Horizontal Line Sampling for Riparian Forests without Land Area Estimation

Thomas B. Lynch

Abstract: Sampling forests in riparian areas presents unique difficulties that can be addressed by judicious application of line sampling. Land areas associated with riparian forests often have highly irregular boundaries, making accurate area determination difficult. Therefore, estimators for total amounts of forest attributes that do not depend on land area are desirable. Such estimators can be obtained by using horizontal line sampling from sample lines that are located at random intervals along a straight baseline and which extend entirely through the riparian area of interest. Horizontal line sampling selects sample trees that subtend an angle whose vertex is located on a length of line. This is conceptually similar to Ståhl’s transect relascope sampling (Ståhl, G. 1998). Transect relascope sampling—A method for the quantification of course woody debris. For. Sci. 44:58–63), which estimates total amounts of coarse woody debris from lines established at fixed intervals. Use of a control variate or importance sampling based on predicted line length (where available) is suggested as a variance reduction technique. An unbiased estimator for sampling at intervals along the line is suggested. Monte Carlo integration estimates of riparian land area are proposed for estimation of per-unit area quantities. These estimators could be applied to any forested area but appear to be most well suited to long, relatively narrow riparian areas. For. Sci. 52(2):119–129.

Key Words: Monte Carlo integration, control variate.

The importance of forested riparian areas has been increasingly recognized in recent decades, yet application of traditional forest inventory techniques can be difficult in these areas. Riparian areas are often characterized by irregular boundaries that make area determination difficult. Riparian areas formed by streams and river corridors may be long and narrow in shape. Traditional forest inventory techniques such as fixed-radius plot, horizontal point, and horizontal line sampling normally require land area estimates to expand per-unit area estimates to forest-level totals. In addition, long, narrow areas with a high ratio of perimeter to area are likely to result in many plot or point locations subject to boundary overlap (sometimes called “sloppover”), which, with traditional point or plot sampling, require specialized techniques such as the mirage method (Schmid-Haas 1969, Avery and Burkhart 2002, Gregoire 1982). Therefore, alternatives to traditional fixed-radius plot or horizontal point sampling may be desirable for long, narrow riparian areas formed by stream and river corridors.

A natural response to the problem of sampling a long, narrow land area might be placement of sample transects or lines roughly perpendicular to the long dimension of the area. This would be somewhat similar to the traditional “strip cruise” (Avery and Burkhart 2002), which selects trees located within a fixed distance of a centerline as sample trees. However, Avery and Burkhart (2002) note several disadvantages of strip cruising as traditionally applied. Also, the traditional application of strip cruising still results in per-unit area estimates that must be multiplied by land areas to obtain forest-level totals. A refinement of strip cruising that selects sampling trees by projecting an angle gauge from a centerline has been termed horizontal line sampling. Strand (1957) first proposed the concept of horizontal line sampling. Horizontal line sampling theory and technique have been further developed and explained by Groenbaugh (1958), Beers and Miller (1976), and Husch et al. (1982, 2003). Horizontal line sampling overcomes several of the disadvantages of strip cruising. Because an angle gauge such as a prism is used to select sample trees, it is not necessary to measure or check strip widths except when “borderline” trees are encountered. Horizontal line sampling selects trees with probability proportional to dbh, which may lead to efficiencies for certain forest populations. However, the traditional applications of horizontal line sample provide per-unit area estimates that must be multiplied by forest land area to obtain totals at the forest level. To avoid the difficulties inherent in land area estimation for long, narrow riparian areas formed by stream corridors, the objective of this study is to develop a variation of horizontal line sampling that does not require land area estimation.

Line Sampling for Riparian Areas

Traditional fixed-radius plot and point sampling methods select trees with probabilities that depend on plot areas. This approach results in estimates expressed as per-unit of land
area, requiring forest land area estimates to obtain forest-level totals. Independence from land area estimation can be achieved by selecting sample trees with probabilities that do not depend on land areas. This can be accomplished by establishing a baseline roughly parallel to the general course of the riparian corridor, as illustrated in Figure 1. Perpendiculars from the baseline endpoints should extend to the “left” and “right” boundaries of the riparian corridor, as indicated in the figure. Sample lines crossing the riparian corridor and perpendicular to the baseline will emanate from points selected randomly from the baseline. Therefore, the probability of sample tree selection will not be based on a plot area or land area, enabling estimation of totals without the necessity for land area estimation. This concept is similar to Ståhl’s (1998) method of transect relascope sampling for estimating amounts of coarse woody debris. Ståhl (1998) proposes transects at fixed intervals for estimating amounts of coarse woody debris. A tree of given dbh \( D \) will be selected if it is close enough to the sample line so that its dbh subtends the angle projected by the angle gauge used. Using the gauge constant \( K \) as defined by Grosenbaugh (1958) and Palley and Horwitz (1961), a tree must be within a limiting distance \( R \) of the sample line to be selected:

\[
R_i = K \frac{D_i}{2},
\]

(2)

\[
K = \frac{2R_i}{D_i} = \csc(\theta/2),
\]

(3)

where \( K \) is the gauge constant, \( \theta \) is the selection angle projected by the angle gauge, \( R_i \) is the limiting distance for selection to tree \( i \) in meters, and \( D_i \) is the dbh of tree \( i \) measured in meters.

Any tree \( i \) having dbh \( D_i \) whose perpendicular distance to the sample line is \( R_i \) or less will be selected as a sample tree, otherwise not. The probability that tree \( i \) is selected from the sample line emanating from randomly chosen point \( x_j \) (Figure 2) is

\[
P_{ij} = \int_{c_j-R_i}^{c_j+R_i} \frac{1}{B} \, dx = \frac{x}{B} \bigg|_{c_j-R_i}^{c_j+R_i} = \frac{2R_i}{B} = \frac{2KD_i}{B},
\]

(4)

where \( P_{ij} \) is the probability of selecting tree \( i \) from random location of \( x_j \). \( c_j \) is the point at which a perpendicular from tree \( i \) intersects the baseline, and \( B \) is the length of the baseline in meters.

Referring to Figure 2, note that one may project a perpendicular from the center of tree \( i \) to the baseline of length \( B \). Consider an interval of width \( 2R_i \) on the baseline centered at the intersection of a perpendicular from tree \( i \) to the baseline. The probability that tree \( i \) will be selected is equivalent to the probability that a point \( x_j \) selected at random from a uniform distribution defined on the baseline of length \( B \) falls within the interval of width \( 2R_i \). This probability is given by Equation 4. Any sample line emanating from a point on the baseline within the previously defined interval of width \( 2R_i \) will select tree \( i \) as a sample tree. Note that the probability of tree selection in Equation 4 does not depend on a plot area, as would be true in the development of traditional horizontal line sampling (e.g.,

\[
f(x) = \begin{cases} 1/B & 0 < x < B, \\ 0 & \text{otherwise}. \end{cases}
\]

(1)
Husch et al. 1982; Grosenbaugh 1958). Instead, the probability of tree selection is proportional to the length of a line segment $2R$ on the baseline of total length $B$. For this reason, estimates for forest totals can be made without using land area estimates. The baseline length $B$ may be thought to act as a surrogate for land area for the purpose of estimation of forest-level totals. The following estimator for total amounts of any tree attribute results:

$$
\hat{Y}_A = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} A_i \frac{B}{2K} \frac{I_i(x_j)}{D_i}.
$$

(5)

where $\hat{Y}_A$ is the estimated amount of attribute $A$; $A_i$ is the amount of attribute $A$ associated with tree $i$; $N$ is the total number of trees in the population; $I_i(x_j) = 1$ if $(c_i - R_i < x_j < c_i + R_i)$ so that tree $i$ is selected by line $j$, otherwise 0; and $n$ is the number of sample points $j = 1, \ldots, n$. In Appendix 1, it is demonstrated that estimator 5 is design-unbiased by using methods similar to those used by Palley and Horwitz (1961) to demonstrate the unbiasedness of point sampling.

When $A_i = 1$ in estimator 5, the estimator for total number of trees results. To obtain the estimator for total volume, let $A_i$ be the individual volume of tree $i$. When $A_i = D_i$ the following estimator for the total sum of diameters is obtained:

$$
\hat{Y}_D = \frac{B}{n2K} \sum_{j=1}^{n} \sum_{i=1}^{N} I_i(x_j).
$$

(6)

Because this estimator is not a function of tree measurements, it may be obtained from tree counts without the need to measure tree dbh except to determine the status of borderline trees. This property has been noted for the horizontal line sampling estimator by Grosenbaugh (1958) and Husch et al. (1982). It may be of interest to note, as indicated by Grosenbaugh (1958) and Husch et al. (1982), that the sum of dbhs is proportional to the sum of tree circumferences.

**Variance of the Estimator**

The following variance of estimator 5 is derived in Appendix 1:

$$
\text{var}(\hat{Y}_A) = \frac{1}{n} \left( \frac{n}{\sum_{i=1}^{N} B \times A_i^2} - \left( \frac{\sum_{i=1}^{N} A_i}{n} \right)^2 \right) + 2 \sum_{i<k} \frac{B \times A_i A_j}{4R_i R_j} (O_{i,k}),
$$

(7)

where $O_{i,k}$ is the overlap length between the sampling intervals $(R_{c_k} - c_k, R_{c_k} + c_k)$ and $(R_k - c_k, R_k + c_k)$ on the baseline of length $B$. To estimate the variance of estimator 5, note that estimator 5 can be expressed as a mean based on $n$ independent sample lines,

$$
\hat{Y}_A = \frac{1}{n} \sum_{j=1}^{n} Y_{Aj} = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{B \times A_i}{2K \times D_i} I_i(x_j),
$$

(8)

Figure 2. Selection of sample tree $i$ using randomly located point $x_j$ on baseline of length $B$.

Figure 3. Stratified sampling for curved riparian area with baselines of length $B_1$ and $B_2$ showing termination of a sample line from $B_2$ at the boundary between strata 1 and 2.
where

\[ Y_M = \frac{B}{2K} \sum_{i=1}^{K} A_i I_y(x_i). \]  

(9)

Therefore, an estimator of the sample variance among lines may be calculated as

\[ S^2_{Y_a} = \frac{\sum_{i=1}^{n} (Y_M - \bar{Y}_a)^2}{(n - 1)}. \]  

(10)

An estimator of the variance for estimator 5 is then

\[ S^2_{Y_a} = \frac{S^2_{Y_a}}{n}. \]  

(11)

Equation 7 indicates that the variance depends on part on the length of the baseline B. Inspection of Figure 1 suggests that the length of the baseline B could be shortened by increasing the angle between B and the direction of the main riparian channel (that is, using a baseline that is not parallel to the direction of the main channel). However, this would tend to lengthen the sample lines and increase the overlap areas O_{ik}, which may have a compensating effect on the variance. Increasing sample line length also results in an increase in time and cost per line. Thus, reduction in the length of B by this method is not likely to increase sampling efficiency.

**Per-Unit Area Estimates**

Riparian forest estimates can be obtained on a per-unit area basis by dividing Estimator 5 by an estimator of land area,

\[ \hat{Y}_{AT} = \frac{\hat{Y}_A}{\hat{T}}, \]  

(12)

where \( \hat{T} \) is an estimator of the land area located in the riparian region that has been sampled. The concept of Monte Carlo integration (e.g., Rubenstein 1981) can be used to obtain the following unbiased land area estimator, if alternative land area estimates are not available:

\[ \hat{T} = \frac{B}{n} \sum_{i=1}^{n} L_j, \]  

(13)

where \( L_j \) is the length of the randomly selected sample line \( j \) within the riparian corridor. Because the \( L_j \) are independent random samples, an estimator of the variance for this land area estimator is

\[ S^2_{T} = \frac{B^2 S^2_{Y_A}}{n}. \]  

(14)

where

\[ S^2_{Y_A} = \frac{\sum_{i=1}^{n} (Y_A - \bar{Y}_A)^2}{n - 1}. \]  

(15)

However, this Monte Carlo integration estimator of land area may not always provide the best estimate available. Its variance depends on the variation in the width of the riparian corridor and the number of sample lines established. Use of Monte Carlo integration for land area estimation has been discussed by Gregoire, T.G. and H.T. Valentine, Sampling strategies for natural resources and the environment (in press). If an approximate land area is available from a map, aerial photograph, or other source, the following estimator could be used to obtain an unbiased estimator of riparian land area with reduced variance:

\[ \hat{T}_C = \frac{1}{n} \sum_{i=1}^{n} \left[ B \times L_j - \beta (B \times L_{Mj} - T_M) \right], \]  

(16)

where \( \hat{T}_C \) is an unbiased estimator of riparian land area using control variates, \( \beta \) is a constant, \( L_{Mj} \) is the length of line \( j \) according to the map, and \( T_M \) is the riparian land area according to the map. Estimator (16) uses the concept of control variates (Rubenstein 1981) to reduce the variance in the Monte Carlo integration estimate of riparian land area. It might be useful for long, narrow riparian areas with complex boundaries that may be too long (e.g., dozens of kilometers) to efficiently walk with a GPS unit. Because \( E(B \times L_{Mj} - T_M) = 0 \), Estimator 16 is also unbiased. Even though a “model” in the form of a map predicting area \( T_M \) is used to reduce the variance of the area estimator, the estimator has the characteristics of design-unbiasedness, because unbiasedness does not depend on the model (map) form. Rubenstein (1981) gives an optimal value for the constant \( \beta \) that depends on the variance of \( L_{Mj} \) and the covariance between \( L_j \) and \( L_{Mj} \). One might attempt to estimate \( \beta \) based on prior empirical studies, but these could not involve data from the current sample lines in the riparian area of interest. In the absence of better information, a value of \( \beta = 1 \) could be used.

Estimator 12 for per-unit area value is essentially a ratio of means estimate. Williams (2001) has indicated that ratio estimators in plot sampling can exhibit biases, especially when substantial edge-effect (that is, when conditions in the interior differ from those near the boundary) is present. The variance of a ratio estimator is often estimated using the standard approximation (e.g., Schreuder et al. 1993),

\[ S^2_{Y_{AT}} = \frac{1}{n \bar{T}^2} \left( S^2_{Y_A} + \hat{Y}_{A}^2 B^2 S^2_{L} - 2B \hat{Y}_{A} S_{AT,L} \right), \]  

(17)

where \( S_{AT,L} \) estimates the covariance between line length \( L_j \) and \( Y_{AT} \),

\[ S_{AT,L} = \frac{\sum_{i=1}^{n} (L_j - \bar{L})(Y_{Aij} - \hat{Y}_A)}{n - 1}. \]  

(18)

This latter covariance will normally be zero if the land area estimate is not obtained by using sample line lengths as from Equation 13, because in that case the land area estimate will be independent of the total forest attribute estimate. Williams (2001) indicates that this variance approximation can perform poorly with plot sampling, especially in the presence of edge-effect.
Boundary Overlap and Slope Correction

Most of the boundary associated with the long, narrow riparian areas formed by streams and rivers will roughly parallel the water course. This does not present a problem because Estimator 5 does not depend on sample line length. The decision regarding inclusion of sample trees depends on whether the perpendicular location of the tree as projected on the baseline of length $B$ falls within the required interval. In the field, one simply extends the sample line beyond the tract boundaries selecting as sample trees any that qualify and are within the tract boundaries. Relatively smaller quantities of boundary occur on the “left” and “right” endpoints of the sample baseline of length $B$. Randomly located lines that fall near the left or right endpoints of the baseline of length $B$ may be subject to boundary overlap. Perhaps the simplest way to handle this situation is to extend the length of the baseline $B$ by an amount $2R^2 = 2KD^2$, where $D^2$ is as large as or larger than the largest tree dbh expected. The length of this extension should be an equal amount $R^2$ on each end of the baseline. From any sample lines that may fall outside the left or right tract boundaries, sample qualifying trees within the boundary. A disadvantage of the latter procedure is that added baseline length should increase the variance of the estimator. Alternatively, one may choose to sample trees only on the interior side of the sample line when the line falls within a specified distance of the left or right baseline endpoint, as suggested by Grosenbaugh (1958) and Beers and Miller (1976). Trees selected by the latter procedure are tallied twice. A third alternative would be similar to the mirage method often used in point sampling (Schmid-Hass 1969, Gregoire 1982). For sample line starting points near the baseline endpoint, one establishes a corresponding sample line on an extension of the baseline at a distance from the endpoint equal to that of the original sample line from the endpoint on the “other side.” Trees tallied from this mirage line that are inside the tract boundary are added to the tally from the original line. As a result, some trees close to a line perpendicular to the baseline and emanating from the endpoint are tallied twice for this line. A fourth alternative would be an adaptation of the walkthrough method (Ducey et al. 2004). It is important to note that, because the population can be thought of as being “projected” onto the baseline, the walkthrough method should be conducted in a direction parallel to the baseline and the “boundary” will be a perpendicular to the baseline emanating from the endpoint, a line that may or may not be the “real” tract boundary. For example, the left boundary of stratum 2 in Figure 3 is not perpendicular to the baseline. However, because trees inside this boundary are projected onto the baseline, the walkthrough evaluation is made with respect to the left baseline endpoint. Of course, trees that lie outside the real tract boundary are not tallied at all. To execute the walkthrough method, one measures the distance from the sample line to the tree in question (which qualifies as a sample tree from that line) along a parallel to the baseline. If this distance is greater than the distance along the same parallel from the tree to a perpendicular passing through the baseline endpoint, the tree is tallied twice. Otherwise, it is tallied only once.

Slope correction may be accomplished by prism rotation, as is commonly done for horizontal point sampling (e.g., Husch et al. 1982). The procedure suggested by Beers and Miller (1976) of laying out horizontal line sample segments along slope cannot be used for the current application. The sample line should remain perpendicular to the baseline regardless of terrain.

**Stratification for Curvature and Branching**

It should not be necessary to alter the baseline on account of minor curvature of streams or rivers. From a theoretical point of view, it would be possible to accommodate any degree of curvature by using a single baseline of appropriate length. However, this may not be practical or desirable. Extensive curvature of stream channels may result in sample lines that are nearly coincident with the direction of the stream channel, and therefore of excessive length. Estimators that use such long sample lines are still unbiased, but these lines may be time-consuming and difficult to measure in the field. Furthermore, a mixture of long and short sample lines may tend to increase the variance of the estimator.

Stratification for curvature of riparian areas is illustrated in Figure 3, where the area is stratified for sampling from two baselines, $B_1$ and $B_2$, which are roughly parallel to the stream channel within each stratum. Separate simple random samples are obtained using Estimator 5 in each stratum, leading to estimates for totals in strata 1 and 2 that can simply be added. This can be extended to any finite number $S$ of strata,

$$\hat{Y}_A = \sum_{k=1}^{s} \hat{Y}_{Ak} = \sum_{k=1}^{s} \frac{B_k}{n_k 2R_k} \sum_{i=1}^{n_k} A_{ki} \frac{I_{kj}(x_{ki})}{D_{ki}},$$

where $\hat{Y}_{Ak}$ is the estimated amount of attribute $A$ in stratum $k$; $A_{ki}$ is the amount of attribute $A$ associated with tree $i$ in stratum $k$; $D_{ki}$ is the diameter of tree $i$ in stratum $k$; $N_k$ is the total number of trees in the population for stratum $k$; $R_k$ and $c_k$ are defined as $R_k$ and $c_k$ above for each stratum $k$; $I_{kj}(x_{ki}) = 1$ if $(c_{ki} - R_k) < x_{kj} < (c_{ki} + R_k)$ so that tree $i$ is selected by line $j$ in stratum $k$, otherwise 0; and $n_k$ is the number of stratum $k$ sample points $j = 1, \ldots, n_k$.

Sampling in each stratum is independent of sampling in other strata, so that the variance of the estimate for the total is the sum of the stratum variances,

$$\text{var}(\hat{Y}_A) = \sum_{k=1}^{s} \text{var}(\hat{Y}_{Ak}).$$

Equation 7 can be used to obtain $\text{var}(\hat{Y}_{Ak})$ for each stratum $k$ because a simple random sample is conducted within each stratum. Estimates of each stratum variance may be computed from Equation 11. Stratified sampling is discussed by Schreuder et al. (1993).

Special attention should be paid to the boundary between adjacent strata indicated by the boundary between stratum 1

**Forest Science 52(2) 2006 123**
and stratum 2 in Figure 3. Because the baselines associated with these strata are not parallel, a sample line near the left end of stratum 2 may intersect the boundary between strata 1 and 2. In this case, no trees from stratum 1 should be sampled for the line emanating from baseline 2, and the methods suggested above for handling boundary overlap should be applied. It may be desirable to mark visibly the boundaries between strata in the field. It is conceivable that, in some situations, GIS and GPS technology could be used to determine a portion of the boundary that may intersect a particular sample line in the field without locating the entire stratum boundary in the field. Note that baseline 2 in Figure 3 should be located so that there is no riparian area that could not be sampled while minimizing overlap of sample area with that of baseline 1. The best location would be found by extending the left endpoint of baseline 2 to a perpendicular that extends to the intersection of the baseline 1 stratum boundary with the riparian area (see Figure 3). This might be determined using a map, GIS layer, or in the field. Where the precise location of the left endpoint is uncertain, it could be extended further leftward. However, this could result in the establishment of null sample lines near the left endpoint of the baseline on which no trees are sampled. Such lines present no theoretical problems but would tend to increase the sample variance for that stratum, as would the unnecessarily large value of the baseline.

Figure 4 illustrates stratification for a riparian area showing a branching pattern. Strata 1 and 2 sample their respective branches from baselines 1 and 2. Stratum 3 begins where the branches join. This procedure would provide separate estimates of totals for each branch that could be summed to provide an estimate of total for the whole system. Where branches exhibit substantial curvature, the orientation of baselines could be adjusted as suggested above. It would be theoretically possible to sample the whole system of Figure 4 with one long baseline. However, this would not provide separate estimates of totals for each branch. Equation 19 could be used with $S = 3$ to estimate totals for the situation of Figure 4.

\[ B_2 = \text{baseline 2 length} \]

\[ B_3 = \text{baseline 3 length} \]

\[ B_1 = \text{baseline 1 length} \]

Figure 4. Stratified sampling for branching riparian features with baselines of lengths $B_1$, $B_2$, and $B_3$.

**Variance Reduction**

Where approximations of riparian area are available from maps, aerial photographs, or other sources, the following control variate estimator $\hat{Y}_{AC}$ could be used to reduce the variance of Estimator 5:

\[ \hat{Y}_{AC} = \frac{1}{n} \sum_{j=1}^{n} \left[ Y_{Aj} - \beta (B \times L_{Mj} - T_M) \right] \]

\[ = \frac{1}{n} \sum_{j=1}^{n} \left[ \frac{B \times A_i}{2K \times D_i} I_{ij} - \beta (B \times L_{Mj} - T_M) \right]. \]

(21)

where $\beta$ is a constant. This is essentially the same control variate previously proposed for land area estimation. Rubenstein (1981) gives an optimal value for $\beta$ that depends on the covariance between $Y_{Aj}$ and $L_{Mj}$ as well as the variance among the $L_{Mj}$. It would probably have to be estimated via empirical studies, but to maintain unbiasedness, the estimate of $\beta$ would have to be before collection of data in the riparian area of interest. One would expect positive correlation between $L_{Mj}$ and $Y_{Aj}$ because longer lines $L_{Mj}$ from the map might be expected to have larger estimates $Y_{Aj}$. The estimator is unbiased for any constant value of $\beta$, and as indicated for the control variate Estimator 16 of land area, should have the characteristics of design-unbiasedness. Note that the actual line lengths $L_i$ were not used here because their true mean is not known. Only the map area $T_M$ is known, which here is assumed to differ from the actual riparian land area.

Other variance reduction strategies suggested by Rubenstein (1981) could also be tried. An importance sampling scheme could be developed using the following probability
density based on the line lengths predicted from a map:

\[ p(x) = \frac{L_M(x)}{T_M} \quad 0 < x < B, \]  

(22)

where \( L_M(x) \) is the line length predicted by the map of area \( T_M \) at point \( x \) on the baseline of length \( B \). Note that

\[ \int_0^B p(x) \, dx = \int_0^B \frac{L_M(x)}{T_M} \, dx = \frac{T_M}{T_M} = 1 \quad 0 < x < B \]  

(23)

The importance sampling estimator \( \hat{Y}_{AI} \) based on this density function is

\[ \hat{Y}_{AI} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N A_i \frac{I_{ij}(x_i)}{2K \times D_i \times p(x_j)} \]

\[ = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N A_i \frac{T_M I_{ij}(x_i)}{2K \times D_i \times L_M(x_j)}. \]  

(24)

The unbiasedness of this estimator is demonstrated in Appendix 1. This is similar to the procedure suggested by Gregoire T.G. and H.T. Valentine (in preparation) for variance reduction with importance sampling to estimate coarse woody debris with line sampling.

Selection of sample lines from the importance sampling density \( p(x) \) is more complex than simple selection at random from the interval \((0, B)\) as was done for Estimator 5. The inverse transform method, often used with importance sampling for tree stem volumes, would be difficult to use here because there generally would be no algebraic equation form to integrate and invert. Conceivably one might select a number at random from the interval \((0, B)\) and within a GIS system compute partial map areas using a straight line perpendicular to the baseline as the right-hand boundary. The point on the baseline where the ratio of partial map area to total map area equals the selected random number would be the starting point for a sample line with importance sampling. A better alternative would probably be von Neumann’s (1951) acceptance–rejection method (Rubenstein 1981), which is outlined in Appendix 2 and was proposed by Gregoire T.G. and H.T. Valentine (in preparation) for use with line intercept sampling of coarse woody debris. Importance sampling would be more likely to select wider sample lines, which should tend to contain more sample trees. To the extent that sample line content (e.g., volume) is correlated with sample line length, importance sampling should result in variance reduction.

Either control variate or importance sampling estimators have the potential to result in variance reduction where approximate riparian area maps can be obtained. Either variance reduction technique will provide unbiased estimates even though the approximate riparian area map may be inaccurate. A possible disadvantage of importance sampling is that longer sample lines should also be more time-consuming and expensive to sample. Selection of random sample line locations is simpler when control variates are used. For these reasons, control variates may have advantages for variance reduction when approximate maps of riparian areas are available.

**Line Sampling at Intervals**

The width of some riparian corridors may make it desirable to sample at intervals along the sample line rather than continuously along the entire line. This could be done with a “systematic random start” procedure on each line. That is, the starting point for tree selection could begin at a randomly selected point on the sampling line based on the sampling interval. Sampling could then be done at fixed intervals along the line (Figure 5). The conditional probability of selection for tree \( i \) given a point \( c_i - R_i < x_j < c_i + R_i \) selected from the baseline of length \( B \) is then

\[ P(I^*_{ij}(x_j) = 1|I_j(x_j) = 1) = \frac{L_{Sj}}{L_j}, \]  

(25)

where \( L_{Sj} \) is the length of line sampled emanating from point \( x_j \) on the baseline; \( L_j \) is the length of line emanating from point \( x_j \) lying with the riparian area; \( I^*_{ij}(x_j) = 1 \) if tree \( i \) is sampled on the length of line \( L_{Sj} \); otherwise 0; and \( P(I^*_{ij}(x_j) = 1|I_j(x_j) = 1) \) is the probability that tree \( i \) will be chosen along \( L_{Sj} \) given that \( x_j \) has been selected on the baseline so that \( c_i - R_i < x_j < c_i + R_i \). Note that

\[ P(I^*_{ij}(x_j) = 1|I_j(x_j) = 0) = 0. \]  

(26)

This suggests the following estimator,

\[ \hat{Y}_{AI}^* = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^N L_j \times B \times A_i \frac{L_{Sj}}{L_j} \frac{I_j(x_j)I^*_{ij}(x_j)}{2K \times D_i \times L_M(x_j)} . \]  

(27)

In Appendix 1, it is demonstrated that the above estimator for line sampling at intervals is design-unbiased if the ratio of sampled to total line length \( L_{Sj}/L_j \) is a constant within the interval for conditioning \( c_i - R_i < x_j < c_i + R_i \). Note that it is not necessary that line lengths are constant (they would not be), only that the ratio between sampled and total line length is constant. This probably would not be precisely true in all applications but would be more nearly true for the wide riparian channels for which using sample intervals might be desirable. In cases where it is possible to measure sample line lengths before implementation in the field (using an accurate map or GIS system) it would be possible to ensure a constant ratio of sampled-to-total line length \( L_{Sj}/L_j \) by design. Where an accurate map is not available, it might be feasible to place two field crew members with GPS equipment on opposite sides of the riparian corridor to measure line length before sample tree selection.

To facilitate estimation of the standard error, Estimator 27 can be written as an average of estimators associated
with each sample line:

$$\hat{Y}_A = \frac{1}{n} \sum_{j=1}^{n} Y_{A_j} = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{L_j \times B \times A_j}{2K \times D_i} I_j(x_j)I_{A_j}(x_j),$$

(28)

where

$$Y_{A_j}^* = \frac{B}{2K} \sum_{i=1}^{N} \frac{L_A}{L_{D_i}} I_j(x_j)I_{A_j}(x_j).$$

(29)

Then an estimate of the sample variance among lines may be calculated as

$$S^2_{Y_{A_j}} = \frac{1}{n-1} \sum_{j=1}^{n} (Y_{A_j}^* - Y_{A_j})^2.$$  

(30)

The sample variance for Estimator 26 can be estimated as

$$S^2_{Y_{A_j}} = \frac{S^2_{Y_{A_j}}}{n}.  

(31)

If the ratio of sampled to total line length $L_{A_j}/L_j$ cannot be assumed to be constant for all $j$, then Estimator 27 is essentially similar to a mean of ratios estimator, because the ratio of sampled-to-total line length would then be a random variable. In this case the form of the estimator would be the same as presented in Equation 27. However, the demonstration of unbiasedness given above would no longer be valid. Schreuder et al. (1993) have discussed general applications of the mean of ratios estimator. Williams (2001) found that a ratio of means estimator performed better than a mean of ratios estimator with plot sampling, but either could exhibit bias in the presence of edge effect. This could provide motivation to implement the design-unbiased estimator by keeping the ratio $L_{A_j}/L_j$ constant. A ratio of Estimator 27 to riparian land area can be used to develop per-acre estimates of riparian area attributes, as was done above in Estimator 12.

**Discussion and Conclusions**

Estimators based on horizontal line sampling have been proposed for trees sampled in riparian areas. However, they differ from traditional horizontal line sampling estimators in that the probability of tree selection is not based on a two-dimensional plot land area, but rather is developed by selecting points from a one-dimensional baseline. Thus, estimates of totals for forest attributes in riparian areas can be developed without the need to estimate riparian land area. It has been indicated that land area could be estimated by Monte Carlo integration using lengths of the randomly located sample lines. Estimator 5 could also be viewed as an example of Monte Carlo integration. One could express the total amount of a forest attribute $A$ as an integral with respect to baseline $B$. The sample lines could then be viewed as random samples of the integrand. Valentine et al. (2001) have shown that many forest sampling techniques can be viewed in terms of Monte Carlo integration. Except for the control variate and importance sampling estimators, the estimators here correspond to crude Monte Carlo integration. Crude Monte Carlo integration can be considered a special case of importance sampling. Monte Carlo integration approaches to certain forest sampling problems have been discussed by Schreuder et al. (1993).

For some riparian areas, increased precision might be obtained by using line sampling as indicated above with a stratified sampling design as was suggested above for riparian areas exhibiting curvature and branching. This would ensure sampling in all subareas delineated as strata along an extensive riparian area. Strata would not be land areas, but rather subintervals on the baseline of length $B$ used to locate sample lines. Therefore, estimates of population totals would not require computation of land area, but would depend on subinterval lengths used in stratified sampling. Sample means and variances within each stratum could be developed as indicated above. These estimates could be combined to obtain overall means and standard errors for the stratified sample.

If it is desired to obtain estimates of coarse woody debris in riparian areas, Ståhl’s (1998) transect relascope sampling could be applied from the same lines used for horizontal line sampling of standing trees. The methods suggested for variance reduction above, including control variates and importance sampling, should be applicable to transect relascope sampling.

These estimators avoid the difficulties that may be involved when attempting to estimate land areas for long, narrow, and irregularly shaped riparian areas. However, in principle, Estimators 5 and 27 could be used for forested areas having any shape. Estimators 5 and 27 provide alternatives to traditional land-area-based sample plots when measurement and/or computation of land areas are difficult or inaccurate.

**Literature Cited**


STÅHL, G. 1998. Transect relascope sampling—A method for the
to derive the variance of the horizontal point sampling
approach similar to that used by Palley and Horwitz (1961)
Variance of the Line Sampling Estimator
Appendix 1. Unbiasedness and Variance
HUSCH, B., C.I. MILLER, AND T.W. BEERS. 1982. Forest mensu-
meling given by Palley and Horwitz (1961).
Forest menseling methods for multiresource forest inventory. John Wiley
and systematic point sampling estimators. For. Sci.
STÅHL, G. 1998. Transect relascope sampling—A method for the
STRAND, L. 1957. “Relaskopisk høyde”-og kubikkmasesbestem-
melse (Relasscopic height and cubic volume determination).
Carlo approaches to sampling forested tracts with lines or
VON NEUMANN, J. 1951. Various techniques used in connection
WILLIAMS, M.S. 2001. Performance of two fixed-area (quadrat)
sampling estimators in ecological surveys. Environmetrics
12:421–436.

Appendix 1. Unbiasedness and Variance
Derivations
Unbiasedness of the Line Sampling Estimator
Estimator 5 is design-unbiased where \( P_{ij} \) is defined by
Equation 4,

\[
E(\hat{Y}_A) = \left[ \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{A_i}{P_{ij}} I_j(x_i) \right] = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{A_i}{P_{ij}} E(I_j(x_i))
\]

\[
= \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{A_i}{P_{ij}} P_{ij} = \sum_{i=1}^{N} A_i
\]

(A1)

and the last term on the right above is the total amount of
attribute \( A \) in the forest population. This is similar to the
demonstration of the unbiasedness of horizontal point sam-
ppling given by Palley and Horwitz (1961).

Variance of the Line Sampling Estimator
The variance of Estimator 5 can be obtained by using an
approach similar to that used by Palley and Horwitz (1961)
to derive the variance of the horizontal point sampling
estimator,

\[
\text{var}(\hat{Y}_A) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{A_i}{P_{ij}} I_j(x_i) \left( \frac{1}{n} \sum_{j=1}^{n} \frac{A_i}{P_{ij}} \right) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{A_i}{P_{ij}} \text{var}(I_j(x_i))
\]

\[
= \frac{1}{n} \left[ \sum_{j=1}^{n} \frac{A_i^2}{P_{ij}} \text{var}(I_j(x_i)) \right] + 2 \sum_{i<k} \frac{A_i A_k}{P_{ij} P_{kj}} \text{cov}(I_j(x_i), I_k(x_j))
\]

(A2)

Because the indicator variable \( I_j(x_i) \) is a Bernoulli random
variable,

\[
\text{var}(I_j(x_i)) = P_{ij}(1 - P_{ij}),
\]

\[
\text{cov}(I_j(x_i), I_k(x_j)) = E(I_j(x_i)I_k(x_j)) - E(I_j(x_i))E(I_k(x_j))
\]

\[
= E(I_j(x_i)I_k(x_j)) - P_{ij}P_{kj}.
\]

(A4)

To evaluate this latter covariance, note that

\[
E(I_j(x_i)I_k(x_j)) = 1 \times 1 \times P(I_j(x_i) = 1, I_k(x_j) = 1)
\]

\[
+ 0 \times 0 \times P(I_j(x_i) = 1, I_k(x_j) = 0)
\]

\[
+ 0 \times 1 \times P(I_j(x_i) = 0, I_k(x_j) = 1)
\]

\[
+ 0 \times 0 \times P(I_j(x_i) = 0, I_k(x_j) = 0)
\]

\[
= P(I_j(x_i) = 1, I_k(x_j) = 1).
\]

(A5)

This last probability is the probability that both trees \( i \) and
\( k \) are sampled by sample line \( j \). This will only occur if \( x_j \)
falls in the overlap length between the sampling intervals
\( (R_k - c_k, R_k + c_k) \) and \( (R_i - c_i, R_i + c_i) \) on the baseline
of length \( B \). Let this overlap length be \( O_{ik} \). Then the expected
value is

\[
E(I_j(x_i)I_k(x_j)) = P(I_j(x_i) = 1, I_k(x_j) = 1) = O_{ik}/B.
\]

The covariance Equation 1.4 is then

\[
\text{cov}(I_j(x_i), I_k(x_j)) = \frac{O_{ik}}{B} - P_{ij}P_{kj} = \frac{O_{ik}}{B} - \frac{4R_kR_i}{B^2}.
\]

(A7)

Substitution results in

\[
\text{var}(\hat{Y}_A) = \frac{1}{n} \left[ \sum_{j=1}^{n} \frac{A_i^2}{P_{ij}} P_{ij} \left( 1 - P_{ij} \right) \right] + 2 \sum_{i<k} \frac{A_i A_k}{P_{ij} P_{kj}} \left( \frac{O_{ik}}{B} - \frac{4R_kR_i}{B^2} \right)
\]

\[
= \frac{1}{n} \left[ \sum_{j=1}^{n} \frac{A_i^2}{P_{ij}} P_{ij} \right] - \sum_{i=1}^{N} \frac{A_i^2}{P_{ij}} + 2 \sum_{i<k} \frac{A_i A_k}{P_{ij} P_{kj}} \left( \frac{O_{ik}}{B} \right) + 2 \sum_{i<k} \frac{A_i A_k}{P_{ij} P_{kj}} \left( \frac{4R_kR_i}{B^2} \right).
\]
noting that
\[
\left( \sum_{i=1}^{N} A_i^2 \right)^2 = \left( \sum_{i=1}^{N} A_i^2 + 2 \sum_{i<k} A_i A_k \right),
\]
and substitution for the probabilities of selecting trees \(i\) and \(k\) result in the following expression for the variance of Estimator 5:
\[
\text{var}(\hat{Y}_i) = \frac{1}{n} \left( \sum_{i=1}^{N} \frac{B \times A_i^2}{2R_i} \right) - \frac{1}{n} \left( \sum_{i=1}^{N} A_i \right)^2 + 2 \sum_{i<k} \frac{B \times A_i A_k}{4R_i R_j} (O_{ij}). \tag{A8}
\]

**Unbiasedness of the Importance Sampling Estimator**

The expected value of the importance sampling Estimator 24 with respect to the density \(p(x)\) where \(0 < x < B\):
\[
I_j(x_j) = 1 \quad \text{if} \quad (c_i - R_i) < x_j < (c_i + R_i), \quad \text{otherwise} \quad 0; \quad \text{and} \quad R_i = K \times D_i.
\]
\[
E(\hat{Y}_{Ai}) = E \left\{ \frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} I_j(x_j) \right\} = \frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} \int_{0}^{B} I_j(x_j) \frac{p(x)}{p(x)} \, dx_j
\]
\[
\frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} \int_{c_i-R_i}^{c_i+R_i} I_j(x_j) \, dx_j
\]
\[
\frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} \int_{c_i-R_i}^{c_i+R_i} I_j(x_j) \, dx_j
\]
\[
\frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} 2R_i
\]
\[
\frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{N} \frac{A_i}{2K \times D_i} 2K \times D_i = \sum_{i=1}^{N} A_i. \tag{9}
\]

Thus the estimator \(\hat{Y}_{Ai}\) is unbiased for the sum of individual tree attributes \(A_i\).

**Unbiasedness of the Estimator for Line Sampling at Intervals**

To investigate the properties of Estimator 27 for line sampling at intervals with respect to unbiasedness,
\[
E(\hat{Y}_i^*) = \sum_{i=1}^{N} E \left\{ \frac{L_{BA_i}}{L_{DS_j}(2K)D_i} I_j(x_j) \bigg| I_j(x_j) \right\}
\]
\[
= \sum_{i=1}^{N} \sum_{j=0}^{1} \frac{L_{BA_i}}{L_{DS_j}(2K)D_i} I_j(x_j)k
\]
\[
\times P[I_j(x_j) = k \big| I_j(x_j)]
\]
\[
= \sum_{i=1}^{N} \left\{ \frac{L_{BA_i}}{(2K)D_i} I_j(x_j) \right\}
\]
\[
= \sum_{i=1}^{N} \left\{ \frac{BA_i}{(2K)D_i} \times 1 \times P(I_j(x_j) = 1) \right\}
\]
\[
= \sum_{i=1}^{N} \left\{ \frac{BA_i}{(2K)D_i} \times 1 \times \frac{(2K)D_i}{B} \right\}
\]
\[
= \sum_{i=1}^{N} \left\{ \frac{BA_i}{(2K)D_i} \times 0 \times \frac{(2K)D_i}{B} \right\}
\]
\[
= \sum_{i=1}^{N} A_i.
\]

This indicates that Estimator 27 is design-unbiased. This argument assumes that the ratio of sampled to total line length \(L_{DS_j}/L_j\) is a constant, at least within the interval for conditioning \(c_i - R_i < x_j < c_i + R_i\).

**Appendix 2: von Neumann’s Acceptance–Rejection Method**

The method involves expression of the importance sampling probability density function (PDF) as
\[
p(x) = Ch(x)g(x) \quad 0 < x < B,
\]
where
\[
C \geq 1, \tag{B1}
\]
\(h(x)\) is a PDF, and \(0 < g(x) = 1\).

Following Algorithm AR-2 (Rubenstein 1981) for this application, define
\[
h(x) = \frac{1}{B}, \quad C = MB, \quad g(x) = \frac{p(x)}{M}, \tag{B2}
\]
\[M \geq \max(p(x)) = \max \left[ \frac{L_M}{M} \right].\]
Note that $M$ is defined to comply with the requirement of the method that $0 < g(x) \leq 1$. $M$ can be obtained by noting the longest possible line length $L_M$ according to the map. Now to generate random variables from $p(x)$, implement the four steps in the following algorithm:

1. Generate uniform random variables $U_1$ and $U_2$ from a uniform distribution on $(0, 1)$
2. $Y = U_2 \times B$
3. If $U_1 \leq g(Y)/M = p(U_2 \times B)/M = L_M(U_2 \times B)/(M \times T_M)$, then accept $Y$ as a random variate from $p(x)$
4. Proceed to step 1.

This algorithm could be executed on an ordinary map or aerial photograph on which the baseline of length $B$ has been drawn. It is conceivable that the process could be automated within a GIS system.