands with dense foliage close to the ground (e.g., Sites 3 and 6), the parametric and non-parametric profiles show the largest discrepancies. For stands with less dense foliage (e.g., Sites 1 and 2), the parametric and non-parametric profiles are much more similar, as the laser hits were able to penetrate further into the canopy.

The parametric LAI estimates reflect these qualitative differences in canopy profile form. The mixed conifer sites had the largest mean LAI estimate, with a value of 4.4 (95% CI=[1.7, 7.0]). The ponderosa pine sites had a substantially higher mean LAI estimate than the ponderosa mix sites, with a value of 4.1 (95% CI=[2.6, 5.6]) compared to 2.4 (95% CI=[1.4,3.4]).

The MaArthur-Horn estimates of LAI were substantially less than the corresponding parametric values (Figure 2.6). The empirical 5% and 95% quantiles for the parametric and non-parametric LAI estimates were (0.91, 9.03) and (0.12, 0.96), respectively, showing large differences in the LAI estimates between the two methods. A linear model regressing the non-parametric LAI values on the parametric LAI values yielded an R^2 value of 0.22.

