

TABLE OF CONTENTS

PART I: TUTORIAL PAPERS

Combining the Error of Sample Plots and Biomass Regressions

- Error of forest inventory estimates: its main components 1
Tiberius Cunia
- An optimization model to calculate the number of sample trees and plots 15
Tiberius Cunia

Error of Biomass Regressions

- Construction of tree biomass tables by linear regression techniques 27
Tiberius Cunia
- Use of dummy variables techniques in the estimation of biomass regressions 37
Tiberius Cunia
- On the error of tree biomass regressions: trees selected by cluster sampling and double
sampling 49
Tiberius Cunia

Error of Sample Plots

- On the error of forest inventory estimates: stratified sampling and double sampling for
stratification 63
Tiberius Cunia
- On the error of forest inventory estimates: two-stage sampling of plots 71
Tiberius Cunia
- On the error of forest inventory estimates: double sampling with regression 79
Tiberius Cunia
- On the error of forest inventory estimates: Continuous Forest Inventory without SPR 89
Tiberius Cunia
- On the error of forest inventory estimates: Continuous Forest Inventory with SPR 99
Tiberius Cunia

PART II: RESEARCH PAPERS

Biomass Regressions and Measurement Error

- An optimization model for subsampling trees for biomass measurement 109
Tiberius Cunia
- Estimating sample tree biomass by subsampling: some empirical results 119
R. D. Briggs, T. Cunia, E. H. White, and H. W. Yawney
- Unbiased estimation of total tree weight by three-stage sampling with probability
proportional to size 129
Harry T. Valentine, Timothy G. Gregoire, and George M. Furnival
- Measurement errors in forest biomass estimation 133
Daniel Auclair

Biomass of Forest Understory Vegetation

- Biomass-dimension relationships of understory vegetation in relation to site and stand
age 141
Paul B. Alaback

TABLE OF CONTENTS

Biomass estimates for nontimber vegetation in the Tanana River Basin of Interior Alaska	149
Bert Mead, John Yarie, and David Herman	
<u>Biomass Functions in the Eastern United States: Regression Models and Application to Timber Inventories</u>	
A summary of equations for predicting biomass of planted southern pines	157
V. C. Baldwin, Jr.	
Summary of biomass equations available for softwood and hardwood species in the southern United States	173
Alexander Clark III	
Methods for estimating the forest biomass in Tennessee Valley Region	189
J. Daniel Thomas and Robert T. Brooks, Jr.	
Areas of biomass research ¹	193
Boris Zeide	
<u>Biomass Studies Outside the United States</u>	
Prediction error in tree biomass regression functions for western Canada	199
T. Singh	
Forest biomass studies in France	209
Daniel Auclair	
Biomass studies in Europe - an overview	213
Dieter R. Pelz	
Subsampling trees for biomass	225
C. Kleinn and D. R. Pelz	
Simple biomass regression equations for subtropical dry forest species	229
Joseph D. Kasile	
<u>Use of Simulation Techniques to Evaluate the Validity of Biomass Regression Functions</u>	
Evaluating errors of tree biomass regressions by simulation	235
Tiberius Cunia	
Estimation of tree biomass tables by cluster sampling: results of a simulation study	243
Andrew J. Gillespie and Tiberius Cunia	
Error of biomass regressions: sample trees selected by stratified sampling	253
Alexandros Arabatzis and Tiberius Cunia	
Error of biomass regressions: sample trees selected by double sampling	269
John Michelakackis and Tiberius Cunia	
Using simulation to evaluate volume equation error and sampling error in a two-phase design	287
David C. Chojnacky	
High order regression models for regional volume equations	295
Joe P. McClure and Raymond L. Czaplewski	

¹Contributed paper, not presented at the workshop.

RESEARCH PAPERS

**Biomass Regressions
And Measurement Error**

Moderator: Timothy G. Gregoire

AN OPTIMIZATION MODEL FOR SUBSAMPLING TREES FOR
BIOMASS MEASUREMENT

Tiberius Cunia

Professor of Statistics and Operations Research
SUNY College of Environmental Science & Forestry
Syracuse, NY 13210

When selecting sample trees for biomass tables construction is customary to estimate their biomass by subsampling rather than measure their actual values. By this procedure one must work with estimated rather than actual tree biomass. One may use an intensive tree subsampling and obtain estimates of the sample tree biomass with high precision. But this implies that for the same sampling costs one must reduce the number of sample trees. On the other hand one may wish to work with more sample trees and less precise estimates of their biomass. An approach using mathematical programming models is shown that optimizes the combination of number of sample trees and intensity of subsampling.

Introduction

Although not explicitly stated when they are applied to forest inventory problems, the tree biomass regression functions are not without error. This error has four main components, each component being associated with a specific source.

There is first the error due to sample tree selection. The same sampling method applied on different occasions results in different sets of sample trees and, thus, in different biomass regression functions. The size of this error component, generally known as the sampling error, is affected by the sampling design by which the sample trees are selected, by the number of sample trees, and by the inherent variation between the biomass of various trees.

Once the sample trees are selected, they must be measured for their biomass. The same tree measured on different occasions may result in different biomass values. We hypothesize the existence of a constant, fixed value which we shall call true biomass of a tree (or of some of its components). What we call the biomass of a tree is the measured value that we have obtained by some measurement or subsampling process. The difference between the true and the measured biomass value will be known here as the measurement error. This is the second error component. For variables that can be measured directly, as for example, the green weight of some small tree component, the measurement error may be negligibly small. For others, like the oven-dry weight of the entire tree which, for practical purposes is measured by subsampling, the error may be quite large.

The third source of error is the statistical model used in data analysis and estimation of the biomass regression function. Different statisticians working with the same sample data may arrive at different biomass regressions. For any given real world problem, one can construct different models, each model having its own assumptions. These assumptions refer to the characteristics of the tree population and the method by which the sample trees are selected. If valid statistical inferences are desired, the assumptions must be sufficiently well satisfied. But even if they are satisfied, different, equally good statistical models would generally yield different biomass regressions. If the models are well selected, however, the difference between their results are expected to be small. This component is known as the error due to statistical modeling. It should not be confused with what is known as statistical error, a term used to denote, in general, the error (from all but subjective sources) associated with the difference between our inferences and the true values of the parameters we want estimated.

The fourth error component is due to the application of the biomass regression function to a specific case and may be known as the error due to application. This is because the regressions are estimated from trees that are no longer members of the population from which they were drawn; they were felled down and destroyed. Furthermore, the populations of trees are dynamic (they change with time) and it is quite common to apply biomass regression functions to forest areas other than those for which the regression functions were calculated. This is generally known as a non-statistical error component since it can hardly be evaluated by statistical means. Note that strictly speaking, this error component is not a part of the error of the biomass regression; only a component due to its use.

We shall not be concerned here with the first, the third or the fourth error component. We shall assume that (i) the sample trees are selected by simple random sampling, the usual assumption of the least squares method, (ii) the statistical model used in the data analysis is appropriate, and (iii) the biomass regression function is being applied to the population for which it was calculated. We shall only be concerned here with the second component, the measurement error.

The measurement error itself has three main subcomponents. There is first the so-called measurement bias due to faulty instruments, to poorly defined or applied measurement procedures, to poor subsampling designs or to selection of biased estimators. There is also the random error due to the precision of the measuring devices or the measurement process itself, as for example, the precision of the weighing scale, caliper or diameter tape, etc. This last error can be assumed to be negligibly small, and together with the measurement bias above will be ignored in this discussion. Our only topic of discussion will be the subcomponent error due to the inherent variation associated with any subsampling process used to determine the biomass (usually oven-dry weight of large tree components) of a sample tree. The same subsampling

procedure, using the same estimators, applied on various occasions, would normally yield different results.

Note that there are no statistical estimation problems related to the subsampling error. As long as (i) the expected value of this error is zero, that is, the estimator of the biomass of the sample tree is unbiased, and (ii) the error is statistically independent of the error component due to sample tree selection, the least squares method takes it automatically into account. The only problem is one of efficiency. Should one select a small number of sample trees and estimate their biomass by an intensive subsampling or is it better to work with a large number of sample trees for which the biomass is poorly measured?

The specific objectives of the present study are to construct a mathematical model for the error of tree biomass regression functions. Because the same error can be obtained from various combinations of number of sample trees and intensity of subsampling, a model is needed to guide us in the selection of the sample and subsample size. Optimum sample and subsample size should be sought; those that would minimize the costs of sampling for desired precision of tree biomass regressions or those that maximize the precision for given allowable sampling costs.

The type of model we shall construct is known as a mathematical programming model. It will consist of either (i) an objective cost function that should be minimized subject to the side condition that the error of the biomass regression function should be smaller than a predetermined value, or (ii) an objective error function that should be minimized subject to sampling costs being below a specified amount. Both cost and error functions will be expressed explicitly in terms of sample size n (of sample trees) and subsample size m (for the estimation of the biomass of the sample trees).

In constructing this model, we shall assume that the selection of sample trees and the estimation of their biomass by subsampling is made by the simple random sampling method. We shall then extend the model to include the case where the biomass of the sample trees is estimated by other sampling methods. We shall not consider, however, the case where the sample trees themselves are selected by methods other than simple random sampling. This is because in order to calculate the regression function, one must then modify the least squares method of estimation.

Expressing the Error of Biomass Regressions

The error of the biomass regression functions can be expressed in many ways, each way being adequate for some applications and quite inadequate for others. It can be expressed as the variance or standard error of (i) regression estimates of the biomass of trees of given size, (ii) regression coefficients, (iii) estimators (of parameters of interest) obtained when the biomass regressions

are applied to forest inventory data, etc. It can also be expressed as confidence intervals (at some predetermined probability level) of (i) the expected biomass of trees of given size, (ii) parameters of interest of the true regression function, (iii) parameters of interest of some forest tree population, etc.

We are interested in the application of the biomass regressions to forest inventory data and wish to work with an expression of the error that can be used within this context. Cunia (1986a, b) has suggested a method to combine the error of the biomass regression with the error of the sample plots (of the forest inventory) so that the calculated error of the inventory estimates would contain the error from both sources. This method assumes that (i) the true regression function of tree biomass on diameter, height, species, etc. is linear, that is, of the form

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = [\beta]' [x]$$

where y = tree biomass, $x_1 = 1$, x_i , $i = 2, 3, \dots, p$ are tree characteristics other than biomass and $[]$ denotes a vector or matrix (with $[]'$ denoting a transpose), (ii) the sample plots (where the trees are measured for attributes other than biomass) are selected by a variety of sampling designs, (iii) the sample of plots and the biomass regression are statistically independent, and (iv) the forest parameters of interest are average biomass per tree, average biomass per unit area (acre or hectare), average growth (or growth components of biomass per unit area, etc.).

To apply Cunia's procedure, one must have an estimate $[b]$ of the vector $[\beta]$ of regression coefficients and an estimate $[S_{bb}]$ of the covariance matrix $[\sigma_{bb}]$ of $[b]$. Because $[\sigma_{bb}]$ can be expressed as a function of (i) the conditional variance of y given $[x]$ and (ii) the sample sizes n and m defined in the previous section, the error expression we shall select for our optimization model is that of $[\sigma_{bb}]$.

To better see our approach, let us review the least squares method as applied to biomass regressions. We shall use the following basic assumptions.

- (1) The true regression function of y on $x_1 = 1, x_2, x_3, \dots, x_p$ is of the linear form

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = [\beta]' [x]$$
- (2) The conditional variance of y given $[x]$, say $\sigma_{yy|x}$, is proportional to some given function of the known tree variables x_1, x_2, \dots, x_p , say

$$\sigma_{yy|x} = \sigma_{uu|v}^w$$
 where $\sigma_{uu|v}$ is the unknown factor of proportionality and w is a known function of x_1, x_2, \dots, x_p .

- (3) The n sets of sample observations $y_k, x_{k1}, x_{k2}, \dots, x_{kp}$ for $k = 1, 2, \dots, n$ are statistically independent of each other.

Then, the best linear and unbiased estimator $[b]$ of $[\beta]$ is the weighted least squares estimator calculated as follows.

Step 1 - Calculate the n sets of transformed sample values $u_k = y_k / \sqrt{w_k}, v_{ki} = x_{ki} / \sqrt{w_k}, i=1,2,\dots,p$ and $k = 1, 2, \dots, n$ and construct the vector $[U]$ and matrix $[V]$ of these values.

$$[U] = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \text{ and } [V] = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{np} \end{bmatrix}$$

Step 2 - Calculate the matrices [T] and [P] of sum of cross-products

[T] = [V]'[V] and [P] = [V]'[U]
and the estimator [b] of [β]

$$[b] = [T]^{-1}[P]$$

Step 3 - Calculate the estimator $S_{uu|v}$ of $\sigma_{uu|v}$

$S_{uu|v} = ([U]'[U] - [b]'[P]) / (n-p)$
and the estimator $[S_{bb}]$ of the covariance matrix $[\sigma_{bb}]$ of [b]

$$[S_{bb}] = S_{uu|v} [T]^{-1}$$

Remarks

(1) If y is the biomass of the main tree bole or the total above ground tree biomass, it has been found empirically that the conditional variance of y given diameter d or given diameter d and height h is approximately proportional to d^4 or $d^4 h^2$ respectively. This means that for the first case u is equal to y/d^2 while u is equal to $y/d^2 h$ for the second case. Similar definitions apply to variables v_1, v_2, \dots, v_p . Of course, for other biomass components the conditional variance of y given d or given d and h may be proportional to known values other than d^4 or $d^4 h^2$, may be related in a different way to d or d and h or may even be homogeneous.

(2) $S_{uu|v}$ is an estimator of the factor of proportionality $\sigma_{uu|v}$ which now can also be interpreted as the conditional variance of $u = y/\sqrt{w}$ for given [v], or what is the same thing, for given [x]. This means that the conditional variance of y given [x] can be estimated by the formula

$$S_{yy|x} = w S_{uu|v}$$

For the biomass y of (1) above, we have

$$S_{yy|x} = d^4 S_{uu|v}$$

or

$$S_{yy|x} = d^4 h^2 S_{uu|v}$$

(3) If the model assumptions stated above are satisfied, the conditional variance of u given [x] is homogeneous, that is $\sigma_{uu|v}$ is the same for the trees of all sizes [x]. This has an extremely important implication. As we shall analyze the conditional variance of y given [x] which varies with the tree size [x], and because $\sigma_{uu|v} = \sigma_{yy|x}/w$ is constant over the trees of all sizes [x], we can work with the variance u and then interpret the conclusions reached later in terms of the variable y.

For example, if the variable x are all defined in terms of d alone, and the conditional variance of y given d is proportional to d^4 , the variable $u = y/d^2$ has a conditional variance which is

homogeneous with respect to d; it is the same for the trees of any diameter size d. This means that if the trees are classified by diameter d, then the variance of u within a diameter class is the same for all diameter classes. Note that $u = y/d^2$ has a physical meaning; it represents "biomass per square inch of breast high diameter." As the basal area of a tree is $\Pi d^2/4$, and $\Pi/4$ is a constant value, u is equivalent to the variable $u^* = (y/\text{basal area})$, that is equivalent to the variable tree biomass per unit of basal area.

Let us express now the covariance matrix $[\sigma_{bb}]$ of [b], defined as

$$[\sigma_{bb}] = \sigma_{uu|v} [T]^{-1}$$

as an explicit function of the number n of sample trees. If the n sample trees are selected completely at random, it can be shown that the expected value of [T] can be written as

$$[T] = n[\bar{T}]$$

where the ij-th element of $[\bar{T}]$ is the expected value of the cross product $v_i v_j$. As the inverse of [T] is

$$[T]^{-1} = [\bar{T}]^{-1}/n = [\theta]/n$$

where [θ] is the inverse of the matrix $[\bar{T}]$, we can finally write that, on the average, we expect to have the approximate relationship

$$[\sigma_{bb}] = (\sigma_{uu|v}/n) [\theta]$$

A Simple Optimization Model 1

Consider a sample of n trees selected by simple random sampling from a relatively large population of N trees. If y is the true biomass of a tree (or some of its components), we shall assume that $\sigma_{yy|x}$ the conditional variance of y given [x] is proportional to some known value w (usually, but not necessarily a function of the known tree values x_1, x_2, \dots, x_p). As shown in the previous section, the variable $u = y/\sqrt{w}$ has a conditional variance for given [x] that remains constant over all values [x]. We shall denote this variance V_1 . Note that V_1 is defined in terms of true values y and $u = y/\sqrt{w}$.

Because the biomass value y of a given tree is not generally measured but estimated by subsampling, let us further assume that a given tree is subdivided into a large number M of approximately equal parts denoted here as biomass samples. These samples are usually, but not necessarily non-overlapping. Assuming non-overlapping biomass samples, let m samples be selected by simple random sampling. If

$$z_k = (\text{biomass of sample } k) / \sqrt{w} = \text{variable measured (without error) on the biomass sample } k$$

$$\mu_z = \Sigma z_k / M = u / M = \text{mean of } z \text{ within a given tree}$$

$$\sigma_{zz} = \Sigma (z_k - \mu_z)^2 / M = \text{variance of } z \text{ within a given tree, } z$$

where Σ is taken over all samples $k = 1, 2, \dots, M$, then, it is well known that for Σ taken now over the subsample values $k = 1, 2, \dots, m$, we have

$$\bar{z} = \Sigma z_k / m = \text{subsample mean of } z = \text{estimator of } \mu_z$$

$$S_{zz} = \Sigma (z_k - \bar{z})^2 / (m-1) = \text{subsample variance of } z = \text{estimator of the variance of } z$$

$\hat{u} = M\bar{z} =$ estimator of the true value of the biomass u of the given tree

$\sigma_{uu}^{\wedge\wedge} = M^2\sigma_{zz}^{\wedge\wedge}/m =$ variance of \hat{u} as an estimator of \hat{u} of the given tree, and

$S_{uu}^{\wedge\wedge} = M^2S_{zz}^{\wedge\wedge}/m =$ estimator of $\sigma_{uu}^{\wedge\wedge}$

Note that for the given tree, the biomass y is estimated by

$$\hat{y} = \hat{u}\sqrt{w} = M\bar{z}\sqrt{w}$$

and the variance of \hat{y} is estimated by

$$S_{yy}^{\wedge\wedge} = wS_{uu}^{\wedge\wedge} = wM^2S_{zz}^{\wedge\wedge}/m$$

Furthermore, under our subsampling assumptions, z , \hat{u} , and \hat{y} are all unbiased and when $\sigma_{uu}^{\wedge\wedge}$, $S_{uu}^{\wedge\wedge}$, and $S_{yy}^{\wedge\wedge}$ were defined, the effect of the finite population correction factor was ignored.

For convenience, we shall write $V_2 = M^2\sigma_{zz}^{\wedge\wedge}$. This means that $\sigma_{uu}^{\wedge\wedge}$ can be written as V_2/m . Recall that V_1 has been defined as the conditional variance of u given $[x]$. As u is not known, we have to work with its estimator \hat{u} . This implies that the variance of \hat{u} about the conditional mean of u given $[x]$ has two, statistically independent components; the first component is V_1 associated with the variation of u about its conditional mean and the second component is V_2/m associated with the variation of \hat{u} about the true value u . Consequently, the conditional variance of u (which in reality is \hat{u}) given $[x]$, which in the previous section was denoted by $\sigma_{uu|v}$, can now be written as

$$V = V_1 + V_2/m = \sigma_{uu|v}$$

As an example of a case where this model can be applied, consider the biomass y of the main bole of randomly selected trees and its regression function on diameter d . Empirical evidence suggests that the regression function of y on d is of the approximate form

$$y = \beta_1 + \beta_2 d + \beta_3 d^2 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 = [\beta]'[x]$$

and that the conditional variance of y given $[x]$ is approximately proportional to $w = d^4$. Consequently, the variable u is defined as

$$u = y/d^2 = \text{bole biomass per square inch (of squared diameter)}$$

Note that y and u are true values of a given tree.

To estimate y and u (of a tree of known d), we shall divide the bole into a large number M of non-overlapping disks. If m disks are selected by simple random sampling without replacement, then we can define the following:

$z = (\text{disk biomass})/d^2 =$ value measured on each disk

$\mu_z = u/M =$ true mean of z

$\sigma_{zz}^{\wedge\wedge} =$ variance of z

$\bar{z} =$ sample mean of the m biomass samples

$S_{zz}^{\wedge\wedge} =$ sample variance of the m biomass samples

$\bar{z} =$ unbiased estimator of μ_z

$\hat{u} = M\bar{z} =$ unbiased estimator of u

$\sigma_{uu}^{\wedge\wedge} = M^2\sigma_{zz}^{\wedge\wedge}/m = V_2/m =$ variance of \hat{u} (as an estimator of u)

$S_{uu}^{\wedge\wedge} = M^2S_{zz}^{\wedge\wedge}/m =$ estimator of $\sigma_{uu}^{\wedge\wedge}$

$\hat{y} = \hat{u}d^2 =$ unbiased estimator of y

$\sigma_{yy}^{\wedge\wedge} = \sigma_{uu}^{\wedge\wedge}d^4 =$ variance of \hat{y} (as an estimator of y)

$S_{yy}^{\wedge\wedge} = S_{uu}^{\wedge\wedge}d^4 =$ estimator of $\sigma_{yy}^{\wedge\wedge}$

Because y and u are not known, we must work with \hat{y} and \hat{u} . The conditional variance of \hat{y} about the regression function of y on $[x]$ (that is, d) is assumed proportional to d^4 , while the conditional variance of \hat{u} about its regression function of u on $[v]$ (that is, d) is homogeneous. Because \hat{u} as an estimate of u is statistically independent of u as an estimate of the expected value of u given $[v]$ (or $[x]$ or d), the conditional variance of \hat{u} given $[v]$ can be written as

$$\sigma_{uu|v} = V_1 + V_2/m$$

since the conditional variance of u given $[v]$ has been denoted by V_1 and the variance of \hat{u} as an estimator of u has been written as V_2/m .

As a second example of where this model can be applied is the method of estimating the tree biomass y by the randomized branch method described by Valentine, Tritton, and Furnival (1985). Each application of the method yields an estimate \hat{y} of y , and because d is known, an estimate $\hat{u} = \hat{y}/d^2$ of $u = y/d^2$. As one can apply the method infinitely many times, M is infinite in size. Nevertheless, we can define the following:

$z_k =$ estimate of u based on the k -th application of the randomized branch method

$\bar{z} = \hat{u} =$ estimate of u , since the expected value μ_z of z is equal to u (the method is unbiased)

$V_2/m = \sigma_{zz}^{\wedge\wedge}/m =$ variance of $\bar{z} =$ variance of \hat{u} .

Note that with this subsampling procedure (i) the biomass samples are overlapping, (ii) the m biomass samples are selected with equal probabilities, even though various parts of the tree enter in biomass samples with unequal probabilities, (iii) each value z is defined so as to be an unbiased estimate of u and (iv) the variance of \hat{u} about the regression function (of u on $[v]$) can be written as $V = (V_1 + V_2/m)$.

Let us define now the sampling costs components,

$c_s =$ average costs directly related to the selection and measurement of a biomass sample, and

$c_t =$ average costs of selecting, processing, and measuring a sample tree for characteristics other than those directly related to the measurement of the biomass samples.

For example, the costs of moving to and from the sample trees, the costs of measuring the tree diameter, height or bole length, the costs of felling, branching and measuring the green weight of the entire tree bole, the costs of the separation (conceptually if not physically) of the tree bole into non-overlapping disks, etc. are all included in c_t .

on the other hand, the actual selection of a sample disk or its location, the measurement of the green or oven-dry weight of a sample disk, etc. are all included in c_s .

If the fixed costs of sampling not directly related to selection and measurement of sample trees and biomass samples (that is, all costs not included in c_t and c_s) are ignored, the variable costs of sampling and subsampling a tree for its biomass can be evaluated, at least approximately, by the linear cost function

$$C = nc_t + nmc_s$$

As an example of some of the fixed sampling costs are those of training costs of the field crews, the costs of writing and debugging computer programs, etc.

It has been shown in the previous section that the expression of the error of biomass regression we shall work with is $[\sigma_{bb}] = (\sigma_{uu}|v/n)[\theta]$. If the selection of the sample tree is done by simple random sampling, it has been shown that $[\theta]$ is a parameter of the tree population of interest. Consequently, it suffices to work only with the expression $(\sigma_{uu}|v/n)$ which is a function of both n and m as the u variables are estimated by the \hat{u} values determined by subsampling. Thus, the problem reduces to that of finding the sample sizes m and n for which the sampling costs C are minimized and the variance

$$\sigma_{uu}|v/n = V/n = (V_1 + V_2/m)/n$$

is equal to some desired value K_0 . Or equivalently, the problem is that of finding m_0 and n for which the value of the variance $K = V/n$ is minimized for given costs of sampling C_0 . More specifically, this optimization problem can be formalized as follows.

Optimization Problem 1. Find m and n that minimize the cost (objective) function

$$C = nc_t + nmc_s$$

subject to the side (variance) condition that

$$(V_1 + V_2/m)/n = K_0$$

where all parameters c_t , c_s , V_1 , V_2 , and K_0 are positive values and m and n are positive integers, and

Optimization Problem 2. Find m and n that minimize the variance (objective) function

$$K = (V_1 + V_2/m)/n$$

subject to the side (cost) condition that

$$nc_t + nmc_s = C_0$$

where c_t , c_s , V_1 , V_2 , and C_0 are all positive and m and n are positive integers.

By using calculus optimization techniques, one can easily find the following optimum solutions.

Solution to Problem 1. Nearest positive integer values

$$m = \sqrt{c_t V_2 / c_s V_1} = (\sqrt{c_t / c_s}) (\sqrt{V_2 / V_1})$$

and

$$n = (V_1 + V_2/m) / K_0 = (\sqrt{c_t V_1} + \sqrt{c_s V_2}) (\sqrt{V_1 / c_t}) / K_0$$

Solution to Problem 2. Nearest positive integer values

$$m = \sqrt{c_t V_2 / c_s V_1} = (\sqrt{c_t / c_s}) (\sqrt{V_2 / V_1})$$

and

$$n = C_0 (\sqrt{V_1 / c_t}) / (\sqrt{c_t V_1} + \sqrt{c_s V_2})$$

We shall show below the proof to the first problem; the proof to the second problem follows the same lines and can be easily done by the reader.

Proof of Solution to Problem 1. To find the optimum values m and n , one can use the method of Lagrange multipliers or the method of substitution. Using the latter method, we substitute the value of n from the side condition

$$n = (V_1 + V_2/m) / K_0$$

for the value of n of the objective function. Then, the new objective function to minimize becomes

$$C = (V_1 + V_2/m) (c_t + mc_s) / K_0$$

Because $K > 0$ and $(V_1 c_t + V_2 c_s) > 0$ are both fixed values, it is equivalent to work with the objective function

$$C^* = K_0 C - (c_t V_1 + c_s V_2) = (V_1 + V_2/m) (c_t + mc_s) - (c_t V_1 + c_s V_2) \\ = c_s V_1 m + c_t V_2 / m$$

To find the optimum solution, we equate to zero the first derivative of C^* with respect to m and solve for m the resulting equation. Then

$$dC^*/dm = c_s V_1 - c_t V_2 / m^2 = 0$$

and the only admissible solution (m is not allowed to be negative) is

$$m = \sqrt{c_t V_2 / c_s V_1}$$

As the second derivative

$$d^2 C^* / dm^2 = 2c_t V_2 / m^3$$

is always greater than zero for the positive solution value of m above, the function C^* attains a minimum. This in turn yields the optimum value of n as

$$n = (V_1 + V_2/m) / K_0 \\ = (V_1 + \sqrt{c_s V_1 V_2 / c_t}) / K_0 \\ = (\sqrt{V_1 / V_2} + \sqrt{c_s / c_t}) (\sqrt{V_1 V_2}) / K_0 \\ = (\sqrt{c_t V_1} + \sqrt{c_s V_2}) (\sqrt{V_1 / c_t}) / K_0$$

Remarks

(1) Because m and n can only be positive integers, the optimum solution above, which in general will be a non-integer solution, must be rounded-off to the nearest integer or to the nearest integer above. This is not a critical factor since the cost values c_t and c_s and variance values V_1 and V_2 are only approximately known.

(2) The optimum subsample size m does not depend on n but only on the variance components V_1 and V_2 and the average costs of sampling per tree and per biomass sample c_t and c_s respectively. Once the optimum m is determined, the value of n is then determined as a function of m and the desired precision K or allowable costs C . This is a very important finding. It means that we can optimize the subsampling first and then seek the value of the sample size n of trees to satisfy the overall requirements of precision or costs.

(3) At least from an intuitive point of view, the optimum solution makes sense. As the variance component V_2 (between biomass samples within trees) decreases with respect to the variance component V_1 (between true value of tree biomass), and as long as the costs c_t and c_s remain the same, we would normally expect m to decrease; we should spend relatively less money for subsampling. For the limiting case where $V_2 = 0$, we must have the limiting case $m = 1$, since this is the positive integer closest to solution $m = 0$; only one biomass sample would give us all the necessary information about the true value u . Similarly, as the costs c_t increase with respect to c_s and the variance components V_1 and V_2 remain the same, we would expect m to decrease; we should spend relatively less money for subsampling. For the limiting case where $c_s = \infty$, we have the limiting case $m = 1$, the positive integer closest to $m = 0$.

(4) We have assumed here implicitly, that the variance component V_2 is the same from tree to tree. This is obviously not true as each sample tree $i = 1, 2, \dots, n$ has its own variance V_{2i} . Because (i) the values V_{2i} are random variables (the trees are selected at random), (ii) knowledge about the probability distribution of V_{2i} is required, and (iii) to construct an optimization model where V_2 itself becomes a random variable is extremely difficult, we must work with some kind of average value V_{2i} , or an estimate of the expected value of V_{2i} , which here is denoted simply as V_2 . This same remark applies as well to the costs c_t and c_s which are also averages of random variables c_{ti} and c_{si} . In general c_{ti} increases with the tree size (felling and processing a large tree is more expensive than felling and processing a small tree), even though the variation of c_{si} (of the costs per biomass sample), for some tree components at least may be relatively small. For practical considerations we must work with the expected values c_t and c_s of the costs c_{ti} and c_{si} respectively.

(5) We have assumed here that a fixed number m of biomass samples are to be selected from each sample tree. A better optimization model may be that which makes m a function of V_2 (or diameter d , if V_2 is a function of d), that is, having an optimum solution m_i as a function of V_{2i} of the sample tree i we happen to select. This type of a model is obviously of much more complexity. However, due to the usually poor knowledge that one may have about the costs c_t and c_s or variance components V_1 and V_2 , it is doubtful that one may gain much by constructing such complex models; use of models based on average values V_2 , c_t and c_s are sufficiently good for all practical purposes.

It is relatively easy to extend the usefulness of the simple optimization model 1 of the previous section by considering a few more common subsampling designs or estimators. For convenience these new models will be known as Models 2, 3, 4, and 5.

Model 2 - Simple Random Subsampling; m is large with respect to M . When m is too large with respect to M one cannot ignore the effect of the finite population correction factor. As the variance of \hat{u} is equal to

$$\sigma_{\hat{u}}^2 = \left(\frac{M-m}{M-1} \right) \left(\frac{M^2 \sigma_{zz}}{m} \right) = \left(\frac{M-m}{M} \right) \left(\frac{M^2 \sigma_{zz}}{m} \right) = (M-m) V_2 / Mm$$

we can write the conditional variance of \hat{u} for given $[x]$ as

$$V = V_1 + (M-m) V_2 / Mm = V_1^* + V_2 / m$$

where $V_1^* = V_1 - V_2 / M$.

As the optimization problem of this model is now identical to that of model 1 of the previous section, the optimum solution m and n is given by the same formulae; only V_1^* is substituted for V_1 .

Model 3 - Simple Random Subsampling; m is large relative to M and the estimator of u is of a ratio type. Assume that an auxiliary variable g_k exists for each biomass sample k and that the corresponding total value G of all M biomass sample values g_k is known without error. If the correlation between z_k and g_k is sufficiently strong, an efficient estimator of u is the well known ratio estimator

$$u_r = Gr$$

where, for Σ taken over the sample units $k = 1, 2, \dots, m$,

$$r = \Sigma z_k / \Sigma g_k = \bar{z} / \bar{g}$$

= estimator of the ratio $\rho (=u/G)$ of the true value u of the variable of interest to the true value G of the auxiliary variable.

The estimator u_r is biased, but the bias is expected to be small when m is sufficiently large. The size of the bias can be approximated by the formula

$$B = (M-m) (\rho \sigma_{gg} - \sigma_{zg}) / m \mu_g$$

where $\mu_g = G/M$ is the expected value of g_k , σ_{gg} is the variance of g_k and σ_{zg} is the covariance of z_k and g_k . The variance of z_g as an estimator of the true value u of the given tree is approximately equal to

$$\sigma_{u_r u_r} = \left(\frac{M-m}{M} \right) (V_2^{**} / m)$$

where, for Σ taken over all tree values $k = 1, 2, \dots, M$,

$$V_2^{**} = M^2 \Sigma (z_k - \rho g_k)^2 / M = M^2 (\sigma_{zz} - 2\rho \sigma_{zg} + \rho^2 \sigma_{gg})$$

By using statistics g , r , S_{zz} , S_{zg} , and S_{gg} calculated from the m pairs of values z_k and g_k one can define similarly estimators of the bias and variance of u_r .

To better see the meaning of u_r , consider the problem of the estimation of the bole oven-dry weight y of a given tree. We are interested in the regression of y on diameter d and empirical evidence suggests that the conditional variance of y given d is proportional to d^4 . This implies that we should work with the transformed variable $u = y/d^2 =$ the bole biomass (oven-dry weight) per square inch of diameter at breast height. Let the biomass samples be disks cut from the tree bole, M be the total number of disks for the given tree and m be the number of sample disks selected at random and without replacement. It is known that the green and oven-dry weights are highly correlated and that $G =$ total green weight of the entire bole divided by d^2 is easy to determine. If we define now the disk values z_k and g_k by the formulae

$$z_k = (\text{oven-dry weight of disk } k) / d^2$$

and

$$g_k = (\text{green weight of disk } k) / d^2$$

we have the following statistics and estimators

$r = \bar{z}/\bar{g} =$ sample ratio of oven-dry/green weight
 = estimator of the ratio $\rho (=u/G)$ of the unknown total oven-dry weight u to the known green weight of the tree bole, and

$u_r = Gr =$ estimator of u , the total oven-dry weight per square inch of the given tree bole

$y_r = d^2 u_r =$ estimator of y , the total oven-dry weight of the given tree bole

We can write now the conditional variance V of u_r given $[x]$ as

$$V = V_1 + \left(\frac{M-m}{M} \right) (V_2^{**}/m) = V_1^{**} + V_2^{**}/m$$

where

$$V_1^{**} = V_1 - V_2^{**}/M = V_1 - M(\sigma_{zz} - 2\rho\sigma_{zg} + \rho^2\sigma_{gg})$$

Consequently, substitution of V_1^{**} and V_2^{**} for V_1 and V_2 of the formulae of the previous section 1 model 1, yields the optimum values m and n for the present model 3. These formulae may take a different form when $M = \infty$, the case of infinitely many, overlapping biomass samples. Then, V_1^{**} becomes equal to V_1 (since $V_2^{**}/M = 0$) and the definition of the variances and covariance of z_k and g_k must change accordingly (in terms of expected values).

Model 4 - Stratified Random Subsampling. We shall assume that (i) the tree is divided into L large sections known as strata, (ii) each stratum $h = 1, 2, \dots, L$ is further subdivided into M_h biomass samples, (iii) $m_h \geq 2$ biomass samples are selected from each stratum h by simple random sampling without replacement, and (iv) the samples from various strata are statistically independent. As before, we define y as the biomass of the tree (or some of its components) and u as (y/\sqrt{w}) , the transformed value of y . In addition, we shall also define the following stratum values

$y_h =$ biomass of stratum h

$u_h = y_h/\sqrt{w} =$ transformed biomass value of stratum h

$z_{hk} =$ (biomass of sample k within stratum h)/ \sqrt{w}

$\mu_{zh} =$ mean of the M_h values z_{hk} within stratum h

$\bar{z}_h =$ mean of the m_h sample values z_{hk} within stratum h

$\sigma_{zzh} =$ variance of the M_h values z_{hk} within stratum h

$S_{zzh} =$ variance of the m_h sample values z_{hk} within stratum h

$\hat{u}_h = M_h \bar{z}_h =$ estimator of u_h

$\sigma_{uuh}^{\wedge} \approx M_h (M_h - m_h) (\sigma_{zzh}/m_h)$

$= (M_h - m_h) (V_{2h}/m_h) / M_h =$ variance of \hat{u}_h

where $V_{2h} = M_h^2 \sigma_{zzh}$

$S_{uuh}^{\wedge} = M_h (M_h - m_h) (S_{zzh}/m_h) =$ estimator of the variance of \hat{u}_h

Using these values, we can define

$$u_s = \hat{u}_1 + \hat{u}_2 + \dots + \hat{u}_L = \Sigma \hat{u}_h$$

= stratified sampling estimator of the true value u of the given tree

and

$$V_s = \Sigma \sigma_{uuh}^{\wedge} = \Sigma M_h (M_h - m_h) (\sigma_{zzh}/m_h)$$

= $\Sigma (M_h - m_h) (V_{2h}/m_h) / M_h =$ variance of u_s

If we define

$$V_1^s = V_1 - \Sigma (V_{2h}/M_h)$$

we can write the conditional variance of u_r given $[x]$ as

$$V = V_1^s + \Sigma (V_{2h}/m_h)$$

If we further define

$c_h =$ average cost of measuring a biomass sample from stratum $h = 1, 2, \dots, L$

the optimization problem can be expressed in the following two forms

Optimization Problem 1. Find values n, m_1, m_2, \dots, m_L that minimize the cost (objective) function

$$C = nc_t + \Sigma n_m c_h$$

subject to the side (variance) condition that

$$(V_1^s + \Sigma (V_{2h}/m_h)) / n = K_0$$

where all parameters c_t, c_h, V_1^s, V_{2h} , and K_0 are positive values and the optimum solution n, m_1, m_2, \dots, m_L are positive integers (greater than two), and

Optimization Problem 2. Find values n, m_1, m_2, \dots, m_L that minimize the variance (objective) function

$$K = (V_1^s + \Sigma (V_{2h}/m_h)) / n$$

subject to the side (cost) condition that

$$nc_t + \Sigma n_m c_h = C_0$$

where c_t, c_h, V_1^s, V_{2h} , and C_0 are positive values and n, m_1, m_2, \dots, m_L are positive integers (greater than two).

To solve these problems, we shall proceed as follows. It is known, see Cochran (1977) among

others, that for a given, overall sample size $m = (m_1 + m_2 + \dots + m_L)$, the optimum allocation of the sample size m_h within stratum h is proportional to the values $m_h \sqrt{V_{2h}/c_h}$. Note that the effect of the stratum size M_h that is shown explicitly in Cochran's procedure is included implicitly in the definition of V_{2h} , since $V_{2h} = M_h^2 \sigma_{zzh}$. This implies that

$$(1) m_h = m \sqrt{V_{2h}/c_h} / \sum \sqrt{V_{2h}/c_h} \\ = mM_h \sqrt{\sigma_{zzh}/c_h} / \sum M_h \sqrt{\sigma_{zzh}/c_h}$$

$$(2) \sum m_h c_h = m \sum c_h \sqrt{V_{2h}/c_h} / \sum \sqrt{V_{2h}/c_h} \\ = m \sum \sqrt{V_{2h} c_h} / \sum \sqrt{V_{2h}/c_h} \\ = m \sum M_h \sqrt{\sigma_{zzh} c_h} / \sum M_h \sqrt{\sigma_{zzh}/c_h} \\ = mc_s$$

where $c_s = \sum m_h c_h / m$ = average cost of measuring a biomass sample from the overall population of biomass samples when the allocation is optimum, and

$$(3) \Sigma (V_{2h}/m_h) = (\Sigma \sqrt{V_{2h}/c_h}) (\Sigma \sqrt{V_{2h}/c_h}) / m \\ = (\Sigma \sqrt{V_{2h}/c_h}) (\Sigma \sqrt{V_{2h} c_h}) / m \\ = (\Sigma M_h \sqrt{\sigma_{zzh}/c_h}) (\Sigma M_h \sqrt{\sigma_{zzh} c_h}) / m \\ = V_2^s / m$$

where

$$V_2^s = (\Sigma \sqrt{V_{2h}/c_h}) (\Sigma \sqrt{V_{2h} c_h}) \\ = (\Sigma M_h \sqrt{\sigma_{zzh}/c_h}) (\Sigma M_h \sqrt{\sigma_{zzh} c_h})$$

Because the cost and variance functions can now be written as

$$C = nc_t + mmc_s$$

and

$$K = (V_1^s + V_2^s/m)/n$$

respectively, the solution to the two optimization problems of this section is the same as the solution of the two optimization problems of the previous section. Of course, V_1^s and V_2^s are substituted for V_1 and V_2 , respectively and the definition of c_s is slightly different.

Remarks

(1) Most of the remarks made in the previous section for model 1 apply here as well. More specifically, (i) all m_h and n values are obtained by the rounding-off (to the nearest or nearest above) integer of the optimum solution values m_h and n obtained by the calculus optimization techniques, (ii) the optimum solution makes sense from an intuitive point of view; m_h increases with the increase of V_{2h} and/or decrease of c_h , and (iii) for practical considerations we have to assume that the variation within trees is the same from tree to tree.

(2) It is extremely difficult, if at all possible, to have good estimates of stratum variances σ_{zzh} , V_{2h} or costs c_h , c_s . Therefore, one can make assumptions that would simplify the formulae. For example, one may assume that $c_h = c_s$ for all h . Then,

$$m_h = m \sqrt{V_{2h}} / \sum \sqrt{V_{2h}}, \text{ and } V_2^s = (\sum \sqrt{V_{2h}})^2$$

Or one may assume that $\sigma_{zzh} = \sigma_{zz}$ in addition to $c_h = c_s$ for all h . Then, for $\sum M_h^2 = M$,

$$m_h = mM_h/M \text{ and } V_2^s = M^2 \sigma_{zz}$$

Finally, if $M_h = M/L$ (equal strata), $\sigma_{zzh} = \sigma_{zz}$ and $c_h = c_s$ for all h , the formulae becomes

$$m_h = m/L \text{ and } V_2^s = M^2 \sigma_{zz} = L^2 V_2$$

since $V_2 = M^2 \sigma_{zz} = M^2 \sigma_{zz} / L^2$

(3) We have defined V_2^s as the quantity

$$V_2^s = m \Sigma (V_{2h}/m_h) = \Sigma (V_{2h}/(m_h/m))$$

where m_h/m were the optimum allocation ratios. Because m_h must be an integer and

$$m_h/m = \sqrt{V_{2h}/c_h} / \sum \sqrt{V_{2h}/c_h}$$

is generally a non-integer, in real life V_2^s , as given by the formula above, will only be an approximation to the optimum value V_2^s . This should not be of any practical concern, however, because m_h is taken as the nearest integer, and with sufficiently large m , the actual allocation ratios m_h/m are not too far from the optimum ratios.

(4) If instead of the optimum allocation ratio m_h/m one uses any arbitrary allocation $w_h = m_h/m$, one can still use our approach. However, in this case we define V_2^s by the formula

$$V_2^s = \Sigma (V_{2h}/w_h)$$

Once m is determined by our optimization formulae, one can apply the allocation ratios w_h to obtain the sample allocations m_h ; that is

$$m_h = (\text{nearest integer } \geq 2) \text{ to } mw_h$$

(5) When for all $h = 1, 2, \dots, L$, the stratum size M_h is large with respect to subsample size m_h , one can simplify the model by making $(M_h - m_h)/M_h = 1$ and, thus, by making $V_1^s = V_1$. The same simplification occurs when the biomass samples are overlapping and M_h is infinite in size. But then, one must use the relative size A_h of stratum h to define $V_{2h} = A_h^2 \sigma_{zz}$; the factors A_h allow one to go from the "per biomass" sample values z to the "per stratum" totals u_h , that is from μ_{zh} to $u_h = A_h \mu_{zh}$.

Model 5 - Stratified Random Subsampling with Ratio Estimators Within Strata.

With this subsampling method, the procedure to select the biomass samples is the same as in Model 4 above but (i) the biomass samples are measured for the auxiliary variables g_{hk} in addition to the variables of interest z_{hk} , (ii) the stratum totals G_h of g_{hk} are known for each h , and (iii) ratio estimators of the form $u_{rh} = G_h r_h$ (defined in Model 3) are used for u_h (instead of \hat{u}_h). Using the procedure and notation of Model 3, we can write

$$r_h = \Sigma z_{hk} / \Sigma g_{hk} = \bar{z}_h / \bar{g}_h$$

= estimator of the ratio $\rho_h (= u_h/G_h)$ of the true value u_h of the variable of interest to the true value G_h of the auxiliary variable for stratum h

$$B = (M_h - m_h) (\rho_h \sigma_{ggh} - \sigma_{zgh}) / m_h \mu_{gh}$$

= bias of ratio estimator u_{rh} of u_h ,

where $\mu_{gh} = G_h/M_h$ = mean of auxiliary variable g_{hk} ,

σ_{ggh} is the variance of g_{hk} , and σ_{zgh} is the covariance of z_{hk} and g_{hk} within stratum h ; and

$$\sigma_{u_{rh}u_{rh}} = \left(\frac{M_h - m_h}{M_h} \right) (V_{2h}^*/m_h)$$

= variance of u_{rh} as an estimator of
 u_h

where

$$V_{2h}^* = M_h^2 \Sigma (z_{hk} - \rho_{g_{hk}})^2 / M_h = M_h^2 (\sigma_{zzh} - 2\rho_{hg} \sigma_{zgh} + \rho_{hg}^2 \sigma_{ggh})$$

Define now

$$u_s = u_{r1} + u_{r2} + \dots + u_{rL} = \Sigma u_{rh}$$

= estimator of the true biomass value u
of the given tree

and

$$V_s = \Sigma \sigma_{u_{rh}u_{rh}} = \Sigma (M_h - m_h) (V_{2h}^*/m_h) / M_h$$

= variance of u_s

If we define

$$V_1^{S*} = V_1 - \Sigma (V_{2h}^*/m_h)$$

we can write the conditional variance of u_{rh} given [x] as

$$v = V_1^{S*} + \Sigma (V_{2h}^*/m_h)$$

The optimization problems for this model are the same as for the previous Model 4, with V_1^{S*} and V_{2h}^* substituted for V_1 and V_{2h} respectively. Consequently, the optimum sample sizes m_h and n can now be determined by the procedure outlined for Model 4. We have

$$m_h = m \sqrt{V_{2h}^*/c_h} / \Sigma \sqrt{V_{2h}^*/c_h} \quad \text{for } h = 1, 2, \dots, L$$

$$\Sigma m_h c_h = mc_s$$

$$\text{where } c_s = \Sigma \sqrt{V_{2h}^*/c_h} / \Sigma \sqrt{V_{2h}^*/c_h}$$

and

$$\Sigma (V_{2h}^*/m_h) = (\Sigma \sqrt{V_{2h}^*/c_h}) (\Sigma \sqrt{V_{2h}^*/c_h}) / m$$

$$= V_2^{S*} / m$$

Consequently, the cost and variance function can be written as

$$C = nc_t + mmc_s$$

and

$$K = (V_1^{S*} + V_2^{S*}/m)$$

respectively, and the optimum values m and n are given by the formulae of Model 1.

Summary Comments

We have considered the problem of optimizing the size of the subsample used to estimate the biomass of sample trees. We have assumed that the n sample trees were selected by simple random sampling, that each sample tree was subdivided into M biomass samples (small sections of the tree) and that m biomass samples were selected by simple random sampling without replacement and measured for the biomass variable of interest z and possibly for a second variable, denoted here as the auxiliary variable g , highly correlated with z . The selection of the n sample trees and the n subsamples (within the n trees) were all assumed to be statistically independent of each other.

To solve this problem, we have constructed an optimization model of the mathematical programming form, having an objective function to minimize and side conditions to satisfy. The decision variables are the sample and subsample size n and m , the parameters of the model are (i) the variance components V_1 (due to the variation of the true biomass about its regression function), and V_2 (due to the variation of the biomass sample values within the sample trees), (ii) the average sampling costs c_t per sample tree and c_s per biomass sample, and (iii) either a desired precision K of some biomass estimators or an allowable cost of sampling C . We have considered the two essentially equivalent problems of (i) minimizing the sampling costs C subject to the side condition that the precision of the parameter of interest is equal to K , or (ii) minimizing the variance of the estimator of interest subject to the side condition that the sampling costs are equal to C . The model had the additional side condition that the optimum sample sizes n and m must be positive integers.

After formulating the mathematical models, we have solved the models by calculus techniques of optimization. The solution shows that the optimum size m of the subsample does not depend on the number n of sample trees; it depends only on the variance components V_1 and V_2 as well as the cost parameters c_t and c_s . Once the subsample size m is determined, then the number n of sample trees is a function of the total variance function V (of the form $V_1 + V_2/m$) and either the desired precision $V/n = K$ or the allowable costs of sampling $n(c_t + mc_s) = C$.

Of course, the subsample size m depends also on the design used to subsample the tree biomass. We have started with the most simple case where the subsampling is done by simple random sampling with replacement or simple random sampling without replacement where the effect of the finite population correction factor can be ignored. We have then extended this model to include the effect of the finite population correction factor (when it cannot be ignored) and consider the case where the estimator of the tree biomass that uses the subsample average \bar{z} is replaced by a ratio type estimator. We have further extended the model by replacing the simple random by the stratified random sampling method of selecting the m biomass samples. We have also considered both estimators of the tree biomass; the one based on the means of various strata and the one based on ratio type estimators of the stratum totals. In all these extensions we have shown how to change the definition of the variance components V_1 and V_2 so that (i) the new components are consistent with sampling design and estimator used and (ii) the mathematical model to optimize does not change and the optimum solution of m and n remains the same function of the (new) components V_1 and V_2 .

The application of this approach to real life problems may present several difficulties or practical limitations. It is first necessary to obtain sufficiently good estimates of the average cost factors c_t and c_s as well as estimates of the variance components V_1 and V_2 . And these may be difficult to obtain. Some sensitivity analysis of the

model may be needed to determine how much variation in the estimates of c_t , c_s , V_1 , and V_2 can be tolerated without critically affecting the optimum solution. Furthermore, to keep the mathematical model sufficiently simple, we had also to assume that the variance component V_2 is the same (or does not change much) from tree to tree. Because (i) V_2 refers to the variance of the ratio $u = y/\sqrt{w}$ of the total biomass y of some tree component to some known weighting factor \sqrt{w} (a function of tree size and possibly species), and (ii) the variance of u is expected to be approximately homogeneous, it may imply that the assumption that V_2 does not vary much from tree to tree is not too critical.

For example, let y be the oven-dry biomass of the main tree bole and assume that the conditional variance of y given diameter d and height h is approximately proportional to d^4h^2 . Then, the variable $u = y/d^2h$, which has an approximately homogeneous variance, has a special meaning that may imply that the variation of V_2 from tree to tree is expected to be small. This can be explained as follows. It is known that the bole volume is approximately equal to some factor k multiplied by the quantity d^2h . Note that $\pi d^2/4h$ represents the volume of a cylinder of diameter d and length h and kd^2h represents the volume of the tree bole as a fraction of the corresponding cylinder. Consequently, dividing y by d^2h is nothing but expressing the biomass value y on a "per volume unit" basis. It is not necessary for the volume unit to be exactly a cubic foot; any unit would suffice. As (i) z is the M -th part of the value u of the main bole, and (ii) z , as a relative value represents "weight per volume unit," varies little from tree to tree, it seems reasonable to expect that the variation of σ_{zz} or $V_2 = M^2\sigma_{zz}$, if any, would be small.

The optimization model requires that M be the same from tree to tree. If m is small relative to M and the effect of the finite population correction factor can be ignored, the formulae of the optimum sample sizes m and n are not a function of M . This seems to imply that as long as M remains large with respect to m , one may use different values of M for different trees. This may affect slightly the variance term σ_{zz} and V_2 but it is not expected to affect it critically.²

The approach shown here can be extended to other methods of subsampling. It suffices to identify the form that V_1 and V_2 will take with these methods. The modification of the sampling technique to select the n sample trees may, however, change critically the optimization model. Instead of one sample size n , the model may have to be written as a function of several sample sizes n_1, n_2, \dots . For example, if the sample trees are selected by cluster sampling, the sample sizes may be n_1 = number of clusters and n_2 = number of trees per cluster, assuming that n_2 is the same from cluster to cluster.

Acknowledgments

This paper is based on research funded by the Research Foundation of the State University of New York, the United States Department of Agriculture Forest Service and the Department of Energy, Grant No. 23-524.

Literature Cited

- Cochran, W. G. Sampling Techniques, third edition. John Wiley and Sons, Inc., New York, NY. 1977.
- Cunia, T. On the error of biomass estimates in forest inventories: Part 1: Its major component. Faculty of Forestry Miscellaneous Publication Number 8, SUNY College of Environmental Science and Forestry, Syracuse, NY. 1986a.
- Cunia, T. On the error of biomass estimates in forest inventories: Part 2: The error component from sample plots. Faculty of Forestry Miscellaneous Publication Number 9, SUNY College of Environmental Science and Forestry, Syracuse, NY. 1986b.
- Valentine, H.T.; Tritton, L. M.; Furnival, G. M. Subsampling trees for biomass, volume or mineral content. Forest Science 30: 673-681. 1984.

ESTIMATING SAMPLE TREE BIOMASS BY SUBSAMPLING:

SOME EMPIRICAL RESULTS

R. D. Briggs, T. Cunia, E. H. White and H. W.

Yawney

Post-doctoral Research Associate, Professor of Statistics and Operations Research, Professor at SUNY College of Environmental Science and Forestry, Syracuse, NY, and Research Scientist, USDA-Forest Service, Northeastern Forest Experiment Station, Burlington, VT respectively

A procedure used to measure the green weight and estimate, by subsampling, the dry weight of the above ground components of 29 randomly selected sugar maple trees is described. Foliage and branch dry weights were determined by direct measurement. Bole wood and bole bark dry weight was estimated by stratified subsampling and subsequent application of ratio-type estimators. An analysis was made of the variation of subsamples within sample trees and variance components are calculated and listed for use in later studies.

Introduction

Regression functions of tree biomass on diameter outside bark at breast height (dbh) are estimated from samples of trees randomly selected from some forest population of interest. Once these trees are selected, their biomass must be measured, if possible, or estimated by subsampling otherwise. If the biomass is expressed as green weight, the tree component can be measured directly as soon as it is separated from the tree. The only possible error may be due to faulty measurement instruments or methods. If the biomass is expressed as dry weight, direct measurement may be too expensive and time consuming for large tree components. The only practical alternative is subsampling; small biomass samples are selected from the tree component by some random procedure. Green and oven-dry weights of these samples are determined in the laboratory and the results extrapolated from biomass samples to the entire tree component.

Note that we define here the "measurement" of biomass as the process of direct determination of the biomass of the entire tree component of interest and the "estimation" of biomass as the process of determination of the biomass by subsampling. The fresh weight of a component is measured if the entire component is weighed or estimated if parts of the component are weighed and the weights of the parts are extrapolated, by some computational procedure, to the entire component. We also define here the measurement error as the difference between the actual value as obtained by the measurement or estimation process and the true value of the parameter of

interest. We shall assume that this error is negligibly small when the biomass is measured but is sufficiently large when the biomass is estimated. The measurement error can be evaluated and expressed in statistical terms if the subsampling procedure is statistical.

The objective of the present paper is to describe the procedure used to measure and estimate the biomass of the aboveground components of 29 sugar maple (*Acer saccharum* Marsh.) trees selected at random from a 21 ha sugar maple stand on the Green Mountain National Forest in the town of Bristol, Addison County, Vermont. These trees were part of a large study designed to evaluate the effects of whole-tree and tree-length harvesting on future site productivity of northern hardwoods. Additional details of this study are provided by Briggs (1985). More specifically, we shall be concerned here with the measurement error due to the process of determining, by subsampling, the biomass of the sample trees. Different people, applying the same or different subsampling procedures will generally arrive at different tree biomass values and thus, arrive at different regression functions.

Subsampling Procedure for Estimating Tree Biomass

The 29 sugar maple trees were all measured first for their diameter at breast height (dbh), total height (h), and merchantable height to a 10 cm minimum top diameter. Each tree was then felled down and ten plastic sheets were placed on the ground surrounding the tree. Beginning at the base of the crown and working towards the top, the tree branches with their leaves attached were removed in groups of ten, so that each branch in a given group had approximately the same amount of branchwood and foliage (as estimated ocularly) and came from a similar height in the crown. One branch from each group was put in a different pile so that for each tree, ten piles were obtained; each pile having a similar distribution of branches and foliage with respect to weight and point of origin from the crown.

For each of the ten piles, all of the foliage was picked from the branches, placed in paper bags and carried to the truck. The branches from each pile were tied together in canvas tarpaulins and carried to the landing. In order to facilitate transportation and oven-drying, each pile of branches was processed separately through a small chipper operated from the PTO (power take-off) shaft of a farm tractor. The chips were collected and bagged separately by pile and transported to the laboratory.

The bole of each sample tree was divided into three sections of equal length. For each section, three integers were randomly selected from 1 to 100. Each of these numbers was multiplied (as a percent) by the section length to obtain the location of a sample disk for the determination of the fresh and dry weight. For example, if the random number was 32 and the section length was 5.0 m, then a disk would be

located at $(.32) \times (5.0) = 1.50$ m from the base of that section. Each disk location was marked with flagging.

Each of the three bole sections was cut into logs of various lengths and weighed on a 90 kg capacity field scale. Disks, approximately 5 cm in width, were removed from the bole (at the marked locations) after each section was completely weighed, placed in double plastic bags and carried to the truck. At the end of each day, the sample disks were weighed in the field office and transferred to paper bags.

All sample material (foliage, branches, disks) was transported to the Forest Service Laboratory in Burlington, Vermont and placed in a greenhouse. The greenhouse was used as a holding area which allowed for partial drying, preventing losses of dry weight due to microbial decomposition. At the end of the field season, all sample material was transported to the SUNY College of Environmental Science and Forestry, Syracuse, N.Y. Foliage, branches and disks were placed in forced air kilns at 65° C until constant weight was obtained. The oven-dry weight was determined

individually for each pile of branches and foliage, as well as for each disk of each individual sample tree. The bark was removed from each disk, dried at 65° C and its weight was recorded.

The measured tree data including dbh, total height and height to a 10 cm minimum top diameter, total bole length and the green weight (wood and bark) for each of the three bole sections are shown in (Table 1.) The green and dry weight data for wood and bark of each of the nine sample disks per tree can be found in Briggs et al. (1986). Finally, the mean oven-dry weights for the ten piles of foliage and branches, along with their variances, are shown in Table 2.

Estimation of Bole Biomass

Determination of the green weight of the tree bole is made by direct measurement of its weight. If the loss of humidity between tree felling and bole weighing is negligibly small and the weighing scale is accurate, the measurement error of the bole biomass, expressed as green

Table 1. Morphological data for the 29 sample sugar maple trees.

Tree	Dbh cm	Total Height m	Height to 10 cm Top m	Bole Length m	Green Weight of Bole (wood + bark)			
					Bot	Mid	Top	Total
-----kg-----								
1	25.1	15.7	15.4	14.8	196.8	142.1	76.0	414.9
2	43.4	19.9	14.0	13.9	595.5	493.9	179.3	1268.7
3	33.0	20.7	15.7	15.5	374.4	316.8	174.3	865.5
4	41.9	20.4	15.2	14.7	544.8	415.9	223.5	1184.2
5	17.6	19.5	11.9	11.5	90.1	77.3	50.1	217.5
6	16.7	19.8	12.4	12.1	75.4	52.5	33.0	160.9
7	38.1	23.5	18.2	17.0	526.1	434.0	204.3	1164.4
8	10.2	16.4	4.4	4.1	10.9	9.1	7.9	27.9
9	32.5	23.2	19.9	19.6	475.5	345.2	131.7	952.4
10	38.6	22.6	18.1	17.6	550.3	408.4	164.6	1123.3
11	41.2	23.4	17.6	17.3	603.0	428.8	239.2	1271.0
12	25.4	21.7	15.2	14.9	196.2	133.7	77.9	407.8
13	32.5	22.0	17.4	17.0	426.8	275.0	142.8	844.6
14	9.1	13.0	2.2	2.2	5.0	5.4	5.4	15.8
15	32.5	24.0	17.8	17.4	417.1	340.4	205.9	963.4
16	25.9	22.0	16.3	16.0	238.0	162.7	92.4	493.1
17	9.1	12.4	1.8	1.5	3.8	3.1	2.8	9.7
18	17.2	18.5	10.7	10.4	74.4	47.2	29.9	151.5
19	39.4	21.8	16.4	16.1	597.5	418.4	235.4	1251.3
20	26.2	22.0	15.6	15.3	233.1	164.5	106.4	504.0
21	37.3	23.4	17.4	17.0	508.3	393.9	189.9	1092.1
22	43.9	22.2	20.8	20.4	812.5	567.9	185.4	1565.8
23	38.6	23.5	18.5	18.0	631.6	493.3	244.5	1369.4
24	9.9	13.7	3.5	3.1	11.3	4.1	6.8	22.2
25	26.4	22.8	16.5	16.1	259.1	175.5	106.3	540.9
26	33.0	22.2	17.6	17.3	428.4	317.3	185.4	931.1
27	42.7	25.2	18.4	18.0	768.6	656.6	420.3	1845.5
28	38.9	23.5	18.4	17.9	589.7	435.4	191.5	1216.6
29	42.2	25.4	19.9	19.5	657.0	487.8	314.4	1459.2

Table 2. Total and average dry weights for the ten piles of foliage and branches for each of the 29 sample sugar maple trees.

Tree	Foliage			Branches		
	Total	Average	Variance Between Piles	Total	Average	Variance Between Piles
	-----kg-----		kg ²	-----kg-----		kg ²
1	7.48	0.748	0.014856	81.56	8.156	3.875084
2	24.10	2.410	0.076900	363.98	36.398	16.106596
3	16.09	1.609	0.113369	126.88	12.688	17.074796
4	16.03	1.603	0.149081	204.44	20.444	6.623464
5	3.48	0.348	0.009996	45.09	4.509	1.248909
6	2.89	0.289	0.006409	28.47	2.847	1.610921
7	12.68	1.268	0.044956	208.28	20.828	13.249076
8	0.61	0.061	0.000369	21.77	2.177	0.282821
9	11.07	1.107	0.053941	130.81	13.081	2.884529
10	13.55	1.355	0.090625	176.90	17.690	4.592060
11	14.23	1.423	0.264621	178.37	17.837	17.018581
12	5.83	0.583	0.020241	64.06	6.406	3.130104
13	8.07	0.807	0.033021	110.69	11.069	4.050209
14	0.68	0.068	0.001416	21.30	2.130	0.880420
15	8.56	0.856	0.014764	87.79	8.779	2.467629
16	4.74	0.474	0.018304	67.50	6.750	3.688480
17	0.56	0.056	0.000384	20.30	2.030	0.458400
18	2.41	0.241	0.002629	26.94	2.694	0.246784
19	25.81	2.581	0.228229	335.42	33.542	32.040436
20	6.47	0.647	0.016781	70.14	7.014	2.392004
21	3.81	0.381	0.006669	121.29	12.129	6.776149
22	23.11	2.311	0.056009	295.50	29.550	5.153040
23	11.96	1.196	0.037364	175.85	17.585	21.395305
24	0.73	0.073	0.000641	19.26	1.926	0.793304
25	6.53	0.653	0.023201	87.24	8.724	5.915184
26	14.06	1.406	0.037444	153.75	15.375	1.502325
27	25.47	2.547	0.300041	482.82	48.282	47.007136
28	8.65	0.865	0.040385	163.33	16.333	16.919021
29	8.85	0.885	0.018745	185.29	18.529	15.937229

weight, is equal to zero. On the other hand, the ovendry weight of the bole is determined by subsampling. As the three bole sections can be considered as strata, and three disks are selected at random from each section, the method of disk selection is stratified random sampling. Because the green weight of the entire bole, individual sections and disks are known and the ovendry weight of the sample disks is measured, one can estimate ovendry weight of the bole by a stratified ratio estimator.

Let us show in detail the estimation procedure for the ovendry weight of the bole (wood plus bark). The estimation of the ovendry weight of wood alone or bark alone would follow the same general lines. It is then sufficient to show only which variables are substituted into the corresponding formulae. The following notation and definitions of estimators are being used:

G = total green weight of the entire tree bole

G_h = green weight of section h (1 for bottom, 2 for middle and 3 for top section)

g_{hk} = green weight of wood and bark of disk $k = 1, 2, 3$ in stratum h

d_{hk} = ovendry weight of wood and bark of disk $k = 1, 2, 3$ in stratum h

$m_h = 3$ = number of sample disks in each section $h = 1, 2, 3$

$\bar{g}_h = \Sigma g_{hk}/n_h = (g_{h1} + g_{h2} + g_{h3})/3$

= average green weight per sample disk within stratum $h = 1, 2, 3$

$$\bar{d}_h = \sum d_{hk} / n_h = (d_{h1} + d_{h2} + d_{h3}) / 3$$

= average oven-dry weight per sample disk within stratum $h = 1, 2, 3$

$$M_h = G_h / \bar{g}_h = \text{conceptual number (not necessarily integer) of disks of average weight } \bar{g}_h \text{ that can be made from section } h = \text{size of section } h$$

$$(M_h - m_h) / M_h = \text{finite population correction factor of section } h$$

$$S_{dd,h} = \sum (d_{hk} - \bar{d}_h)^2 / (m_h - 1) = \text{sample variance of the } m_h \text{ disk values } d_{hk} \text{ within section } h$$

$$S_{gg,h} = \sum (g_{hk} - \bar{g}_h)^2 / (m_h - 1) = \text{sample variance of the } m_h \text{ disk values } g_{hk} \text{ within section } h$$

$$S_{dg,h} = \sum (d_{hk} - \bar{d}_h)(g_{hk} - \bar{g}_h) / (m_h - 1) = \text{sample covariance of the } m_h \text{ pairs of disk values } d_{hk} \text{ and } g_{hk} \text{ within section } h$$

$$r_h = \bar{d}_h / \bar{g}_h = \text{estimator of the ratio of the oven-dry weight to the green weight of section } h$$

$$D_h = G_h r_h = M_h \bar{d}_h = \text{ratio estimator of the oven-dry weight of section } h$$

$$B_h = (M_h - m_h)(r_h S_{gg,h} - S_{dg,h} / m_h \bar{g}_h) = \text{estimator of the bias of } D_h$$

$$S_{DD,h} = M_h (M_h - m_h) (S_{dd,h} - 2r_h S_{dg,h} + r_h^2 S_{gg,h}) / m_h = \text{estimator of the variance of } D_h$$

Note that, in the formulae above, the section (stratum) size M_h is a random variable; it is defined as a function of the random variable \bar{g}_h . Furthermore the disk material is not selected with equal probability; the material from the smaller end diameter of the bole is sampled with a higher probability. Consequently, our "ratio" estimator is somewhat different from the classical ratio estimators as they are commonly defined in standard textbooks, where M_h is fixed and the sampling is done with equal probability.

Because D_1 , D_2 and D_3 are statistically independent random variables, the oven-dry weight of the entire tree bole and its error can be estimated by the following formulae:

$$D = \sum D_h = D_1 + D_2 + D_3 = \text{stratified ratio estimator of the oven-dry weight of the wood and bark of the tree bole}$$

$$B = \sum B_h = B_1 + B_2 + B_3 = \text{estimator of the bias of } D$$

$$S_{DD} = \sum S_{DD,h} = \text{estimator of the variance of } D$$

$$\sqrt{S_{DD}} = \sqrt{\sum S_{DD,h}} = \text{estimator of the standard error of } D, \text{ and}$$

$$D \pm 2\sqrt{S_{DD}} = \text{approximate 95 percent confidence limits of the oven-dry weight of the tree bole}$$

If we define now:

$$d_{hk,w} = \text{oven-dry weight of the wood of disk } k \text{ of section } h, \text{ and}$$

$$d_{hk,b} = \text{oven-dry weight of the bark of disk } k \text{ of section } h, \text{ and}$$

if these values are substituted for the variable d_{hk} of the formulae above, one can define the estimators D_w and D_b , the stratified ratio estimators of the oven-dry weight of wood and bark, respectively, of the tree bole, as well as the corresponding estimators of their biases, variances and confidence limits.

In order to conserve space and still provide the necessary details required to demonstrate the estimation procedure for sample tree biomass and its associated error, intermediate statistics are provided for the bole component of only 3 of the 29 sample trees (1, 14, 29). The statistics \bar{g}_h , \bar{d}_h , $S_{dd,h}$, $S_{dg,h}$, $S_{gg,h}$, r_h , D_h , B_h and $S_{DD,h}$ of the oven-dry weight of the wood and bark for the bottom, middle and top sections of the bole for the three selected sample trees are shown in Table 3. The summary statistics D , B , S_{DD} , $\sqrt{S_{DD}}$ and the approximate 95 percent confidence limits for the bole are shown for all 29 trees in Table 4. The summary statistics for the bolewood and bole-bark oven-dry weights are shown in Tables 5 and 6, respectively.

Analysis of Biomass Data and Their Error

This study is concerned with the measurement of the biomass of tree foliage, branches, bole-wood, bole-bark and bole (wood + bark). Only the last tree component was measured for its green weight. We expect a small loss of humidity, if any, between the time of tree felling and bole weighing, as well as a small loss, if any, of wood or branch material due to bole sectioning prior to measurement of green weight. We also expect a sufficiently high accuracy and precision of the field scale used and we do not expect that any blunders were made by the field people when the bole was weighed, (i.e., misreading the scale or error in recording weights). Consequently, we should assume here that the green weight of the main tree bole as shown in Table 1 represents a true value without any measurement error.

Similarly, we shall also assume that the oven-dry weight of foliage and branches as shown in Table 2 has no measurement error. This means that (i) the possible loss of foliage or branch material from the time that the tree was felled in the field to the time of measurement in the laboratory is negligibly small, (ii) the measuring instruments have high precision and accuracy and (iii) no blunders were made when the oven-dry weight was measured and recorded.

Because the foliage and branch material was divided into ten piles and the biomass of each pile was measured and recorded separately by pile, it is possible to make inferences about the

Table 3. The statistics associated with the estimation of the bole oven-dry weight for the three sections of three of the twenty nine sample trees.

Tree	\bar{d}_h	\bar{g}_h	$S_{dd,h}$	$S_{dg,h}$	$S_{gg,h}$	r_h	D_h	B_h	$S_{DD,h}$
	-----kg-----			-----kg ² -----			-----kg-----		kg ²
Bottom									
1	1.13173	1.79877	0.10058230	0.14796983	0.21774336	0.629	123.82	-0.216	2.2534
14	0.22237	0.35087	0.00035754	0.00033079	0.00044217	0.634	3.17	-0.001	0.0062
29	1.48857	2.45100	0.03355816	0.07364185	0.16421700	0.607	399.02	0.941	110.8305
Middle									
1	0.77747	1.25107	0.00398492	0.00903385	0.02048281	0.621	88.31	0.109	2.7933
14	0.35707	0.57377	0.01694665	0.02708867	0.04334362	0.622	3.36	0.000	0.0003
29	1.35097	2.15000	0.07199929	0.13392440	0.25207900	0.628	306.51	0.849	54.5814
Top									
1	0.47660	0.77373	0.01720093	0.02637653	0.04049172	0.616	46.81	-0.059	0.2181
14	0.22793	0.37467	0.00020024	0.00029754	0.00044804	0.608	3.29	0.000	0.0002
29	0.54330	0.90333	0.18934599	0.32075240	0.54350933	0.601	189.09	0.781	4.9239

Table 4. Summary statistics associated with the calculation of D = estimate of the oven-dry weight for the entire tree bole (wood + bark) for the 29 sample sugar maple trees.

Tree	D	Bias	S_{DD}	$\sqrt{S_{DD}}$	95% Confidence Limits	
					Lower	Upper
	-----kg-----		kg ²		-----kg-----	
1	258.94	-0.166	5.2648	2.29	254.36	263.52
2	782.25	-0.189	38.8910	6.24	769.77	794.73
3	563.30	0.321	599.3665	24.48	514.34	612.26
4	716.41	11.652	1633.0545	40.41	635.59	797.23
5	134.28	0.070	1.3178	1.15	131.98	136.58
6	101.92	0.035	1.6132	1.27	99.38	104.46
7	734.93	-1.154	39.3400	6.27	722.39	747.47
8	17.70	0.000	0.0104	0.10	17.50	17.90
9	599.80	1.032	140.7951	11.87	576.06	623.54
10	727.80	-0.030	83.8443	9.16	709.48	746.12
11	809.70	1.022	48.4170	6.96	795.78	823.62
12	254.41	0.033	8.8941	2.98	248.45	260.37
13	575.41	-1.143	44.5967	6.68	562.05	588.77
14	9.82	-0.001	0.0067	0.08	9.66	9.98
15	600.49	-0.931	338.7310	18.40	563.69	637.29
16	318.11	-0.065	10.5999	3.26	311.59	324.63
17	6.22	-0.001	0.0007	0.03	6.16	6.28
18	90.27	0.003	0.3705	0.61	89.05	91.49
19	777.32	-0.799	59.1263	7.69	761.94	792.70
20	316.66	0.485	30.1318	5.49	305.68	327.64
21	714.69	0.341	192.0900	13.86	686.97	742.41
22	994.01	1.104	185.1251	13.61	966.79	1021.23
23	856.00	1.113	45.3786	6.74	842.52	869.48
24	14.06	-0.050	0.0187	0.14	13.78	14.34
25	339.61	0.372	27.6807	5.26	329.09	350.13
26	591.61	1.260	346.5537	18.62	554.37	628.85
27	1158.86	-1.661	230.1460	15.17	1128.52	1189.20
28	1767.45	0.178	140.3759	11.85	1743.75	1791.15
29	894.62	2.571	170.3358	13.05	868.52	920.72

Table 5. Summary statistics associated with the calculation of D = estimate of the oven-dry weight for the entire tree bolewood for the 29 sample sugar maple trees.

Tree	D	Bias	S _{DD}	$\sqrt{S_{DD}}$	95% Confidence Limits	
					Lower	Upper
	-----kg-----		kg ²		-----kg-----	
1	234.80	-0.253	5.8342	2.42	229.96	239.64
2	710.88	-0.998	47.6966	6.91	697.06	724.70
3	506.01	0.314	591.9527	24.33	457.35	554.67
4	639.79	8.419	1090.6481	33.02	573.75	705.83
5	122.11	0.028	0.5103	0.71	120.69	123.53
6	90.55	0.020	1.1864	1.09	88.37	92.73
7	647.23	-1.378	46.9841	6.85	633.53	660.93
8	16.08	-0.001	0.0007	0.03	16.02	16.14
9	545.02	0.153	84.5232	9.19	526.64	563.40
10	659.37	-0.506	29.4932	5.43	648.51	670.23
11	731.39	0.778	42.1788	6.49	718.41	744.37
12	226.39	-0.161	10.4214	3.23	219.93	232.85
13	505.28	-0.594	32.1162	5.67	493.94	516.62
14	8.33	-0.002	0.0054	0.07	8.19	8.47
15	535.40	-1.491	326.4250	18.07	499.26	571.54
16	288.13	-0.182	11.6678	3.42	281.29	294.97
17	5.59	-0.001	0.0007	0.03	5.53	5.65
18	77.99	-0.030	0.7554	0.87	76.25	79.73
19	690.23	-0.548	21.2706	4.61	681.01	699.45
20	287.61	0.306	27.1924	5.21	277.19	298.03
21	630.19	0.072	167.6352	12.95	604.29	656.09
22	904.14	0.876	182.4555	13.51	877.12	931.16
23	768.70	1.241	54.5529	7.39	753.92	783.48
24	12.36	-0.049	0.0188	0.14	12.08	12.64
25	306.27	0.236	25.2890	5.03	296.21	316.33
26	535.29	0.876	266.2677	16.32	502.65	567.93
27	1040.24	-1.599	252.8985	15.90	1008.44	1072.04
28	690.02	-0.212	141.0389	11.88	666.26	713.78
29	799.75	1.767	131.0079	11.45	776.85	822.65

Table 6. Summary statistics associated with the calculation of D = estimate of the oven-dry weight for the entire tree bolebark for the 29 sample sugar maple trees.

Tree	D	Bias	S _{DD}	$\sqrt{S_{DD}}$	95% Confidence Limits	
					Lower	Upper
	-----kg-----		kg ²		-----kg-----	
1	24.14	0.087	0.2401	0.49	23.16	25.12
2	71.37	0.809	10.2646	3.20	64.97	77.77
3	57.29	0.008	0.7573	0.87	55.55	59.03
4	76.63	3.232	91.8978	9.59	57.45	95.81
5	12.18	0.041	0.2591	0.51	11.16	13.20
6	11.36	0.015	0.2485	0.50	10.36	12.36
7	87.70	0.226	6.4143	2.53	82.64	92.76
8	1.61	0.001	0.0069	0.08	1.45	1.77
9	54.78	0.880	31.8935	5.65	43.48	66.08
10	68.43	0.477	44.3247	6.66	55.11	81.75
11	78.31	0.244	1.5695	1.25	75.81	80.81
12	28.02	0.194	0.8577	0.93	26.16	29.88
13	70.12	-0.549	60.7034	7.79	54.54	85.70
14	1.49	0.001	0.0001	0.01	1.47	1.51
15	65.09	0.560	4.1230	2.03	61.03	69.15
16	29.97	0.116	0.4512	0.67	28.63	31.31
17	0.63	0.000	0.0000	0.00	0.63	0.63
18	12.28	0.033	0.1117	0.33	11.62	12.94
19	87.09	-0.250	18.0980	4.25	78.59	95.59
20	29.06	0.178	0.7317	0.86	27.34	30.78
21	84.50	0.268	1.5386	1.24	82.02	86.98
22	89.87	0.228	10.6607	3.27	83.33	96.41
23	87.31	-0.128	1.2501	1.12	85.07	89.55
24	1.69	0.000	0.0000	0.00	1.69	1.69
25	33.35	0.136	0.9099	0.95	31.45	35.25
26	56.32	0.385	7.1535	2.67	50.98	61.66
27	118.63	-0.061	11.8567	3.44	111.75	125.51
28	77.43	0.391	4.8309	2.20	73.03	81.83
29	94.87	0.804	7.4899	2.74	89.39	100.35

measurement error when a few rather than all piles are being measured. If $m < 10$ piles are selected by simple random sampling without replacement, the biomass of the tree foliage or branches can be estimated by $Y = 10 \bar{y}$, where Y is the estimate of the tree biomass and \bar{y} is the average of the biomass of the m sample piles. The variance of Y can be estimated by the formula

$S_{yy} = (10-m) S_{yy}/10 m$, where S_{yy} is the variance of the estimator Y and S_{yy} is the variance of the biomass values y of the 10 individual piles. Let us analyze more closely the behavior of S_{yy} for both foliage and branch biomass. Because the piles are formed more or less arbitrarily, and it is natural to expect variation in the pile biomass from tree to tree, the values S_{yy} can be viewed as random variables.

Casual examination of the values S_{yy} for the foliage and branch biomass of the 29 sample trees suggests that S_{yy} and tree size are correlated; as the tree increases in size, so does its variance (Tables 1 and 2). This is not surprising since it is empirically known that large things tend to vary more than small things. To better see this relationship, foliage variance S_{yy} was plotted against dbh (Figure 1) and against foliage biomass Y (Figure 2). In both cases the variance S_{yy} seems to increase exponentially with tree size.

It was also felt that the variance between pile biomass may be proportional to the fourth power of tree dbh. As this hypothesis is identical to the hypothesis that the standard deviation is proportional to squared dbh, Figure 3 shows the graph of standard deviation against the squared dbh value. As the reader can verify, this hypothesis seems to be approximately true; a straight line passing through the origin of the two axes seems to be a sufficiently good approxi-

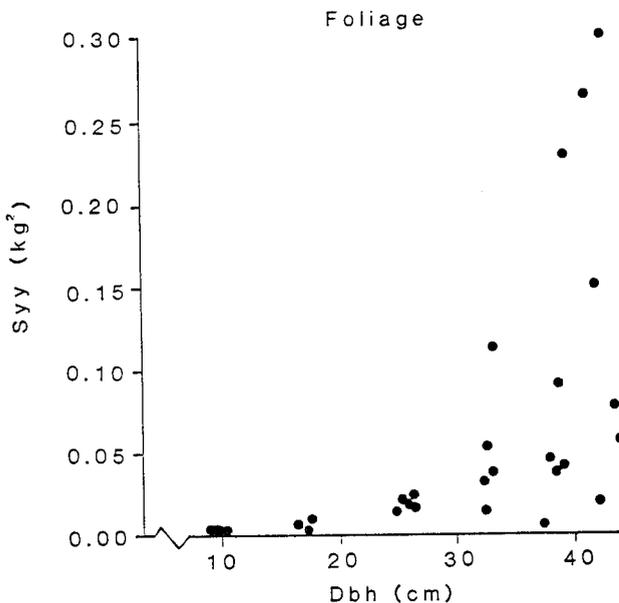


Figure 1. The variance of the foliage biomass (ovendry weight) of the ten piles plotted against the tree diameter.

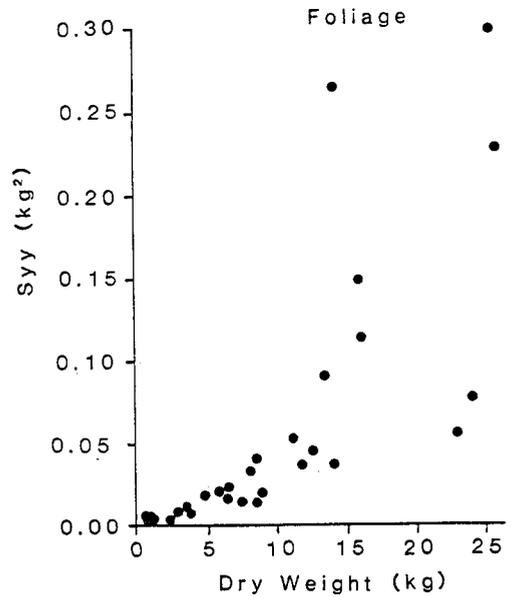


Figure 2. The variance of foliage biomass (ovendry weight) of the ten piles plotted against the total ovendry weight of the tree foliage.

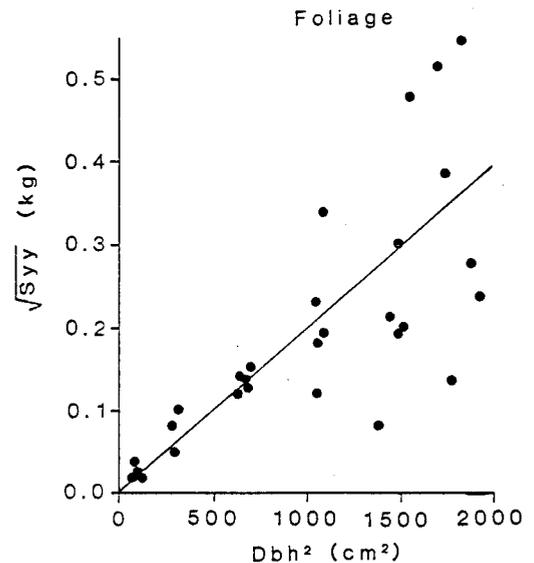


Figure 3. Standard deviation of the foliage biomass (ovendry weight) of the ten piles plotted against the squared value of tree diameter.

mation of this relationship. Similar conclusions can be drawn from the analysis of the ovendry weight of branches.

Let us analyze now the variation of the stratified ratio estimator of the bole ovendry weight (wood + bark). The variance ranges from 0.0007 to 1633.0545, corresponding to the smallest and largest sample trees, respectively (Table 4). As was the case with foliage and branch biomass, we have plotted the variance of the estimator of the bole biomass against tree dbh

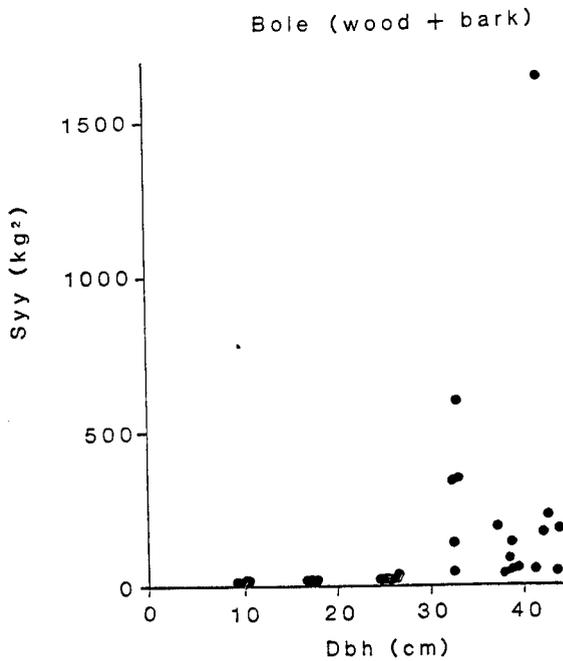


Figure 4. The variance of the stratified ratio estimator of the bole oven-dry weight (wood + bark) plotted against the tree diameter.

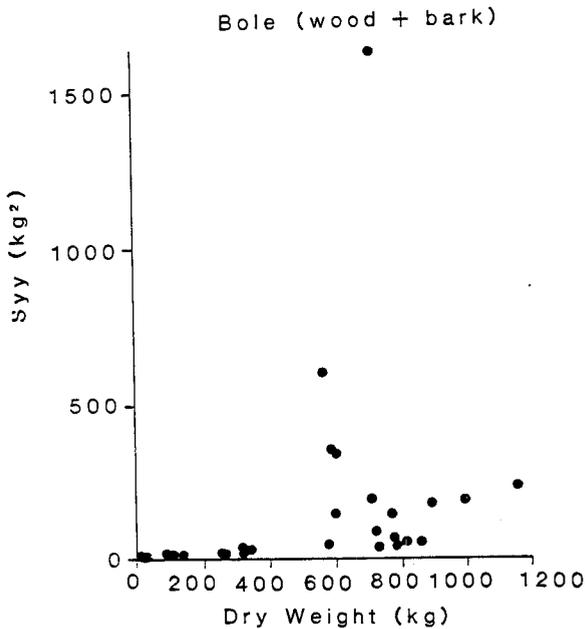


Figure 5. The variance of the stratified ratio estimator of the bole oven-dry weight (wood + bark) plotted against biomass.

(Figure 4) and against bole biomass itself (Figure 5). The graph of the standard error of the biomass estimator against the squared value of dbh (Figure 6) shows that the assumption that the variance estimator of bole biomass is approximately proportional to the fourth power of dbh is reasonable. Similar analyses for the bolewood and bole-bark components lead to the same conclusions.

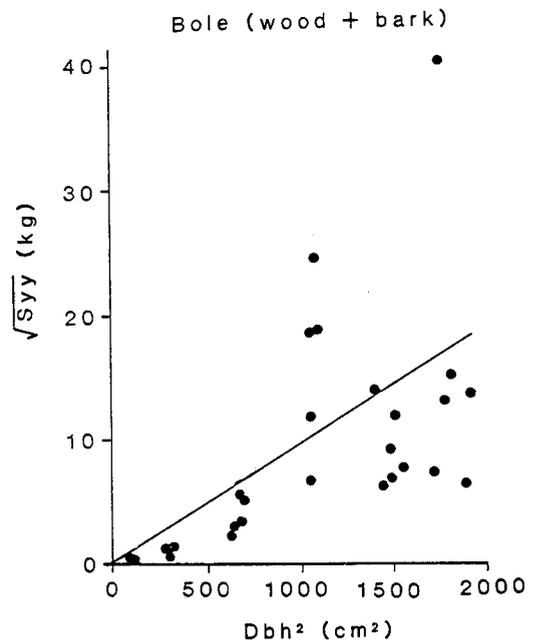


Figure 6. The standard error of the bole oven-dry weight (wood + bark) plotted against the squared value of tree diameter.

It is known that the ratio estimators are biased. As the estimator of the bole biomass is based on a sample size of nine disks, we would expect the bias to be small. Results from Table 4 verify this expectation; with very few exceptions the bias is much less than one percent of the biomass estimate. In terms of the standard error, however, the estimated bias is about seven percent for the first tree, about one percent for the fourteenth tree, and about 20 percent for the twenty-ninth tree. Whether a bias of this size can be tolerated is a question for management to answer. We feel that in terms of the resulting biomass table we construct, this bias is negligibly small. Similar conclusions can be reached from the corresponding analyses for the bolewood and bole-bark oven-dry weight data.

Concluding Remarks

It is generally expensive and time consuming to determine the biomass of sample trees by direct measurement. Most of the time, it is sufficiently precise to estimate the biomass by subsampling. This can be done in many ways. We have described the specific subsampling methods used in a study designed to evaluate the effects of some harvesting procedure on future site productivity.

The oven-dry weight of the tree foliage and branches was found by direct measurement. All the material was oven-dried and weighed. To facilitate the handling of this material, but much more importantly to obtain data that can be used later in the optimization of the subsampling method, the foliage and branches were divided into ten piles and the oven-drying, weighing and

recording were done separately by pile. By considering the estimation error when $m < 10$ piles only are sampled, one can make a decision as to how many sample piles would be sufficient to satisfy precision requirements.

The biomass of the main tree bole expressed as green weight can be determined easily by direct measurement; there is no need for subsampling. It requires the use of a field scale, and only when the tree is large requires sectioning, with each section being weighed separately. The measurement error is expected to be equal to zero if the weighing is carefully done. The biomass of the same bole, expressed as oven-dry weight, is however, much more time consuming and expensive to measure. For this reason it is necessary to determine oven-dry weight by subsampling.

The subsampling method used in this study was stratified random sampling and the estimator used was that of the stratified ratio type. The variance within strata was estimated for both green and oven-dry weight of the subsample units (disks) as well as their covariance and ratios. These statistics can be used in later study to decide the amount of subsampling that one may need to obtain estimates of the oven-dry weight of the main bole, should one decide to use stratified sampling and ratio estimators of the type used in this study.

It is in this sense that one must view the value of this paper. It provides information about the variation within sample trees. Quantitatively expressed, this information can be used in later studies to determine the amount of subsampling needed. Then, optimization models of the type suggested by Cunia (1986) can be used to find out at what point one should stop subsampling and add more sample trees so as to obtain maximum information for given costs of sampling or to minimize these costs and obtain estimates of desired precision.

Acknowledgements

This paper is based on research funded by the Research Foundation of the State of New York, the United States Department of Agriculture Forest Service, and the Department of Energy Grant No. 23-524 and Grant No. 23-708.

Literature Cited

- Briggs, R. D.; Cunia, T.; White, E. H.; and Yawney, H. W. Estimating sample tree biomass by subsampling. Syracuse, NY: SUNY College of Environmental Science and Forestry; Faculty of Forestry Misc. Publ. In Review. 1986.
- Briggs, R. D. Estimating biomass and nutrient removals from whole-tree and tree-length harvesting of a northern hardwood stand. Syracuse, NY: Ph.D. Thesis; 1985. 264 p.

Cunia, T. An optimization model for subsampling trees for biomass measurement. Paper presented at the Syracuse, NY, May 26-30, 1986 Workshop on Tree Biomass Regression Functions and Their Contribution to the Error of Forest Inventory Estimates. 1986.

UNBIASED ESTIMATION OF TOTAL TREE WEIGHT BY
THREE-STAGE SAMPLING WITH PROBABILITY
PROPORTIONAL TO SIZE

Harry T. Valentine, Timothy G. Gregoire, and
George M. Furnival

Research Forester, U.S. Forest Service, Hamden,
CT 06514, Assistant Professor of Forest
Biometrics, Virginia Polytechnic Institute &
State University, Blacksburg, VA 24061 and J.P.
Weyerhaeuser, Jr., Professor of Forest
Management, Yale University, New Haven, CT
06511.

A three-stage procedure for estimating the total tree weight in a plot employs discrete sampling with probability proportional to size, randomized branch sampling, and importance sampling--a continuous analogue of sampling with probability proportional to size. Published regression functions that predict tree weight from diameter breast height can be used to assign probabilities of selection to the trees for the first-stage sampling.

Introduction

In ecological or forestry research, a need often arises for an estimate of the total weight of the trees in a research or survey plot. The biomass literature is rife with regression functions that estimate tree weight from diameter. Consequently, an investigator might be tempted to measure the diameters of the trees in a plot, use published regression functions to estimate their individual weights, and then sum these weights to obtain the needed estimate of total tree weight. Unfortunately, the model-based estimate of total tree weight obtained by this simple procedure may be biased if the regression functions do not derive from the population to which they are applied. Bias also may arise if the population of current interest is a subpopulation of the parent population from which the regression functions derive. This bias may be large or small and of either sign, and the absolute magnitude of this bias does not decrease with increasing sample size.

In this paper, we describe a procedure that furnishes a sample-based unbiased estimate of the total weight of the trees in a plot. Initially, a sample of trees is selected at random with probability proportional to size (PPS). The weights of the trees in this sample are either measured or unbiasedly estimated in optional second and third stages of sampling. The second and third stages employ procedures for subsampling trees developed by Valentine et al. (1984).

Published regression functions may be used in connection with the first-stage PPS sampling. The probability of selection of a tree in the plot can be made proportional to weight as predicted by a regression function. Any regression function(s) could be used for this purpose. However, the precision of the PPS estimator depends on the degree of correlation between the predicted weights and the actual weights. Better correlation gives better precision.

First-Stage Sample

The use of PPS sampling for the estimation of the total tree weight in a plot has become feasible with the advent of programmable pocket calculators and field-worthy portable microcomputers. To proceed with PPS sampling, the N trees in a plot are numbered consecutively as their diameters are measured. Regression functions that predict tree weight from diameter are programmed into a calculator or computer and used to predict the weight of each tree. The predicted weight of the tree number i (i = 1, 2, ..., N) is denoted by \hat{w}_i . The conventional model-based estimate of the total (fresh or dry) weight of the trees in the plot is obtained by summing the individual (fresh or dry) predicted weights. To obtain our sample-based estimate, we select a sample of n trees at random (with replacement) with probability proportional to predicted weight as follows:

The total predicted weight of the trees number 1 through i (C_i is accumulated for $i = 1, 2, \dots, N$):

$$C_i = \sum_j^i \hat{w}_j \quad i=1, 2, \dots, N \quad (1)$$

We also define $C_0=0$. To select the kth of n trees in our sample, a random number, u_k , is drawn from a uniform [0,1] distribution. The tree numbered i is selected as the kth tree in our sample if

$$C_{i-1} < u_k C_N \leq C_i \quad i=1, 2, \dots, N$$

We denote the tree (fresh and dry) weight of this tree numbered i by w_i . An unbiased estimate of the total tree weight in the plot, based on this kth tree in our sample, is

$$\tilde{w}_k = C_N w_i / \hat{w}_i \quad (2)$$

An unbiased estimate of the total tree weight based on a sample of n trees is

$$\tilde{W} = \sum_k^n \tilde{W}_k / n \quad (3)$$

and a sample-based estimator of the variance (var) of \tilde{W} is

$$\text{var}(\tilde{W}) = \sum_k^n (\tilde{W}_k - \tilde{W})^2 / [n(n-1)] \quad n \geq 2 \quad (4)$$

Proofs of unbiasedness of PPS estimators can be found in most sampling texts (e.g., Cochran 1963, Sukhatme and Sukhatme 1970).

By performing PPS sampling as was described, we take advantage of the information extant in published regression functions, and we also can take advantage of a built-in safeguard against error. If there is a suspiciously large difference between \tilde{W} and the conventional model-based estimate, C_N , then we should check for errors in measurement or entry of data into the calculator or computer.

There may be occasions when published regressions are not available for the species of interest, in which case other information might be used to advantage in a PPS sampling. For example, in 1927, C.D. Murray reported that the weight of a tree above a given cross section was proportional to the diameter (d) of that cross section raised to a power of 2.5. The report was based on 116 measurements of whole trees, branches, and leaves from nine tree species. In 1981, J. White cited 50 publications in which tree weight has been reported as proportional to diameter breast height (D) raised to a power (b). With just two exceptions, all values of b were between 2 and 3. White, like Murray 54 years before, concluded that 2.5 is a reasonable approximation of b for many tree species. White also noted that tree volume is correlated with $D^{2.5}$. Thus, in the absence of regression functions, we could sample with probability proportional to $D^{2.5}$ in pure stands, or with probability proportional to $gD^{2.5}$ in mixed stands, where g is an approximation of the (fresh or dry) weight per unit volume of the wood for a given species.

Second-Stage Sample

The PPS estimator furnishes an unbiased estimate of the total weight of the trees in the plot from measurements of the true weights of the trees in the PPS sample. If total fresh weight is the only parameter of interest, then the trees in the sample could be felled, weighed, and forgotten. Often, investigators are interested not only in total fresh weight but also in dry weight, mineral content, and volume. A subsampling procedure developed by Valentine et al. (1984) can be applied to the trees in the first-stage PPS sample to furnish

estimates of all of these parameters and their sampling errors. This procedure also eliminates the laborious task of weighing whole trees.

The subsampling procedure of Valentine et al. has two stages, the first stage employing randomized branch sampling (Jessen 1955), and the second stage employing importance sampling--a technique of Monte Carlo integration (e.g., Rubinstein 1981, Gregoire et al. 1985). The first and second stages of this subsampling procedure become, respectively, the second and third stages of the present procedure for estimating the total weight of the trees in a plot.

Randomized branch sampling is used to select a path that consists of connected branch segments (internodes), starting at the butt and ending at a terminal shoot of a tree. For operational purposes, a branch is defined as the entire stem system that develops from a single bud (lateral or terminal), whereas a segment is a part of branch between two consecutive nodes. We make no operational distinction between the main stem and side branches.

The first segment of any path extends from the butt (defined as the first node) to the second node. The second segment extends from the second node to a third node, and so on to the last (m th) segment of the path, a terminal shoot. The number of possible paths in a tree equals the number of terminal shoots.

Because the first segment of any path is fixed at the outset, its probability of selection (denoted by q_1) equals 1. To determine the remainder of our path we select one of the branches emanating from the second node at random with probability proportional to size, e.g., $d^{2.5}$. Selection of a branch fixes the second segment of the path. Its probability of selection is denoted by q_2 . The second segment is followed to a third node where a branch (and the third segment of the path) is selected with probability q_3 . This procedure is repeated until a terminal shoot is selected with probability q_m .

The probability q_j ($j=1,2,\dots,m$) is the conditional probability of selecting the j th segment of the path given that the path has reached the node to which the j th segment is attached. The unconditional probability of selection for the j th segment included in the path is:

$$Q_j = \prod_{r=1}^j q_r \quad (5)$$

Suppose that randomized branch sampling has been used to select a path on a tree numbered i in the first-stage sample. Denoting the weight of the j th segment in the path by b_j , an unbiased estimate of the weight of this tree is (Valentine et al. 1984):

$$\tilde{w}_i = \sum_j^m b_j Q_j^{-1} \quad (6)$$

This estimate could be substituted for w_i in (2) to furnish an unbiased estimate of the total weight of the trees in the plots.

Third-Stage Sample

The reciprocal of the unconditional probability of selection of the j th segment included in a path, Q_j^{-1} , is called an inflation factor. The product, $b_j Q_j^{-1}$, is called the inflated weight of the j th segment, and the sum of the inflated weights of the segments in a path--the unbiased estimator of tree weight--is called the inflated weight of the path. To measure the inflated weight of the path, we must separate and weigh the segments, some of which could be quite heavy. However, by employing the importance sampling procedure developed by Valentine et al. (1984), we can avoid weighing these segments. In fact, we can estimate the inflated weight of a path from the weight of a single disk. We also obtain a convenient specimen (the disk) for estimation of the inflated dry weight or mineral content of the path, which are, respectively, unbiased estimators of the dry weight or mineral content of the tree.

Importance sampling is a continuous analogue of PPS sampling and is used for estimating integrals. The inflated path weight, \tilde{w}_i , can be defined as the integral:

$$\tilde{w}_i = \int_0^\lambda Y(L) dL \quad 0 \leq L \leq \lambda \quad (7)$$

where $Y(L)$ is the inflated weight per unit length at point L along the path, and λ is the length of the path. (Any quantity measured at a point on the j th segment of the path is inflated by dividing the quantity by Q_j .)

By analogy to discrete PPS sampling, our task is to select a sample of points along the path independently and at random with probability as nearly proportional to inflated weight per unit length as possible. As was noted, tree weight is highly correlated with volume. Therefore, inflated weight per unit length, $Y(L)$, is highly correlated with inflated cross-sectional area $[A(L)]$, because cross-sectional area is volume per unit length. Although $A(L)$ is unknown, it is easy to build an *ad hoc* proxy function that accurately approximates $kA(L)$ (where k is an arbitrary constant).

For example, Valentine et al. (1984) described how to build a segmented-linear function, $S(L)$, that interpolates inflated d^2 measured at $L=0$, $L=\lambda$, and at arbitrary points in between. The resultant function,

$S(L)$, approximates $1.27A(L)$, is positive in the range $0 \leq L \leq \lambda$, and integrates to give

$$V(\lambda) = \int_0^\lambda S(L) dL \quad (8)$$

Selection of the z th of M points along the path at random with probability proportional to $S(\theta_z)$ involves finding the root, θ_z , of

$$V(\theta_z) - u_z V(\lambda) = 0 \quad 0 \leq \theta_z \leq \lambda; z=1, 2, \dots, M \quad (9)$$

where u_z is a uniform $[0,1]$ random number. An unbiased estimate of w_i based on a sample of M points is

$$\bar{w}_i = [V(\lambda)/M] \sum_{z=1}^M Y(\theta_z)/S(\theta_z) \quad (10)$$

Substitution of \bar{w}_i for w_i in (2) gives an unbiased estimate of total weight of trees in the plot.

In application we find that inflated weight per unit length can not be measured at a point, θ_z . Instead, we must measure weight per unit length of a disk centered at θ_z , thereby creating a trivial technical bias. However, it is possible to measure inflated cross-sectional area at θ_z , and substitution of this quantity for $Y(\theta_z)$ in (10) gives an unbiased estimate of the volume of the tree that can be substituted for w_i in (2) to give an unbiased estimate of the total aboveground woody volume of the trees in the plot.

The subsampling procedure involving randomized branch sampling and importance sampling is described in detail by Valentine et al. (1984). Results of a field test are reported by Valentine et al. (1984) and in somewhat greater detail by Tritton et al. (1983).

Discussion

Our procedure for estimation of total tree weight on a plot requires some on-site calculations, but very little office work. We have attempted to design a procedure whereby an investigator can enter a plot with a programmable calculator or portable computer and leave with an essentially unbiased estimate of fresh weight (or volume) and an estimate of sampling error, if desired. If the parameter of interest is dry weight or chemical constituency, then an investigator can leave the plot with specimens for drying or analysis and their respective inflation factors.

Invariably, complications arise in any forest sampling. In a randomized branch sampling of a broad-leaved tree, we may have to

deal with epicormic branches. If there are not too many, they can be removed and weighed separately. If there are a lot, we can ignore them while selecting the path by randomized branch sampling. An estimate of the weight of the epicormic branches on the tree can be calculated as the sum of the inflated weights of the epicormic branches on the segments of the path. Dead branches can be treated the same way.

In a randomized branch sampling of a coniferous tree, we may find a whorl containing a large number of small side branches at the j th node on the main stem. Consequently, assignment of selection probabilities based on diameter measurements becomes tedious. We can attempt to avoid measuring these small branches by performing the selection of the path segment in two stages. Let d_a and d_b denote, respectively, the main-stem diameter above and below the j th node. Based on Murray's (1927) observations, the sum of $d^{2.5}$ for the branches in the whorl plus $d^{2.5}$ should be approximately equal to $d_b^{2.5}$. If $d_a < d_b$ as expected, we can randomly select the main stem with probability $q_j = (d_a/d_b)^{2.5}$ or the side branches collectively with probability $1 - (d_a/d_b)^{2.5}$. The side branches need not be measured if the main stem is selected. If the collection of side branches is selected, we select one of the side branches at random with probability q'_j , whence $q'_j = q_j / [1 - (d_a/d_b)^{2.5}]$.

The techniques of importance sampling should be applicable for estimation of many quantities of interest to foresters and ecologists. Since its seminal use in a forestry problem by Valentine et al. (1984), importance sampling has been employed in procedures for estimation of bole volume (Gregoire et al. 1985, 1986), bole-volume increment (Valentine et al. 1986), and log volume (Furnival et al. 1986). Each of these procedures eliminates the bias that arises when a volume formula is used for estimation in a population other than that from which the formula derives.

Literature Cited

- Cochran, W. G. Sampling techniques. New York: John Wiley & Sons; 1963. 413 p.
- Furnival, T. G.; Valentine, H. T.; Gregoire, T. G. 1986. Estimation of log volume by importance sampling. For. Sci. 32(4): 1073-1078.
- Gregoire, T. G.; Valentine, H. T.; Furnival, G. M. Estimation of volume with a taper function: A sampling approach. In: Use of auxiliary information in natural resource inventories; 1985 October 1-2; Blacksburg, VA. Soc. Am. For. Publ. No. SAF86-01. 1985: 164-170.
- Gregoire, T. G.; Valentine, H. T.; Furnival, G. M. 1986. Estimation of bole volume by importance sampling. Can. J. For. Res. 16: 554-557.
- Jessen, R. J. Determining the fruit count on a tree by randomized branch sampling. Biometrics II: 99-109; 1955.
- Murray, C. D. A relationship between circumference and weight in trees and its bearing on branching angles. J. Gen. Physiol. 10: 725-729; 1927.
- Rubinstein, R. Y. Simulation and the Monte Carlo method. New York: John Wiley & Sons; 1981. 278 p.
- Sukhatme, P. V.; Sukhatme, B. V. Sampling theory surveys with applications. 2nd rev. ed. Ames, IA: Iowa State University Press; 1970. 452 p.
- Tritton, L. M.; Valentine, H. T.; Furnival, G. M. A new procedure for estimation of tree biomass and nutrient content. In: Mesures des biomasses et des accroissements forestiers; 1983 October 3-7; Orleans, France. Les Colloques de l'INRA, No. 19; 1983: 335-341.
- Valentine, H. T.; Tritton, L. M.; Furnival, G. M. Subsampling trees for biomass, volume, or mineral content. For. Sci. 30: 673-681; 1984.
- Valentine, H. T.; Gregoire, T. G.; Furnival, G. M.; Solomon, D. S. Unbiased estimation of bole increment. In: Environmental influences on tree and stand increment. IUFRO (S4.01-06); 1985 September 23-27; Durham, NH. [in press].
- White, J. The allometric interpretation of the self-thinning rule. J. Theor. Biol. 89: 475-500; 1981.

MEASUREMENT ERRORS IN FOREST

BIOMASS ESTIMATION

Daniel AUCLAIR

INRA, Centre de Recherches d'Orléans,
Station de Sylviculture
ARDON - 45160 OLIVET (France)

by estimating the fresh weight of the whole tree, most often separated in several components, and collecting samples for water content.

However in some cases we have also estimated biomass of larger trees, from traditional high forest or coppice with standards, where the stem volume was estimated geometrically, and samples were collected to convert volume into dry weight.

This paper concerns estimations of measurement errors on :

When sampling for biomass regressions, one source of error can come from the error made on the independent variables. The order of magnitude of this error has been estimated on several variables : girth at breast height, tree height, relative water content of samples, volume of samples, ... Their contribution to the error on the estimation is discussed.

- girth at breast height
- total tree height
- water content of samples
- weight of samples
- volume of samples
- the use of basic density for biomass estimation.

The data comes either from original mensurations collected especially for this purpose, or from other studies by AUCLAIR and METAYER (1980) or by BISCH (1985).

Introduction

For estimating biomass of sample trees, to be subsequently applied to regressions in order to determine prediction equations, many procedures have been developed by various forest scientists. These have been well described by YOUNG (1976), PARDE (1980), SATOO and MADGWICK (1982), and one of the annual meetings of IUFRO's S4.01-00 subject group on "Mensuration, Growth and Yield" was devoted to "Problems in forest biomass mensuration and growth and yield studies" (AUCLAIR, 1983).

1 - Girth and Tree Height

a - Method : On twenty Chestnut trees (*Castanea sativa* Mill.) situated in a traditional coppice stand of the INRA estate, aged from four to thirty five years, the girth at breast height was measured, using a metal tape graduated in millimetres. Fifteen different persons, forest research professionals, technical staff, and also some non-professionals, measured the same trees, giving the girth with an accuracy of 1 cm.

Most methods use diameter (or girth) at breast height, or tree height, or a combination of both, as independent variables for biomass regressions. The weight of sample trees can be estimated by several methods, including either a conversion factor from fresh weight to dry weight or a conversion factor from volume to dry weight. This estimation can be done on the whole tree, or if it is too big on subsamples (branches) which may be used once more for establishing regressions for individual tree biomass estimation.

On the same trees the total tree height was estimated :

- 1) with a graduated rod on trees smaller than 8 m., with an accuracy of 1 cm.
- 2) with a Blume-Leiss hypsometer for trees 8 to 15 m high, with a 10 cm accuracy. On these trees a possible bias was estimated by one single measurement with the graduated rod.

At each of the steps in the process of determining biomass regression equations, some measurement errors occur. These may be random errors, often assumed to be normal, or systematic errors.

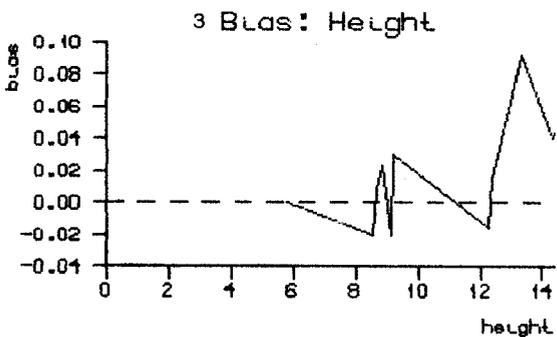
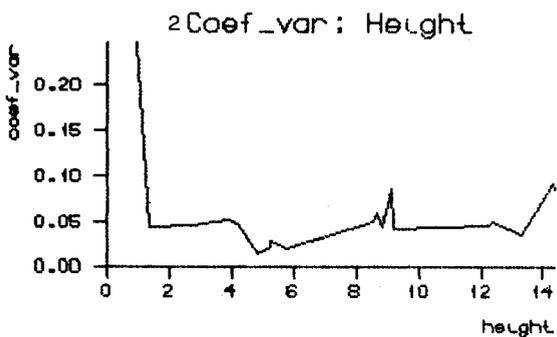
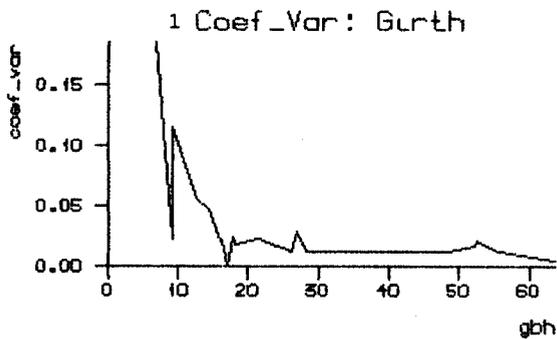
To verify whether or not there was any large difference between operators, we used the non-parametric test of Kendall's coefficient of concordance. In all cases it showed no difference, at a very high degree of significance, both for girth and height measurements.

In order to obtain the best possible estimations, the present study was undertaken to determine the various contributions of measurement errors in the procedures used in our laboratory for estimating biomass. The practical aim was to minimize the most important errors, at a "reasonable" cost.

b - Results : The coefficients of variation and the bias for height measurements are shown in Figures 1 to 3, plotted against the mean of either girth-at-breast-height, or total height.

Most of our studies concern trees treated as coppice : traditional coppice with 20- to 30-year rotations, or short-rotation coppice going down to annual rotations. These are therefore small trees, whose biomass can be determined

The two high values of the coefficient of variation for smaller girths (Fig. 1) can be explained by the data collected by one operator who overestimated the girth by 2 cm compared to the mean. These figures however still remain around 0.10 if this operator is excluded, but this corresponds to absolute differences of ± 1 cm.



Figures 1 to 3 : coefficient of variation of girth or height plotted against girth (cm) or height (m), and percentage of bias for tree height.

The very high value for the smallest tree in Figure 2 comes from an error in transcription of the data where a tree 1.00 m high was noted 0.30 m. If this data is excluded the coefficient of variation becomes 0.09, which corresponds to an absolute difference of ± 15 cm.

In Figure 1, after 20 cm the coefficient of variation for girth smooths out at a value under 0.02, whereas that for height (Fig. 2) seems to increase slightly with increasing height, but remains around 0.05.

The bias estimated on height measurements (Fig. 3) seems to increase with height, but it is much more correlated with the fact that the crown of a tree is more or less visible : a "visibility index" was applied to each of the measured trees, and the coefficient of variation and bias for height were compared to this index :

Visibility	Coef-var (%)	bias (%)
good	5.7	- 0.3
medium	3.6	+ 0.7
poor	7.2	+ 5.4

The height of poorly visible trees seems to be overestimated, but no conclusion can be reached about the coefficient of variation.

c - Conclusions : The variation in girth and height measurements is quite high for very small trees, however their contribution to the total biomass of a stand is usually rather small. The variation for girth levels out quickly for larger trees, but variations in height measurement remain at quite a high value, and may increase for the bigger trees.

Height measurements are very time-consuming, and it would be rather costly to try to increase precision by improving the accuracy of these measurements for all the trees. We therefore suggest to focus on accurate "top height" measurements and use the "top height" of a stand as an independent variable for biomass regressions. This improves the precision compared to regressions with girth or diameter as the only independent variable (PAGES, 1985).

The bias which was estimated here may seem unimportant, as the same bias should be obtained when establishing the regression and when using it for prediction. However the sampling should take into account the fact that the tree crown is more or less visible, as a "poor" visibility seems to provoke an overestimation.

2 - Water content of samples

a - Variation within a tree : In a previous report, AUCLAIR and METAYER (1980) studied the water content of three series of samples on trees aged 25 to 40 (Birch, Oak, and Hornbeam) divided in five diameter classes : 0-2.5 cm, 2.5-4 cm, 4-7 cm, 7-15 cm, + 15 cm. The results (Fig. 4) show that the water content increases in the smaller components.

Several authors have tried to use only one single mean sample for estimating total dry biomass, but as the proportion of different components varies between species, or between trees (Fig. 5), it seems most important to apply the appropriate conversion factor to each of the components.

Water content

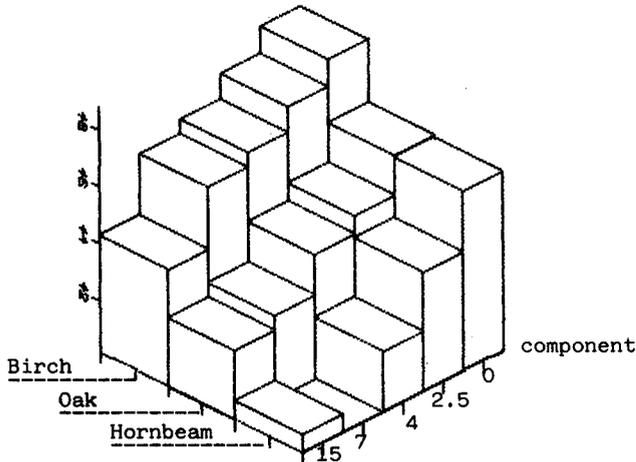


Figure 4 : Water content (percent fresh weight) for four components in Birch, Oak, and Hornbeam.

Biomass distribution : Birch

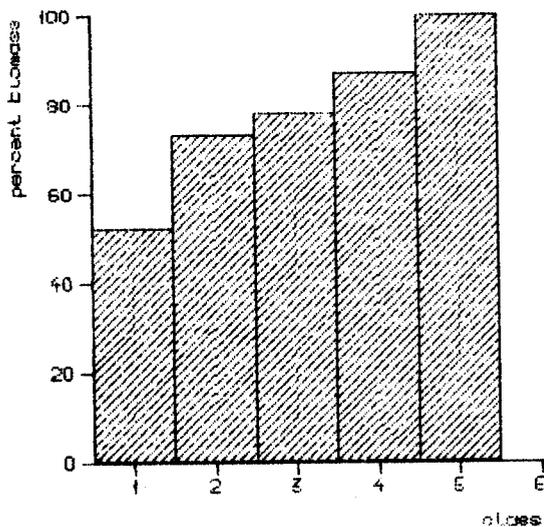


Figure 5 : Percentage of biomass in each class for 25-year Birch coppice :

- 1 : > 7 cm
- 2 : "stacked wood"
- 3 : > 4 cm
- 4 : > 2.5 cm
- 5 : total

b - Drying temperature : For drying samples to estimate water content, the usual procedure is to leave them in an oven until their apparent weight becomes constant. This usually lasts two to five days.

However, some authors suggest using an oven temperature of 60 or 65 degrees C., others 85 de-

grees, and others 105 degrees. The lowest temperature is used mostly for drying leaves, or when nutrients are analyzed ; at the highest temperature all the water evaporates but some volatile compounds may also disappear.

On a series of samples, divided in four diameter classes, we tested the relative difference in apparent dry weight after drying at 65, 85, or 105 degrees :

Diameter	$\frac{W_{65}-W_{85}}{W_{65}}$ %	$\frac{W_{85}-W_{105}}{W_{85}}$ %
0-2.5 cm	3.9	1.3
2.5-4 cm	4.5	2.4
4-7 cm	4.7	6.4
+ 7 cm	3	8

In all cases the samples still loose some weight between 85° and 105°. The smaller diameter classes (< 4 cm) loose less than 2.5 percent, but the larger ones loose between 5 and 10 percent.

It therefore seems quite difficult to give a single conversion factor from the dry weight at one oven temperature to the dry weight at another, for the same reason as above, as the distribution of diameter classes inside each tree is most variable.

c - Water loss of samples : In the process of biomass estimation, it is often impossible to estimate the fresh weight of samples immediately at the same time as the whole tree or compartment biomass estimation. The type of balance used is often not sufficiently precise for weighing small samples. If the fresh weight is to be estimated at the laboratory, we often keep the samples in hermetically closed plastic bags.

To estimate a possible error due to this practice, we measured the fresh weight of samples in the field, and after keeping them in plastic bags for 1, 2, 3, or 10 days. Figure 6 shows the distribution of relative differences at the four sampling dates.

After one day, most samples lose 1 or 2 percent of their fresh weight, but after two or three days the weight loss can reach 6 percent. After ten days the distribution of error smooths out, the weight loss varies from 1 to 8 percent, and in one case 10 percent.

As in the study concerning different drying temperatures, it is not possible here to give a single correction factor due to the delay in weighing samples. However, if the samples are weighed inside the plastic bag still hermetically closed, their weight loss is compensated by water kept inside the bag, and the error is negligible.

number of samples

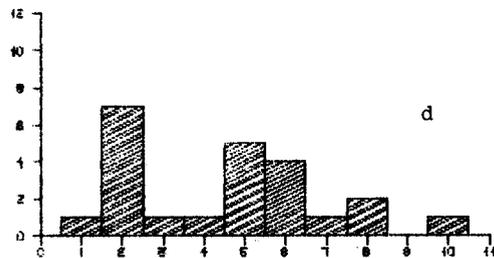
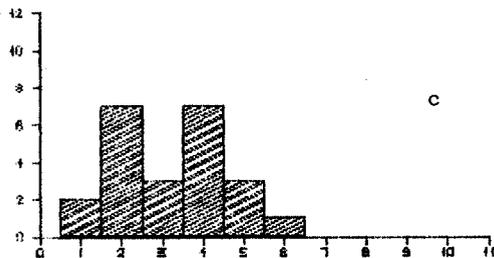
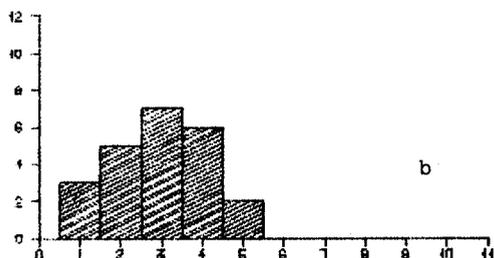
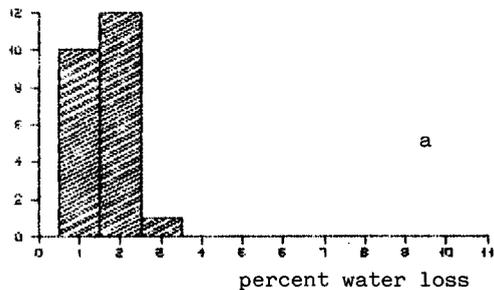


Figure 6 : Histogram of water loss of samples left in plastic bag for :

- a : 1 day
- b : 2 days
- c : 3 days
- d : 10 days

3 - Weight of samples

For biomass estimation in the field, we usually weigh tree compartments on a very precise balance, whose scale is graduated every 20 grams, up to 60 kg.

Although this balance is sent every year to have a complete check-up, we have verified in various occasions the values indicated. One exam-

ple is given in figure 7. It shows the absolute value of relative error, plotted against total weight. All through the scale there is a slight underestimation. For lower values the underestimation is 40 g, and for higher values it can reach 80 g. This leads to quite a small relative error, less than one percent, but it is however possible to give a correction factor for this systematic bias if such curves as that in figure 7 are regularly verified.

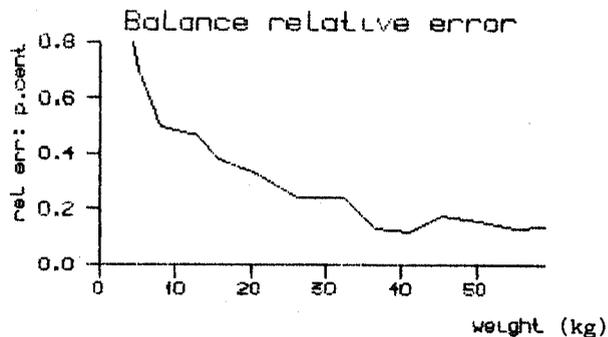


Figure 7 : Relative error (x 100) of balance used for biomass estimation.

4 - Volume of samples

For estimating biomass of very large trees, a method often applied is to estimate the bole volume, and use a conversion factor estimated from a sample, or several samples taken from various levels on the bole.

Such a study was undertaken by BISCH (1986), who showed that there is often a bias in the estimation of the volume of disks, due to the fact that the Archimedean principle is often incorrectly used. The fact that the quantity of water absorbed by a sample during its immersion is not taken into account, produces a systematic underestimation of the volume. On oak disks he showed that the mean bias was above 1 percent, and can reach 2 percent. This seems highly correlated with the bark volume of the samples.

However, as the bole volume of a tree is estimated geometrically, it seems more important to also estimate geometrically the volume of samples. This was done by measuring the disk girth and the disk width as a mean of four different measurements, with a metal tape.

The geometric volume gives an overestimation compared to "real volume" estimated by weight of water displacement. The relative bias is 5 percent \pm 0.4 percent, estimated on 74 disks from two different oak species (*Q. robur* and *Q. petraea*). The author noticed no difference between species, nor between silvicultural treatments.

5 - The use of basic density

In his study of biomass distribution, BISCH (1985) estimated bole volume with two different methods, and estimated wood basic density on four

disks from the base of the tree, 4 m, 8 m, and 12 m upwards.

Volume 1 is estimated with the traditional method of the french national forest survey, dividing the bole in two parts : under, and above 2.6 m, and measuring the girth at the middle of each of these sections.

Volume 2 is estimated by measuring diameters every meter along the bole.

Four different total stem biomass estimations were made :

B1 : volume 1 and a mean of the basic densities of the four samples ;

B2 : volume 2 and a mean basic density ;

B3 : volume 1, and basic density from the base of the tree being applied to the first section, a mean of the others to the above section ;

B4 : volume 2, divided in four compartments, 0-1 m, 1-6 m, 6-10 m, > 10 m.

The mean differences are the following :

$$\frac{B1 - B4}{B4} \times 100 = - 2.13$$

$$\frac{B2 - B4}{B4} \times 100 = + 0.59$$

$$\frac{B3 - B4}{B4} \times 100 = - 2.53$$

The use of the "traditional" forest survey volume underestimates the total stem biomass. The basic density varies up the stem (Fig. 8), and using the mean of four disks seems to give a reasonably good estimation. The use of one single sample for basic density estimation seems insufficient.

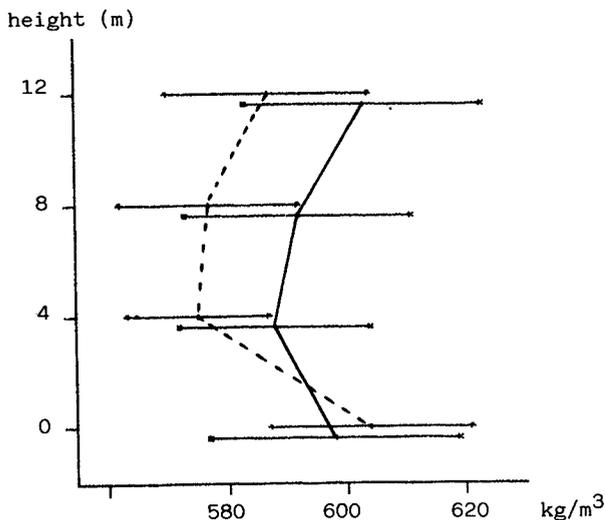


Figure 8 : Mean basic density of samples of Oak at different levels in the stem. Horizontal bars indicate confidence intervals at 5 % :
 — standards from coppice with standards
 ---- trees from high forest

Conclusions

In our studies on biomass estimation, we have attempted to minimize measurement errors, and especially bias. Several phases in the process of biomass estimation have been studied here, and suggestions can be given to improve the accuracy of measurements. However, the quality/cost ratio may in some cases become very low, and some errors are difficult to minimize.

The same type of study has already been undertaken for volume estimation (SCHMID-HAAS et al., 1980 ; WINZELER, 1986) suggesting the best type of material to use or measurements to be done. The present study brings a contribution to measurement errors on biomass estimation.

The effect of various types of error on the results of regression utilization has been studied by GERTNER and DZIALOWY (1984) and GERTNER (1985), who showed that for example a 10 % bias in DBH may cause a 25 % error in predicted basal area.

This type of bias is quite in the same order of magnitude as the data presented here. It must however be outlined that the present study concerns mostly traditional coppice, with very bad stem form, most often on very old root systems. In other cases, for example conifers who have much straighter stems, or with more homogeneous plant material, as genetically controlled plantations, the errors may be reduced.

Literature cited

- AUCLAIR, Daniel (Ed.). Problems in forest biomass mensuration and growth and yield studies. Proceedings, IUFRO S4.01-00 ; 1983 October 3-7 ; Orléans (FRA). Les colloques de l'INRA, 19 ; I.N.R.A., Paris ; 1983. 356 p.
- AUCLAIR, Daniel ; METAYER, Sylvie. Méthodologie de l'évaluation de la biomasse aérienne sur pied et de la production en biomasse des taillis. Acta oecol., Oecol. appl. 1 (4) : 357-377 ; 1980.
- BISCH, Jean-Luc. Influence du traitement sylvicole (futaie et taillis sous futaie) sur la répartition de la biomasse dans le Chêne en région Centre. INRA, Station de Sylviculture, Orléans (FRA) ; 85/42 ; 1985. 26 p.
- BISCH, Jean-Luc. Le taux de conversion de volumes de bois frais en biomasse : Amélioration de ses méthodes d'estimation. Ann. Sci. for. 43 (3) : in press ; 1986.
- GERTNER, George Z. Regressor variable errors and the estimation and prediction with linear and nonlinear models. October 1985. 15 p.
- GERTNER, George Z. ; DZIALOWY, Paul J. Effects of measurement errors on an individual tree-based growth projection system. Can J. For. Res. 14 : 311-316 ; 1984.

- PAGES, Loïc. Le taillis de Robinier du Val de Loire - croissance, biomasse, régénération. Université de Paris-Sud (Orsay) ; thèse (Dr. es Sci.) ; 1985. 74 p.
- PARDE, Jean. Forest biomass. For. Abstr. 41 (8) : 343-362 ; 1980.
- SATOO, T. ; MADGWICK, H. A. I. Forest biomass. The Hague : Martinus Nijhoff/Dr. W. Junk ; 1982. 152 p.
- SCHMID-HAAS, P. ; WERNER, J. ; BAUMANN, E. Fehler bei der Rundholzmessung. Schweiz. Z. Forstwes. 131 (9) : 801-820 ; 1980.
- WINZELER, Klemens. Volumenbestimmung durch Messungen am stehenden Baum. Eidg. Anst. forstl. Versuchswes., Mitt. 62 (1) ; 1986. 99 p.
- YOUNG, Harold E. (Ed.). Oslo biomass studies. Proceedings, IUFRO S4.01-00 ; 1976 June 22 ; Oslo. College of life sciences and agriculture, University of Maine at Orono, ME ; 1976. 302 p.

RESEARCH PAPERS

Biomass of Forest Understory Vegetation

Moderator: Delbert R. Mead

BIOMASS-DIMENSION RELATIONSHIPS OF
UNDERSTORY VEGETATION IN RELATION TO SITE
AND STAND AGE.

Paul B. Alaback

Research Ecologist, USDA Forest Service,
Pacific Northwest Research Station,
Forestry Sciences Laboratory, P.O. Box 909,
Juneau, AK 99802.

Three woody shrubs, two ferns, and three herb species in a wide range of stand types in southeast Alaska and the Pacific Northwest were examined. Biomass regression equations were developed for each of the species in recent clearcuts, mature second growth forests, old growth, and, for one species, in three thinning treatments. All but one of the fern equations were significantly different between age classes ($P < .05$). Herb biomass per unit cover was greatest in clearcuts and least in second growth and old growth. Woody biomass per stem size-class of shrubs was greatest in clearcuts and least in old growth. Leaf and twig production had the opposite relation. Differences between shrub biomass equations were generally related to shrub density and openness of the overstory canopy; changes in herb biomass were mostly related to openness of overstory canopy. Thinning experiments confirmed these basic patterns of change in biomass-dimension relations with overstory density. Use of equations from other regions gave significant errors in estimates of biomass compared with estimates made from equations developed in southeast Alaska. Equations developed for the same species in geographic regions with similar overstory or climate gave no more accurate estimates of biomass for southeast Alaska than equations developed in distant regions or from regions with contrasting climates. Developing equations for local populations or for specific stand structures is recommended for forest understory species.

Introduction

Understory biomass has been used to contrast plant community habitat types, to distinguish subtleties in forest microclimate, and to indicate quality of wildlife habitat. For more than two decades, regression equations have been used to predict understory biomass, principally for production, nutrient cycling, and studies of wildlife habitat (Alaback 1982, Ohmann and Grigal 1985, Tappeiner and John 1973, Whittaker 1961, Yarie 1980). Almost all of the equations developed were based on sampling from one study area or from several study plots

within a general study area. Most authors stressed the requirement that their equations be used on similar sites, but equations are often used in sites significantly different from the original sampling areas because of poor documentation of the original site conditions and locations.

Improper use of biomass equations developed from dimension analysis has been furthered by compilations of libraries of regression equations, usually with little information on how they were developed, what their limitations are, how they might be best applied, and what magnitude of error or bias is likely if they are applied to different sites (Gholz et al. 1979, Smith and Brand 1983, Stanek and State 1978). Some users of these libraries suggest an average of coefficients be used where more than one equation is available for a species, which may compound problems of potential error and bias (Ohmann and Grigal 1985, Pastor et al. 1984).

In this study, equations are presented for understory species growing in different sites in southeast Alaska and in comparable forest types in British Columbia, Montana, Minnesota, and Oregon. Objectives were: (1) to estimate the magnitude of bias in biomass estimates arising from using equations developed from different stand age-classes and overstory structures in southeast Alaska (2) to document the general pattern of variation in biomass-dimension relations for a given species across a gradient of overstory density and understory productivity (3) to examine the bias in biomass estimates made from equations developed in other geographic regions for a given species.

Methods

Sites with relatively homogeneous overstory structure, gently sloping terrain, and well-drained to relatively well-drained soils were selected in southeast Alaska (Alaback 1982, 1986). In each site, samples were randomly selected by plant size-class so that samples were more or less equally distributed across the full range of plant-size classes. Sampling was originally done during the summers of 1977, 1978 and 1979 coincident with a large-scale study of postlogging succession (Alaback 1982). Sampling was done between July 1 and September 1 of each year when all species had attained peak biomass (Alaback 1986). In the summers of 1982 to 1984 additional sampling was done to develop equations for new species, and to provide for sufficient samples in each of three stand-age classes to allow for direct comparison of equations. In addition, one site that had been precommercially thinned 4 years before biomass sampling was

studied. Biomass samples were taken from three of the thinning treatments (control, 8-foot spacing, and 16-foot spacing (Alaback 1984)).

Non-destructive measurements were taken on individual plants before destructive sampling. Samples were separated into components in the field, then oven-dried at 70 °C for 24-48 hours. Samples were weighed to the nearest 0.001 gram. Independent variables that were easy to measure in the field and were thought to bear a precise relation to biomass were selected: for shrubs, basal diameter; for herbs, percent ground cover on 0.1 m² plots; for ferns, frond length was used. Both ln-ln transformed and untransformed linear models were applied to the data for each species and biomass component. Models with the highest correlation coefficient and the least mean square error were selected as the best fit. Correction for bias was used for the ln-ln transformed data (Baskerville 1972).

Equations were compared by use of successive analyses of covariance with dummy variables (Cunia 1973). For each species, a regression model that included slopes and intercepts for all sites and treatments was used. All possible equations for a species were tested for differences in slopes and intercepts. If differences were significant then all possible pairs of equations were compared to determine which were significantly different. Pairs of equations with calculated F statistics greater than F at P = 0.05 were judged to be significantly different. Equations between regions for which original data were not always available were compared by calculating biomass estimates at an arbitrary value of the independent variable, usually at the upper limit of sampling where the greatest differences occurred.

Nomenclature for trees, shrubs, and herbs follows Hulten (1968).

Results and discussion

Herbs

Equations for specific age-classes or stand types tended to have coefficients of determination and mean squared error within the range of or better than regional biomass equations (Alaback 1986, Table 1). Herb cover equations were the most precise; shrub foliage and twig production were generally least precise. The independent variable for fern equations was frond length rather than percent ground cover because of the large variation in frond orientation with frond size or plant age. For both Athyrium and Dryopteris,

regression coefficients for open clearcuts and old growth were similar and are probably consistent over a wide geographic range. The primitive growth form and the lack of variation in branching or canopy structure probably contribute to the relative site independence of these equations.

For each of the three herb species, equations based on percent cover gave similar magnitudes and direction of bias for stand age-class (Figures 1-3). In recent clearcuts, herbs are generally taller and have denser foliage. In closed forests where light and temperature may be limiting factors, herb biomass is as little as 20-25 percent of biomass per unit cover in clearcuts at maximum cover values. Plants in closed forests are generally shorter with sparser canopies. The biomass equations also predict that plants growing in second-growth forests have slightly less biomass per unit cover than plants in old growth (P < 0.05, Table 2), consistent with the hypothesis of light limitation.

Shrubs

All three shrub species showed a similar pattern of change in biomass-dimension relations across the three age-classes (Table 1). In clearcuts, woody biomass was much greater per unit stem diameter than in the closed forest types (Figure 4). Leaf and twig production per unit stem diameter was much less than in the old-growth sites (Figure 5). Although the largest differences were between equations for recent clearcuts and closed forest types, significant differences were detected between biomass equations for old- and second-growth sites (Table 2).

In general, wood production per unit stem diameter was proportional to solar radiation, plant vigor, and shrub density, but leaf and twig production per unit stem diameter responded inversely to shrub density and incident radiation. Shrubs growing in open areas with plenty of light thus put more energy into height growth and stem development than shrubs growing in closed forests, which presumably had less competition with other understory plants and more need for expanded leaf area and canopy development to capture limited incident solar radiation (Gholz 1978). Equations from second-growth stands were probably influenced by the low vigor of residual plants, slowly being shaded out by overstory species (Alaback 1982). Many of these plants were tall with sparse canopies and were restricted to growing in small gaps in the overstory canopy. The dramatic change in absolute biomass of each component and proportional allocation between components argues strongly for reporting shrub biomass by component rather than simply as total aboveground biomass, which confounds these differences.

Table 1. Biomass equations for species in southeast Alaska growing in different stand types or age classes. All equations are significantly different at $P < .05$ using the techniques of Cunia (1973).

Species	a	b	r ²	mse	n	model**	range
HERBS TOTAL BIOMASS							
<u>Athyrium felix-femina:</u> *							
Clearcut	-3.226	0.047	.967	0.061	18	LL	18-105
Second growth	-2.568	0.034	.902	0.200	15	LL	32-147
<u>Cornus canadensis:</u>							
Clearcut	-2.665	1.278	.951	0.188	8	LC	2-100
Second growth (1)	-3.694	1.183	.834	0.205	17	C	3-60
Second growth (2)	0.020	0.050	.820	0.145	17	C	3-60
<u>Coptis aspleniifolia:</u>							
Clearcut	-2.567	0.293	.731	31.202	11	C	2-95
Second growth	-0.060	0.049	.955	0.009	8	C	2-25
Old growth	0.054	0.032	.972	0.006	7	C	5-40
<u>Dryopteris austriaca:</u>							
Clearcut	-9.036	2.286	.766	0.545	25	LL	9-95
Second growth	-8.841	2.178	.942	0.182	109	LL	4-95
<u>Rubus pedatus:</u>							
Clearcut	-0.578	0.100	.963	0.182	11	C	5-68
Second growth	0.103	0.028	.939	0.038	13	C	1-90
Old growth	0.244	0.011	.674	0.033	10	C	1-60
SHRUBS TWIG AND LEAF BIOMASS							
<u>Oplopanax horridum:</u> ***							
Clearcut	-6.314	3.684	.969	0.121	5	LD	4-27
Second growth	-3.205	2.128	.639	0.639	11	LD	5-30
Old growth	-5.488	3.579	.957	0.116	10	LD	5-20
<u>Rubus spectabilis:</u>							
Control	-3.754	2.842	.791	0.466	20	LD	2-15
Light thinning	-2.141	2.122	.690	0.322	11	LD	4-15
Heavy thinning	-1.114	1.923	.450	0.602	19	LD	6-25
<u>Vaccinium alaskaense:</u>							
Clearcut	-2.684	1.820	.660	0.816	82	LD	1-26
Second growth	-2.961	1.841	.810	0.433	75	LD	1-22
Old growth	-4.301	2.803	.730	1.262	7	LD	4-23
SHRUBS WOODY BIOMASS							
<u>R. spectabilis:</u>							
Control	-6.710	4.523	.914	0.295	10	LD	2-15
Light thinning	-2.680	2.636	.919	0.102	10	LD	4-15
Heavy thinning	-3.632	2.930	.733	0.402	20	LD	6-25
<u>V. alaskaense:</u>							
Clearcut	-2.247	2.553	.519	2.132	52	LD	1-26
Second growth	-1.612	1.701	.561	1.194	40	LD	1-22
Old growth	-3.231	3.210	.793	1.164	7	LD	4-23

* clearcut = 0-25 years after logging, second-growth = 26-200 years, old growth > 200 years.

** LL: $Y = \exp(a+b*\ln(\text{length,cm}))$
 C: $Y \text{ (g/1000 cm}^2\text{)} = a+b*(\text{percent cover})$
 LC: $Y = \exp(a+b*\ln(\text{percent cover}))$
 LD: $Y = \exp(a+b*\ln(\text{basal stem diameter,mm}))$

*** total above ground biomass

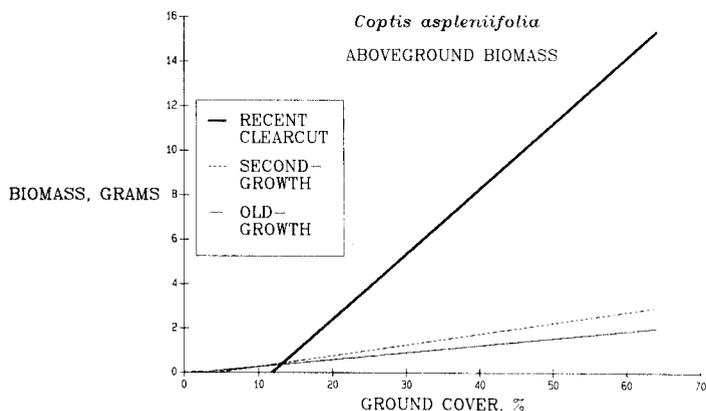


Figure 1. Biomass equations for Coptis asplenifolia in relation to stand age.

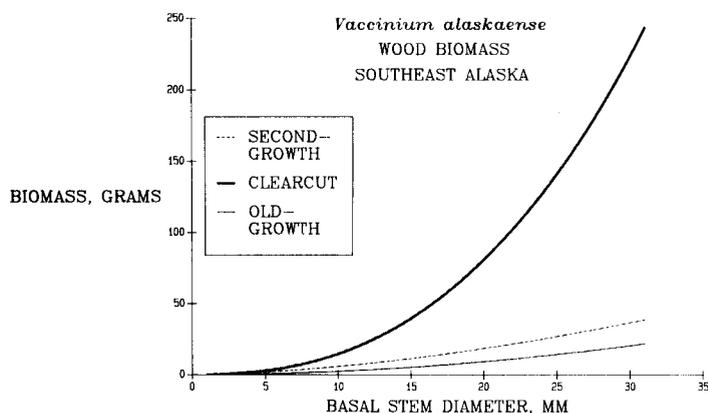


Figure 4. Wood biomass equations for Vaccinium alaskaense in relation to stand age.

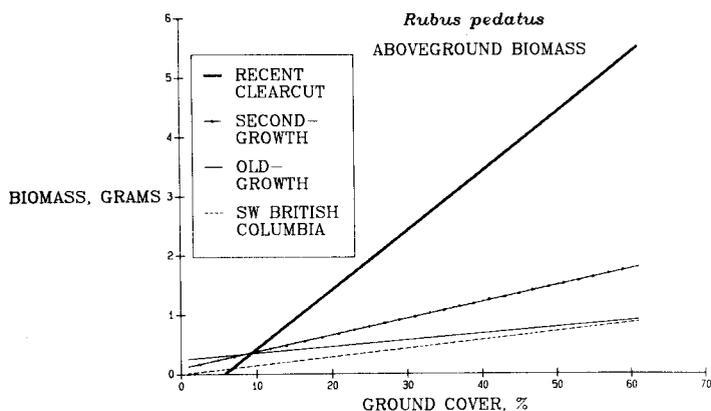


Figure 2. Biomass equations for Rubus pedatus in relation to stand age.

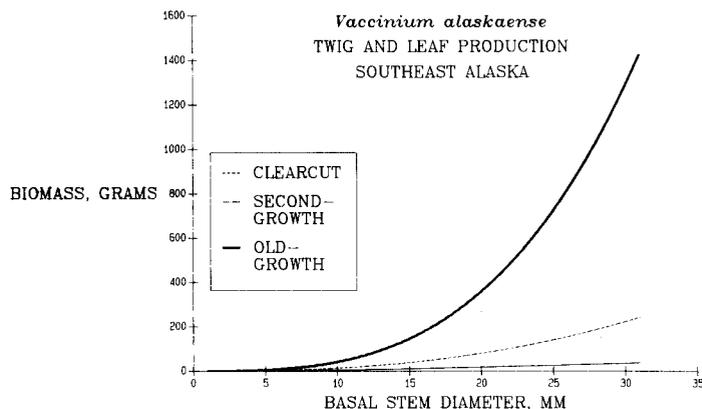


Figure 5. Twig and leaf production equations for Vaccinium alaskaense in relation to stand age.

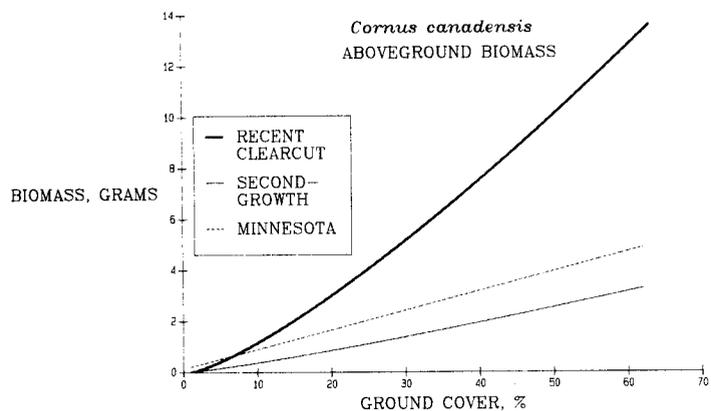


Figure 3. Biomass equations for Cornus canadensis in relation to stand age.

Effects of thinning

Results from experimental manipulations of overstory canopy density generally agreed with results from site age-class comparisons. Thinning resulted in increased vigor of Rubus spectabilis reflected in the equations for woody biomass per unit stem diameter (Table 1, Figure 6). Twig and foliage production decreased per unit stem diameter in response to thinning (Figure 7). The pattern, therefore, generally agreed with that observed for biomass equations from other shrub species across the gradient of overstory canopy openness and shrub density. The largest difference was between the unthinned (control) and the thinned plots equations.

Table 2. Comparisons of biomass equations using F statistics from analyses of covariance for southeast Alaska sites.

Comparison	Species	F statistics ***	
		slopes	intercepts
Clearcut vs second growth vs old growth:			
	<u>C. aspleniifolia</u>	**7.304	-0.950
	<u>C. canadensis</u>	*4.780	** -1.655
	<u>O. horrida</u>	*-4.530	** -8.220
	<u>R. pedatus</u>	*4.030	0.655
	<u>V. alaskaense</u>	**11.550	**23.160
Second growth vs old growth:			
	<u>C. aspleniifolia</u>	*-4.770	** -16.730
	<u>O. horrida</u>	2.040	** -18.700
	<u>R. pedatus</u>	** -9.990	** 1.170
	<u>V. alaskaense</u>	** -12.770	** -55.980
Clearcut vs second growth:			
	<u>A. felix-femina</u>	2.110	*2.620
	<u>C. aspleniifolia</u>	2.330	*5.530
	<u>C. canadensis</u>	*8.010	**92.360
	<u>D. austriaca</u>	**119.100	**121.370
	<u>O. horrida</u>	** -18.416	** -20.819
	<u>V. alaskaense</u>	**58.366	** -82.153
Control vs light thinning vs heavy thinning:			
	<u>R. spectabilis</u>	** -32.500	** -13.800
Control vs light thinning:			
	<u>R. spectabilis</u>	** -37.100	** -39.100
Light vs heavy thinning:			
	<u>R. spectabilis</u>	** -58.600	** -30.700

* significant at P < 0.95

** significant at P < 0.99

*** F values significantly greater than the critical value indicate in column 1 that slopes are significantly different or in column 2 intercepts are significantly different.

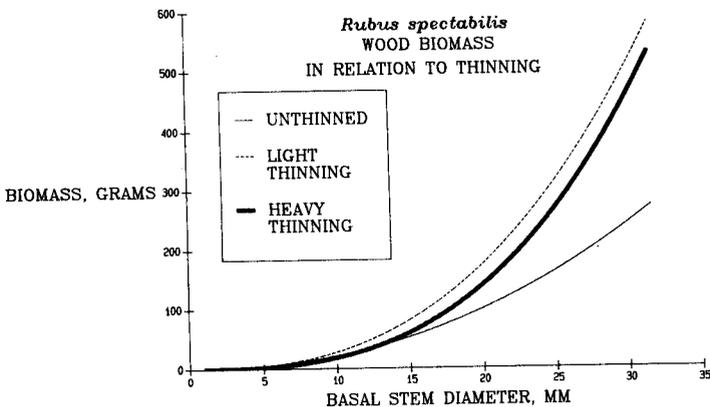


Figure 6. Woody biomass equations for Rubus spectabilis in relation to thinning.

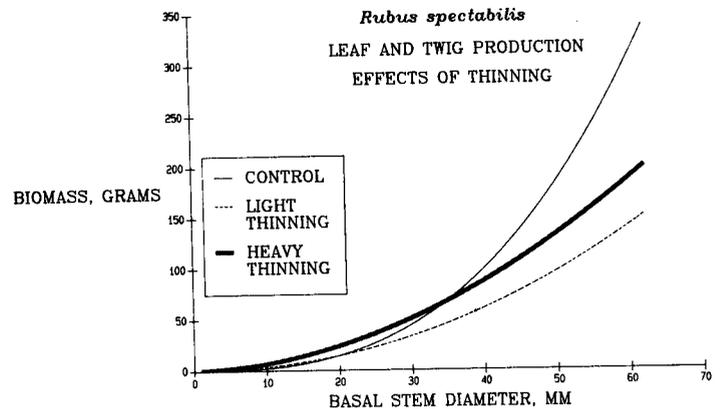


Figure 7. Leaf and twig production equations for Rubus spectabilis in relation to thinning

If woody biomass accumulation per unit stem diameter were proportional to shrub density and light intensity, the heavy thinning rather than the light thinning would be expected to yield the highest biomass estimates. The fact that both light and heavy thinning produced generally similar equations compared with the control (Figure 7) suggests that the effect of thinning on biomass dimension relations may not operate over the restricted range of light and shrub-density values tested, or that other site factors may be involved. For example, although the heavy thinning resulted in 14 times the biomass of the control, and the light thinning resulted in only 1.5 times the biomass of the control for a stem of 25 mm, R. spectabilis growing on the lightly thinned site had the highest proportion of woody biomass per unit stem diameter (Alaback 1980). On the lightly thinned site the second most abundant species V. alaskaense was 17 times more productive than on the heavily thinned site, perhaps offering less competition for canopy development of R. spectabilis because of its more open shade-tolerant canopy than the dense shade-intolerant canopy of R. spectabilis in the heavily thinned site (Alaback 1980). Thus, the balance between woody and canopy components of biomass in shrubs could be a function of light, soils, and competitive interactions with other species.

Regional comparisons

In contrast to many overstory species, most forest understory species are relatively restricted in geographic range so that comparisons between regions are usually difficult to make. Biomass equations were found for some of the species studied in low-elevation Tsuga

heterophylla forests in Alaska in Tsuga mertensiana forests in southwest British Columbia (Yarie 1980), in Picea sitchensis-Tsuga heterophylla forests in coastal Oregon and Pseudotsuga menziesii forests in the Western Cascades (Gholz et al. 1979) mixed conifer forests in northern Idaho and western Montana (Brown 1976), subalpine conifer forests in Glacier National Park, Montana (Gholz et al. 1979) and mixed hardwood-conifer forests in northern Minnesota (Ohmann et al. 1981). Although original data from these other studies were not available for direct statistical comparisons, most biomass estimates from these equations had biases at least as large between regions as was found in this study between age-classes within southeast Alaska, and were therefore assumed to be significantly different.

Only in southwest British Columbia and northern Minnesota were biomass equations developed for herbaceous species found in southeast Alaska. The equations for British Columbia were pooled from three habitat types, which were not significantly different ($P < 0.05$) (Yarie 1980). The data from Minnesota were also based on several habitat types. Only the biomass predictions for the fern Gymnocarpium dryopteris were within 10% of the predictions that used the southeast Alaska equations at a cover value of 50% (Table 3). Although the sites in British Columbia were geographically closer and were also growing in a cool, maritime climate, the biomass equations for those areas were no more similar to those in southeast Alaska than the equations from sites in the extreme continental climate of northern Minnesota. In both regions, equations were found that overestimated biomass for some species and underestimated biomass for others, generally with equal magnitude.

Table 3. Comparisons of biomass equation estimates at 50% cover for herbs from three geographic regions.

Species	----- % of estimate from southeast Alaska -----	
	Northern Minnesota *	southwest British Columbia **
<u>Cornus canadensis</u>	+14.8	--
<u>Gymnocarpium dryopteris</u>	--	-1.61
<u>Maianthemum dilatata</u>	-17.6	--
<u>Rubus pedatus</u>	--	-30.1
<u>Streptopus roseus</u>	+268.4	+199.0
<u>Tiarella trifoliata</u>	--	+225.8

* Ohmann et al. (1981)

** Yarie (1980)

Equations for biomass components of woody shrub species from other geographic regions gave more consistent bias compared with equations from southeast Alaska than did equations for the herbs (Figures 8 and 9). Equations from the northern Rocky Mountains underestimated total biomass for both the riparian species *Oplopanax horridum* and the upland forest species *Menziesia ferruginea*. Biomass equations from Oregon and British Columbia predicted double or more the woody biomass of *Rubus spectabilis* and *Vaccinium alaskaense*. Prediction of twig and foliar biomass from Oregon and British Columbia equations generally underestimated biomass in southeast Alaska for *R. spectabilis* and *V. alaskaense*. The only exception was twig and foliar production of *V. alaskaense* in Oregon, which overestimated biomass relative to the equations from southeast Alaska. The increased biomass and vigor of shrubs from lower latitudes should not be too surprising because these forests are more open and generally grow on better drained, more productive sites (Gholz 1982, Grier and Logan 1977, Yarie 1980).

Conclusions

The dimension analysis technique has been widely accepted as the most efficient and precise means for estimating biomass of both trees and understory vegetation, but more attention needs to be focused on the potential errors of its application. Because understory plants are responsive to short-term changes in microclimate and localized site conditions, a much larger effort is needed in expanding existing libraries of biomass equations to include more site-specific equations. Some of these difficulties with site-specific or area-specific equations could be alleviated by using more precise independent variables, and multiple independent variables such as height x cover, canopy volume, or vigor. In places such as southeast Alaska, however, where understory shrubs often form dense thickets and have highly irregular canopy shapes, estimates of these other variables would likely make the regression technique much less efficient and useful. The simplest and most straightforward approach is to develop new equations for habitat types, age classes, or stand structures that will be the subject of intensive nondestructive sampling. Much of the existing data may also benefit from re-analysis, especially if it can be re-aggregated according to logical ecosystem groups or productivity classes. If more precision in biomass estimation equations is developed, more may be learned of the variation in vigor and structure of understory plants in response to environmental gradients, and more accurate and sensitive analyses of variation in understory growth and abundance will be possible. This increased precision will be of valuable in wildlife habitat assessment and process-oriented ecosystem studies.

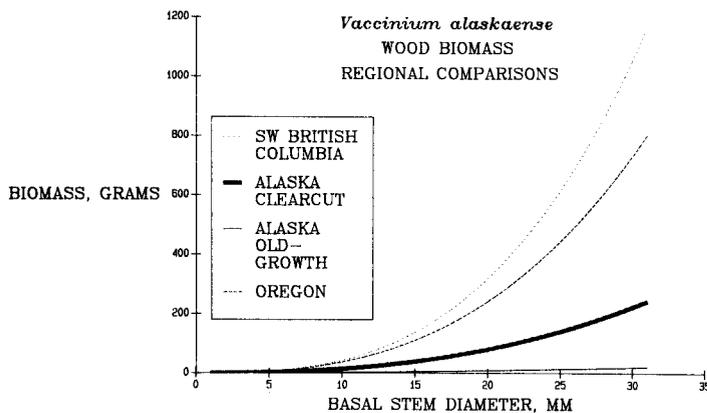


Figure 8. Woody biomass equations for *Vaccinium alaskaense* from different geographic regions.

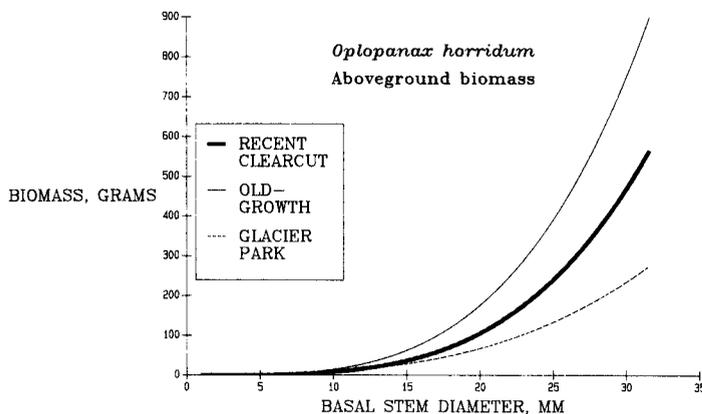


Figure 9. Total aboveground biomass equations for *Oplopanax horridum*

Literature Cited

- Alaback, P.B. Biomass and production of understory vegetation in seral Sitka spruce-western hemlock forests of southeast Alaska; Ph.D. diss.; Oregon State University, Corvallis, Oregon. 79p.; 1980.
- Alaback, P.B. Dynamics of understory biomass in Sitka spruce-western hemlock forests of southeast Alaska. *Ecology* 63: 1932-1948; 1982.
- Alaback, P.B. Plant succession following logging in the Sitka spruce-western hemlock forests of southeast Alaska: implications for management. Gen. Tech. Rep. PNW-173. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station; 1984.

- Alaback, P.B. Biomass regression equations for understory plants in coastal Alaska: effects of species and sampling design on estimates. *Northwest Science* 60:90-103; 1986.
- Baskerville, G.L. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* 2:49-53; 1972.
- Brown, J.K. Estimating shrub biomass from basal stem diameters. *Can. J. For. Res.* 6:154-158; 1976.
- Cunia, T. Dummy variables and some of their uses in regression analysis. In: T. Cunia, K. Kuusela, and A.J. Nash, editors, *Proceedings of the June 1973 meeting of IUFRO subject group 54.02*. Vol 1. Nancy, France; SUNY College of Environmental Science and Forestry, Syracuse, New York; 1973.
- Gholz, H.L. Assessing stress in *Rhododendron macrophyllum* through an analysis of leaf physical and chemical characteristics. *Canadian Journal of Botany* 56:546-556; 1978.
- Gholz, H.L. Environmental limits on aboveground net primary production, leaf area, and biomass in vegetation zones of the Pacific Northwest. *Ecology* 63:469-481; 1982.
- Gholz, H.L., C.C. Grier, A.G. Campbell, and A.T. Brown. Equations and their use for estimating biomass and leaf area of plants in the Pacific Northwest. *Res. Pap. 41.*, Corvallis, OR: Forest Research Laboratory, Oregon State University; 1979. 37 p.
- Grier, C.C., and R.S. Logan. Old growth *Pseudotsuga menziesii* communities of a western Oregon watershed: biomass distribution and production budgets. *Ecological Monographs* 47:373-400; 1977.
- Hulten, E. *Flora of Alaska and neighboring territories. A manual of the vascular plants.* Stanford, CA: Stanford University Press; 1968. 1008 p.
- Ohmann, L.F., and D.F. Grigal. Biomass distribution of unmanaged upland forests in Minnesota. *Forest Ecology and Management* 13: 205-222; 1985.
- Ohmann, L.F., D.F. Grigal, and L.L. Rogers. Estimating plant biomass for undergrowth species of northeastern Minnesota. *Gen. Tech. Rep. NC-61.* St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1981. 10p.
- Pastor, J., J.D. Aber, and J.M. Mellilo. Biomass prediction using generalized allometric regressions for some Northeast tree species. *Forest Ecology and Management* 7:265-274; 1984.
- Smith, W.B., and G.J. Brand. Allometric biomass equations for 96 species of herbs, shrubs, and small trees. *Res. Note NC-299.* St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station; 1983. 8 p.
- Stanek, W., and D. State. Equations predicting primary productivity (biomass) of trees, shrubs, and lesser vegetation based on current literature. *Inf. Rep. BC-X-183.* Victoria, BC: Canadian Forest Service Pacific Forest Research Centre; 1978.
- Tappeiner, J.C., II., and H.H. John. Biomass and nutrient content of hazel undergrowth. *Ecology* 54:1342-1348; 1973.
- Whittaker, R.H. Estimation of net primary production of forest and shrub communities. *Ecology* 42:177-180; 1961.
- Yarie, J. The role of understory vegetation in the nutrient cycle of forested ecosystems in the mountain hemlock biogeoclimatic zone. *Ecology* 61:1498-1514; 1980.

BIOMASS ESTIMATES FOR NONTIMBER VEGETATION IN THE
TANANA RIVER BASIN OF INTERIOR ALASKA

Bert Mead, John Yarie, and David Herman

Bert R. Mead is a forester and David Herman is a computer programmer, Forest Inventory and Analysis for Alaska Forestry Sciences Laboratory, Pacific Northwest Research Station, USDA Forest Service, 201 East Ninth Avenue, Suite 303, Anchorage, Alaska 99501. John Yarie is an assistant professor of silviculture and forest ecology, Forest Soils Laboratory, University of Alaska, Fairbanks, Alaska 99775-5500.

Vegetation on sample inventory plots in the Tanana River basin was described using horizontal-vertical profile descriptor techniques to show percentage of foliar cover and average height by species. These profiles were used in conjunction with species biomass equations to develop total biomass estimates for each plot. Equations were previously developed for all major species in the area, minor plants were grouped into lifeforms, and the most similar major species equation was applied for weight estimation. Sample weight profiles and a preliminary biomass summary by vegetation type are given. Usefulness of such biomass information for wildlife habitat analysis is discussed.

Introduction

The research unit, Forest Inventory and Analysis for Alaska, USDA Forest Service, has been describing and evaluating all vegetation (timber and nontimber) on inventory plots since 1978. We no longer sample timber to the exclusion of other vegetation resources. This makes good sense for Alaska as much of interior Alaska is currently only marginal in producing wood products; the main economic value of the land may be as fish and wildlife habitat and for oil and mineral production. Identifying areas of good wildlife habitat for a particular animal requires sufficient knowledge of the vegetation to quantify preferred plants in its composition and to identify beneficial patterns of vegetation layering. Monitoring change caused by oil and mineral production requires good baseline vegetation data to adequately assess the impact on the surrounding environment. Also, air pollution originating in industrialized cities in the Soviet Union and Europe is a concern in arctic areas. Plant composition and total biomass changes in various communities may be good indicators of changes created by pollution.

Two of Alaska's first large-area inventories that included nontree vegetation, were done in the 3.6-million-hectare (9-million-acre) Porcupine River basin and the 6.4-million-hectare (16-million-acre) Susitna River basin from 1978 to 1980. Measurement of nontimber vegetation in the Susitna River basin was directed toward estimating annual production of herbage and browse rather than estimating biomass. Annual production was estimated by a double-sampling clip and weigh technique at each sample plot (USDA 1976). Recorded weights were related to foliar cover estimates to project the amount of browse and herbage production for each plot (USDA In press). It became clear that these plot-by-plot weight estimates of annual growth were varying more because of the date of sampling than for any real differences in site productivity.

The problem of obtaining good estimates of productivity is magnified by the short, compressed growing season in Alaska and the elevational differences encountered in large-area river basin inventories. River basin sample units always include extremes of elevation from river bed to surrounding mountainous terrain, whereas conventional timber inventories are generally restricted to the more productive lowlands of Alaska. Attempts were made to adjust the timing of sampling in an already short field season so that plots were measured near the peak of their vegetative development; however, the costs incurred in returning to remote areas to pick up a few isolated high-elevation plots later in the season quickly precluded this as a reasonable solution. The cost of access to much of the sample area also precluded a separate phenological development study whereby subsampled plots would be visited repeatedly to monitor and adjust for observed seasonal differences in phenology. Because of these problems, we concluded that estimates of current annual growth on nontree vegetation were of questionable value. A more appropriate assessment of vegetation seemed to be to describe composition and plant biomass. Data on plant biomass may also be more meaningful for wildlife habitat assessments, which examine total cover and total available browse. Future studies may be able to correlate a site's annual productivity with its biomass and related site factors, such as elevation and aspect.

Biomass estimates have an advantage over browse inventories because the former describe the total resource. No subjective judgments are made on what is available, based on height and twig size considerations, because these vary with yearly snow depth and with differences in use by individual animals. Biomass estimates do not eliminate the sampling problems associated with phenological development but they do minimize them. Estimates are made from percent cover values and applied to equations which were constructed by sampling at sites with maximum seasonal development. Cover estimates on inventory sites which have not reached their full seasonal development are more accurately adjusted than are weight samples of annual growth.

The decision to estimate herbaceous and shrub biomass on our inventory plots was made at the same time we were totally revising our sampling system. We changed from sampling 0.4-hectare (1-acre) homogeneous plots to 8-hectare (20-acre) heterogeneous ground plots, which could include several forest and nonforest vegetation types in each 8-hectare sample. The large plots allowed for collection of more information at a site and were determined to achieve better sampling of edge conditions. It was also more cost effective to sample multiple conditions at fewer sites in a remote area that required helicopter access. We adapted the horizontal-vertical vegetation profile technique developed by the Southeastern Forest Research Station (Cost 1979) as part of the new sample design. This technique, describes vegetation in vertical layers and gives a more complete picture of total browse under a variety of snow conditions; and, it is more descriptive of vegetation in general.

It was against this background that a cooperative study was initiated between FIA and the University of Alaska to develop biomass equations for the Tanana River basin--an area equivalent in size to the States of New York and Connecticut (13.8 million hectares, 34.5 million acres) and with an elevation range from 120 meters (400 feet) at the confluence of the Tanana and Yukon Rivers to 6,096 meters (20,320 feet) at the top of Mount McKinley. See figure 1.

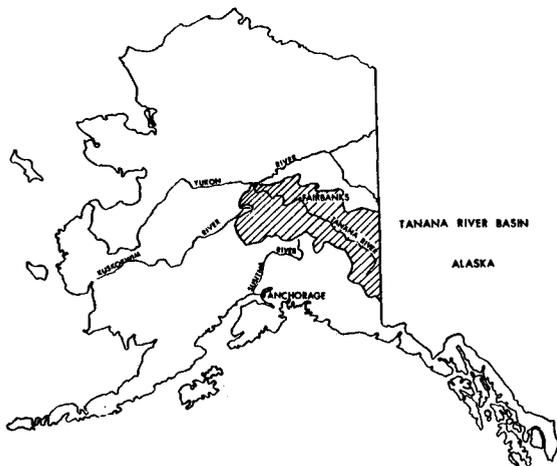


Figure 1. Map of the Alaska showing location of the Tanana River basin.

Methods

Ten of the most common vegetation types in the Tanana River basin were selected to represent the range in vegetation from closed forest to more open shrubland and grassland. Sample locations were selected randomly among the inventory plots that contained these vegetation types. Two transects were located in each sampling location. The first began at a random

starting point. Each transect was 120 meters (400 feet) long; the second line was 50 meters (165 feet) from the first and ran parallel to it. Rectangular biomass sampling plots 0.91 by 0.61 meters (2 by 3 feet) were positioned along the transects every 20 meters (66 feet). A vertical profile was defined by hanging ropes from an aluminum ladder in a technique similar to that described by Harcombe and Marks (1977). The ropes were color coded to delineate 0.3-meters (1 foot) vertical segments.

The ground layer was sampled using a 0.093-square-meter (0.102-square yards) subplot. This sample included all mosses and lichens and any other small plant not exceeding 2.5 centimeters (1.0 inch) in height when fully grown. Once the entire plot was established, the percentage of cover was estimated within each vertical segment starting at 5.3 m (16 feet) and progressing to the ground layer. The material hanging within each segment was clipped. The samples were oven dried and the weight of leaves and twigs determined separately. Frequent checks were made between field crews obtaining biomass information and those working on horizontal-vertical profile plots to determine if significant bias in estimating the percentage of cover was occurring. Results demonstrated repeatability within 10 percent with some of the variation caused by minor differences in vertical layer breaks.

Equations of the form, biomass for 10 cm layer = percent cover multiplied by regression coefficient, were developed from these data by using a conditioned linear regression equation to define the relationship between the percentage of cover and the biomass. Equations were developed for 58 species.

Biomass estimates were made for each inventory plot by using the percentage of cover and height estimates taken on horizontal-vertical profile plots and applying the equations developed for that plant species or by matching its lifeform (as determined by its average size, shape, and weight) to the most similar species for which an equation was available. These data can be summarized in numerous ways: by vegetation types, by natural layers, by arbitrary height classes, or by lifeform grouping.

Discussion

By sampling all the plants in each major vegetation type in the river basin, we were able to develop equations for the most common plants and the plants significant to major wildlife species. Some groups of plants, such as ferns and tall forbs, were missed using this sampling procedure and additional work is needed to develop equations for those lifeforms. Additional study is also needed on shrub biomass of stems and twigs larger than 5 millimeters in diameter. Biomass of these components is highly

variable in relation to the percentage of cover; therefore, they were not included in the initial studies. Estimates of percentage of cover of larger stems and twigs is highly dependent on twig orientation, and biomass of these components may be more accurately assessed by stem size counts.

Tables 1 and 2 show typical plot summaries charted by plant species and by layer. They provide very detailed species-level weight data, which are useful in quantifying occurrence of preferred plant species in wildlife habitats. For example, fruticose and squamulose lichens such as Cetraria and Cladonia are common forage for caribou (Rangifer tarandus) because of the high carbohydrate content and easy digestibility. Other lichens, however, are seldom eaten by caribou. Moose (Alces alces) supplement their browse diets, which consist primarily of willow (Salix sp.) and birch (Betula sp.), with foliose lichens such as Peltigera as well as a variety of other plants. Small mammals, such as snowshoe hare (Lepus americanus), are also known to selectively browse low shrubs and have a strong preference for Vaccinium uliginosum and Ledum groenlandicum. Vaccinium vitis-idaea is an important forage supplement for moose and grizzly bear (Ursus arctos). Although not appearing in the sample

profiles presented, an example of a preferred and less abundant plant in grizzly bear habitat may be Hedysarum whose roots provide essential forage in early spring and late fall. Boykinia is another summer forage plant important to the grizzly bear (United States Department of Interior 1980).

Further summarization of plots within vegetation types is shown in Table 3. Lichens appear to be a common source of forage among all vegetation types. Willow and birch shrubs, which are favored by moose, also represent a substantial portion of the biomass depicted. Preliminary willow and birch biomass estimates for the Tanana River basin are considerably higher than annual browse production figures reported in other studies. This difference is accounted for in older material that is not part of the current year's growth. This second- and possibly third-year material may not be as palatable for moose but is within the twig size reportedly used (up to 4.0 millimeters) (Wolff 1976). Average total biomass estimates from our study appear reasonable when compared with figures of total annual growth reported by Hanley (1984) of 3,266 kilogram/hectare (3,952 pounds/acre) in the understory of an old-growth coniferous forest in Washington.

Table 1. Biomass profile for a single plot in a closed deciduous forest, by plant species and layer.

Species	Ground Layer 2 Layer 3 Layer 4 (vertical extent of layer in decimeters)			
	0	0 - 1	1 - 4	4 - 20
(kilograms/hectare)				
Mosses:				
<u>Brachythecium</u> spp.	118	--	--	--
<u>Hylocomium splendens</u>	144	--	--	--
<u>Pleurozium schreberi</u>	347	--	--	--
<u>Polytrichum</u> spp.	322	--	--	--
<u>Ptilium crista-castrensis</u>	69	--	--	--
Grasses:				
<u>Calamagrostis canadensis</u>	t ^{a/}	t	133	--
<u>Poa alpigena</u>	t	t	33	--
Shrubs: ^{b/}				
<u>Linnea borealis</u>	30	30	--	--
<u>Ledum groenlandicum</u>	t	t	395	--
<u>Rosa acicularis</u>	t	t	35	--
<u>Vaccinium vitis-idaea</u>	t	461	--	--
Trees: ^{b/}				
<u>Alnus crispa</u>	--	--	72	450
<u>Betula papyrifera</u>	--	--	73	--
<u>Picea mariana</u>	--	--	141	1,411

^{a/} Species occurs at trace amounts (<1 kg/ha) with no significant biomass in these layers.

^{b/} Does not include woody material >5 mm in diameter or above 5 meters high.

Table 2. Biomass profile for a single plot in open deciduous forest, by plant species and layer.

Species	Ground	Layer 2	Layer 3	Layer 4	Layer 5
	(vertical extent of layer in decimeters)				
	0	0 - 2	2 - 4	4 - 11	11 - 50
(kilograms/hectare)					
Lichens:					
<u>Cladina</u> spp.	121	--	--	--	--
<u>Cladonia</u> spp.	216	--	--	--	--
Mosses:					
<u>Mnium</u> spp.	26	--	--	--	--
<u>Polytrichum</u> spp.	129	--	--	--	--
Clubmoss:					
<u>Lycopodium annotinum</u>	48	--	--	--	--
Forbs:					
<u>Epilobium angustifolium</u>	t ^{a/}	14	14	--	--
<u>Mertensia paniculata</u>	t	23	--	--	--
<u>Polygonum</u> spp.	t	17	--	--	--
Shrubs: ^{b/}					
<u>Betula occidentalis</u>	t	216	198	1,323	3,510
<u>Betula papyrifera</u>	t	18	37	321	--
<u>Ledum groenlandicum</u>	t	329	691	--	--
<u>Ledum palustre</u>	t	185	83	--	--
<u>Rosa acicularis</u>	t	18	4	--	--
<u>Salix glauca</u>	t	52	35	426	--
<u>Spiraea Beauverdiana</u>	t	1	1	--	--
<u>Vaccinium vitis-idaea</u>	t	146	--	--	--

^{a/} Species occurs at trace amounts (<1 kg/ha) with no significant biomass in these layers.

^{b/} Does not include woody material >5 mm in diameter or above 5 meters high.

Table 3. Average biomass by plant grouping and vegetation type for sixteen sample plots.

	Forest	Scrub	Tall Shrub	Low Shrub	Dwarf Shrub	Herbaceous
	(kilograms/hectare)					
Lichen	926	993	425	1,167	2,006	572
Moss	1,057	1,524	593	1,197	1,276	870
Grass	164	222	359	283	138	1,209
Forbs	124	60	204	69	62	272
Low shrub	384	333	108	216	234	89
Willow, birch ^{a/}	2,396	1,172	4,914	3,390	894	960
Alder ^{a/}	1,682	766	1,899	212	--	309
Tree seedlings ^{a/}	5,892	11,066	1,350	778	--	396
TOTAL	12,625	16,136	9,552	7,312	4,610	4,677
Number of plots	127	23	23	25	7	17

^{a/} Does not include woody material >5 mm in diameter or above 5 meters high.

Biomass estimates for the river basin provide a good starting point for further wildlife habitat analysis. For instance, willow and birch shrubs occur in most vegetation types; however further characterization shows some of this biomass is above or below browsing reach, of insufficient density, or in areas where terrain or snow depth would prevent utilization. The remaining areas can be evaluated based on knowledge of moose consumption requirements. Studies have shown moose consumption at approximately 5 kilograms per day whereas willow and birch biomass in the Tanana ranges from 100-2,500 kilogram/hectare, depending on the vegetation type. Canadian estimates over extensive areas indicate actual utilization may be in the range of 5-10 percent of the estimated total browse (Telfer 1981). Applying those estimates to the Tanana basin, a productive hectare might provide up to 50 days of browse for moose. However, year-round densities of moose over large areas seldom exceed 12-15 per 1000 hectare (2,500 acres) and are more commonly in the range of 4-6. Because low-protein content limits digestibility of browse, large biomasses may not be food unless additional high-protein forage is also obtainable (Telfer 1981). Our plant species profiles within individual vegetation types will aid in identifying those areas with high-protein forage and with sufficient browse.

The examples given above are not intended to oversimplify the complex task of describing good habitat for a wildlife species. Vegetation descriptions by themselves cannot describe preferred habitats but used in conjunction with terrain descriptions of slope, elevation, and aspect and with snow surveys, they further describe areas likely to provide suitable habitat. Amount of hiding cover, interspersions of forest/nonforest edge, and distance to water; data which are obtained from our photo sampling, will further refine our area estimates of good habitat. This should allow us to assign each plot with wildlife habitat value ratings for several important wildlife species. These value ratings, when tied to vegetation types and other site descriptors, will enable us to produce broad estimates of area of suitable and unsuitable habitat for a particular animal. Further sampling of good habitat areas by interested cooperators would allow them to make estimates of carrying capacity for each wildlife species of interest.

This large data-base of vegetation data could also be applied to the study of habitat requirements of the animals themselves. Very little information is known about the habitat requirements of Alaska's major wildlife species with the possible exception of moose. Analytical study of vegetation and other factors in preferred habitats could identify the important factors in habitat selection and lead to further understanding of the animals themselves.

Conclusions

Biomass and cover estimates provide a useful way to quantify and describe Alaska's vegetation. This presents the new challenge of identifying the vegetation biomass requirements of various wildlife species. These could be expressed as: amount of browse or forage required, amount of preferred plants needed, or as some desirable form of vegetative layering and composition, depending on the wildlife species being studied. This will allow us to make more refined assessments of wildlife habitat than were previously possible. This work has only begun but promises big payoffs in the future.

Literature Cited

- Cost, N. D. Ecological structure of forest vegetation. In: Frayer, W. E., ed. Forest resource inventories, workshop proceedings. 1979 July 23-26; Fort Collins, CO. Fort Collins, CO: Colorado State University; Dept. of Forest and Wood Sciences, 1979: 29-37. Vol. 1 of 2 vol.
- Hanley, T. A. Habitat patches and their selection by wapiti and black-tailed deer in a coastal montane coniferous forest. *Journal of Applied Ecology*. 21:423-436; 1984.
- Harcombe, P. A.; Marks, P. L. Understorey structure of a mesic forest in southeast Texas. *Ecology*. 58: 1144-1151; 1977.
- Telfer, Edmund S. Browse inventories: techniques and evaluation. In: Miller, F. L.; Gunn, A., eds. Symposium on census and inventory methods for population and habitats: northwest section, The Wildlife Society: Proceedings. 1980 April 10; Banff, Alberta, Canada. Contribution 217. Moscow, ID: University of Idaho; 1981: 67-82.
- U.S. Department of Agriculture, Soil Conservation Service. Susitna River basin vegetation report. Anchorage, AK: U.S. Department of Agriculture, Soil Conservation Service; [in press].
- U.S. Department of Agriculture, Soil Conservation Service. National range handbook. Washington, DC: U.S. Department of Agriculture, Soil Conservation Service; 1976. 131 p.
- U.S. Department of the Interior, Fish and Wildlife Service. Terrestrial habitat evaluation criteria handbook--Alaska. Anchorage, AK: U.S. Department of the Interior, Fish and Wildlife Service, Division of Ecological Services; 1980. 850 p.
- Wolff, J. O. Utilization of hardwood browse by moose on the Tanana flood plain of interior Alaska. Res. Note PNW-267. Portland, OR: U.S. Department of Agriculture, Forest Service; 1976; 7 p.