STATISTICAL METHODOLOGY IN FORESTRY

Timothy G. Gregoire
Yale University, New Haven, CT, USA

Keywords: Basal area, biomass, breast height, *dbh* (diameter at breast height), forest inventory, taper models, forest growth and yield models, site quality, stand density

Contents

Summary

Forest biometrics comprises the use of statistical methods for the quantitative assessment of forest resources. Forest surveys are conducted with the use of probability sampling methods ranging from a simple random selection of plot locations to multistage and multiphase designs making use of auxiliary information from aerial photography and satellite imagery. Characteristics of individual trees often are modeled using regression techniques, some of which make use of allometric considerations. For example, the diameter of a tree bole may be used as a predictor of the tree's height or the biomass of the bole. Because of the traditional interest in the woody material and fiber in the tree bole, the morphology of the bole has been modeled much more extensively than other parts of the tree. Increasing interest in forest ecology, e.g. driven by the need to understand the effect of elevated CO_2 on tree physiology, has motivated a increased modeling of foliage and root morphology and growth, too. Statistical models also have been used to quantify the growth and dynamics of aggregations of trees in forest stands. These models range from simple, single-equation linear models to multiple and interdependent nonlinear models. Statistically designed experiments are conducted in order to gain a better understanding of tree, stand, and forest responses to planned

manipulations, (e.g. fertilization, genetic improvement, site preparation) and natural events (insect infestations, ice storms, hurricanes, temperature change).

1. Introduction

Forestry can be characterized as the deliberate management of forests to achieve stakeholder objectives. The perceived stakeholders have changed over time but rarely are limited to forest owners, even when the latter cohort exercise principal authority in setting management policy. Likewise, objectives vary among stakeholder groups. Oftentimes these objectives may be economically related to resource extraction, but not always: preservation of biodiversity, hydrological resources, recreation amenities, or even religious and social values may take precedence over the extraction of resources in many settings. Presently, forest management to achieve multiple objectives increasingly recognizes that the multiple objectives are more complementary than competing. In nearly all regions of the globe informed forest management is based on a foundation of the biological, physical, and social sciences. Hence, forestry as broadly defined here is an interdisciplinary pursuit. While trees, perhaps, may be the defining resource of forests, sustainable forest management requires much more than a knowledge of the biological basis of tree growth and dynamics.

Statistics has played a vital role in the development of science that undergirds informed forest management. Defining forest biometry as the development and application of statistical methods to assess, estimate, and evaluate biological characteristics and processes of forests, forest biometry has had a pervasive influence on forestry science. The sampling of forests in order to estimate the abundance of one or more forest resources has long had a probabilistic basis. Regression models of varying complexity have been fitted to portray the expected size or growth of individual trees conditionally upon some other measure of size, age, species, or other features of the plant. Scaling up from the individual to an aggregate of trees, homogenous areas of forest, called forest stands, have been modeled for purposes of characterizing, *inter alia*, expected growth, mortality, and composition. Indices have been developed to characterize the productive capacity of a forest stand, its diversity, and the level of competitive interaction among trees and other plants of the forest. Statistically designed experiments, both in the forest and in the laboratory, have guided our understanding of forest regeneration and response to management intervention.

2. Forest Inventory

While an inventory of forest resources does not preclude a census, the sheer magnitude of the task almost always requires the use of sampling techniques. Evolving from strip surveys that apparently originated in Scandinavia and elsewhere in Northern Europe, many forest sampling designs rely on the random or systematic arrangement of plots (quadrats) of some fixed, predetermined area. From a practical standpoint, a random placement of plots is more difficult to locate than a systematic placement, at least historically, before the development of global positioning systems. Moreover, systematic placement of sampling locations ensures their more uniform spatial distribution, a feature that has broad appeal. Extensive studies have established that estimators of forest characteristics have sampling distributions very closely

approximated by their distributions under simple random sampling, so long as reasonable care is taken to ensure that i) the distance between plots is sufficient to nullify spatial autocorrelation, and ii) transects along which plots are aligned do not follow topographical gradients that affect the resource to be estimated.

Both the size of the plot, as measured by its area, say *a*, and its shape typically are held constant for the entire inventory. Circular and rectangular plots predominate in current use. A variant of this practice occurs when one size plot is used to sample trees, and a smaller, nested plot is used to sample smaller vegetation such as seedlings, herbaceous plants, wildlife scat, litter, mast, etc.

Given the probabilistic selection of plot (sampling) locations, each item sampled on the plot has a corresponding and equal probability of being selected, namely $p = a/A$, where A is the area of the region within which plots can be located.

With the advent of aerial photography in the early decades of the 20$^{\text{th}}$ century, forest inventory specialists were quick to seize upon the benefits of stratifying the forest into regions of homogeneous forest cover, and then sampling independently within each stratum. In a stratified survey, the selection probability of any item within stratum *h*, say, is $p = a/A_h$, where A_h is the area of the stratum. Increasingly, strata may be delineated on the basis of satellite imagery.

Forest inventory has almost always relied on the sampling design as the basis of statistical inference, although this has not always been recognized and abundant confusion over design-based and model-based inference, in particular, has been evident.

To preserve design-unbiasedness of estimation, stratification requires that each point of forestland be assigned to one stratum only, and that every point be assigned. In contrast, a plot sample rarely can be designed so that the region, or stratum thereof, is exhaustively partitioned into *N* plots from which a sample of *n* is selected. Consequently, the areal frame is more appropriately viewed as an infinite number of plot locations (*e.g*, centers of circular plots or a pre-assigned corner of a rectangular plot) of which *n* are selected at random or systematically with a random start with probability density $1/A$.

Forest inventories typically are intended to provide estimates of multiple attributes of the forest. For sake of example, suppose interest focuses on the estimation of the total aboveground biomass, say τ, of a particular plant species.

Let Y_i be the aboveground biomass of this species measured on the *i*th sample plot, $i = 1, \cdots, n$. There are many ways in which the sample data $\Delta = (Y_1, \cdots, Y_n)$ can be used to estimate τ, but the usual expansion estimator is the most commonly used. With this estimator, each Y_i is expanded by the factor $p = a/A$ to yield $\hat{Y}_i = Y_i/p$; and then the $\hat{Y}_i$ are averaged, *viz*:

$$\hat{\tau} = n^{-1}\sum_{i=1}^{n}\hat{Y}_i = n^{-1}\sum_{i=1}^{n}Y_i p^{-1} = n^{-1}a^{-1}A\sum_{i=1}^{n}Y_i.$$

An estimator of aboveground biomass per hectare is obtained as $\hat{\tau}^* = \hat{\tau}/A.$

Inventories with a particular emphasis on tree and woody plants provide the basis for estimating tree frequency and other attributes (aggregate basal area, biomass, volume) in each stratum. These commonly would be disaggregated into a two-way layout showing the estimated amount of the resource in each diameter class for each species. Specifically, let Y_{ids} denote the aboveground biomass of species s, $s = 1, \cdots, S$, in diameter class d, $d = 1, \cdots, D$, measured on the ith plot. Then the estimated biomass of the forest in the dth row and sth column of the table is computed as

$$\hat{t}_{ds} = n^{-1}a^{-1}A\sum_{i=1}^{n}Y_{ids},$$

and $\hat{\tau}_{ds}^* = \hat{\tau}_{ds}/A$ is the estimated value on a per hectare basis. These estimators are additive in the sense that $\hat{\tau} = \sum_{d=1}^{D}\sum_{s=1}^{S}\hat{\tau}_{ds}.$

If there are plants so near to the edge of the forest region or stratum being sampled that their selection probability is less than $p = a/A$, then the above estimator is biased. Numerous solutions to this ``boundary bias'' problem have been devised.

Plot sampling selects plants and other discrete forest organisms into the sample with equal probability, barring edge effect as noted above. When trees are the sole focus of forest inventory, a method of unequal probability sampling, introduced by Bitterlich as angle count sampling and expanded upon by Grosenbaugh as horizontal point sampling, is widely used. With this method, a tree is selected into the sample with a probability proportional to its basal area: trees with larger diameters are selected with greater probability than trees with smaller diameters. It is a very efficient method of sampling for estimation of those forest resources that are well and positively correlated with tree basal area; *e.g*, aggregate bole volume or biomass.

Poisson sampling, introduced into the forestry literature by Grosenbaugh as 3P sampling, is yet another method of sampling trees with unequal selection probabilities. This sampling method is free of the areal basis for selection that characterizes plot and horizontal point sampling.

For understory vegetation, downed trees, logging slash, and coarse woody debris, the method of line intersect sampling has often been used.

3. Modeling Individual Tree Characteristics

Many tree species feature a main stem or bole from which there is, more or less, lateral branching supporting foliage. The crown of a tree is that part bearing foliage and live

branches. Species displaying deliquescent or decurrent crowns are those that are formed from lateral branches that grow as fast as tree height, often resulting in repeated forking of the main stem, whereas species with excurrent crowns are those whose height growth exceeds lateral growth. Many northern conifer species exhibit excurrent crowns, and many deciduous species exhibit decurrent crowns.

The diameter of the tree bole is perhaps the most important measure of tree size. Typically the diameter of the bole decreases with increasing height aboveground. Moreover, very near the ground, tree boles exhibit flaring as the result of added tissue to provide mechanical support and resistance to being blown down by wind gusts. In view of the relationship between bole diameter and the height, h, at which it is obtained, bole diameter is symbolized as $d(h)$.

If needed for clarity, $d_o(h)$ is used to signify diameter including bark or overbark diameter, and $d_u(h)$ is used for underbark diameter. A reference height at which to measure bole diameter, which has been adopted almost universally, is 1.37 m, commonly known as breast height. Diameter at breast height is, by convention, the overbark diameter at breast height, $d_0(1.37)$, and commonly abbreviated as *dbh* or as D (both notations are used here). Diameter at breast height is the most fundamental of all quantitative measurements in forestry. Diameter is measured at breast height rather than at groundline because the swelling near the base of the tree bole for mechanical support usually subsides at breast height. For some tropical tree species which have buttressed support extending much higher than breast height, *dbh* may be of limited usefulness as an omnibus measure of tree size; oftentimes, instead, the diameter above buttress is used as the reference measure of tree diameter.

The basal area of a tree is the cross-sectional area of the tree stem at breast height. Normally expressed in m^2, basal area is computed from a measurement of *dbh* in cm as $\pi \times dbh^2 / 40000$.

The height of the tree is perhaps the second most important measurement of size. For excurrent trees, height is identical to the length of the bole, whereas for decurrent trees the height of the tree may be regarded as the length of the bole plus the height of the crown vertically above it. In the latter case, the length of the main bole to the point where it forks into the crown-supporting branches may be of more relevance (*e.g.*, if the purpose is to relate height to the volume or biomass of fibers contained in the bole), but tree height may be measured nonetheless because it can be done with less error of measurement.

In distinction to the very uniform practice of measuring tree diameter by *dbh*, the measurement of height varies considerably more by region of the globe, and by the need for merchantable rather than total tree or bole height. Merchantable height, $H_m = h(d_m)$, is the height of the bole to the point where the bole diameter has tapered to a minimum diameter below which the bole has too little economic value to be worthwhile processing or merchandising; d_m may be stipulated as an overbark or underbark diameter.

-
-
-

Bibliography

Avery, T. E. and Burkhart, H. E.. (1994). *Forest Measurements* (4th ed.) 408 pp. [This book has a good review of empirical growth and yield models developed for forest management purposes].

Bitterlich, W. (1948). Die Winkelzählprobe. *Allgemeine Forst--und Holzwirtschaftliche Zeitung* 59 (1/2), 4-5. [The introduction of angle count sampling into forestry].

Carmean, W. H. (1975). Forest site quality evaluation in the United States. *Advances in Agronomy* 27, 209-269. [A review paper of historical interest that provides a broad acquaintance with the many ways that site quality has been estimated and measured].

Gray, H. R. (1956). *The Form and Taper of Forest-Tree Stems*, Institute Paper No. 32, Imperial Forestry Institute, University of Oxford. 79 pp. [An informative and historical paper in the systematic study of stem taper].

Gregoire, T. G. and Scott, C. T. (1990). Sampling at the stand boundary: a comparison of the statistical performance among eight methods. *Research in Forest Inventory, Monitoring, Growth and Yield*} (Proceedings of the International Union of Forest Research Organizations XIX World Congress, Montreal, Canada, 5 - 11 August, 1990) (eds., H.E. Burkhart, G.M. Bonnor, J.J. Lowe), pp.78-85. Publication FWS-3-90, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University. [A statistical treatment of the boundary bias problem in forest inventory].

Gregoire, T. G., Schabenberger, O., and Barrett, J. P.. (1995). Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent plot measurements. *Canadian Journal of Forest Research* 25, 137-156. [Forest yield modeling with longitudinal data].

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research* 28, 1429-1447. [A lucid presentation of competing bases of inference in survey sampling].

Gregoire, T. G., O. Schabenberger, and F. Kong. (2000). Prediction from an integrated regression equation: a forestry application. *Biometrics* 56, 414-419. [This paper derives the statistical properties of a prediction of bole volume by integrating a linear regression model of bole taper].

Grosenbaugh, L.R. (1958). Point-sampling and line-sampling: probability theory, geometric implications, synthesis. *USDA Forest Service Occasional Paper 160*, 34p. [This report established the probabilistic basis for selecting trees with unequal probabilities].

Husch, B., Miller, C. I. , and Beers, T. W. (1982). *Forest Mensuration* (3rd ed.), 492 pp. New York: Wiley. [A good summary of quantitative measurements and methods useful for forest management].

Kaiser, L. (1983). Unbiased estimation in Line-Intercept sampling. *Biometrics* 39, 965-976. [A thorough presentation of line intersect sampling from a design-based perspective].

Maguire, D. A., Moeur, M., and Bennett, W. S.. (1994). Models for describing basal diameter and vertical distribution of primary branches in young Douglas-fir.

Forest Ecology and Management 63, 23-55. [A very readable account of efforts to model the size and distribution of primary branches on young Douglas-fir trees].

Parresol, B. R. (1999). Assessing tree and stand biomass: a review with examples and critical comparisons. *Forest Science* 45, 573-593. [An excellent overview of models and methods used for estimation of tree and stand biomass in forestry and ecology].

Schreuder, H. T., Gregoire, T. G., and Wood, G. B.. (1993). *Sampling Methods for Multi-resource Forest Inventory*, 446 pp. New York: Wiley. [A comprehensive text describing methods of probability sampling widely used in the conduct of forest inventories].

Biographical Sketch

Timothy G. Gregoire received a Bachelor of Science degree in engineering from Princeton University, 1971; an M.S. of forest resources from the University of New Hampshire, 1980; and a Ph.D. in statistical modeling of forest yeield equations from Yale University in 1985. From 1985 until 1998 he was a member of the faculty in the Department of Forestry at Virginia Polytechnic Institute and State University. Since 1998 he has been the J. P. Weyerhaeuser Jr. Professor of Forest Management at the School of Forestry and Environmental Studies at Yale University.

His research interests include sampling methods, modeling longitudinal data, nonlinear mixed-effects models, sample designs for natural resource assessment, and statistical inference. He has served as an elected officer in two capacities for the Section on Statistics and the Environment of the American Statistical Association. He has served on the Council of The International Biometric Society (TIBS), as well as on its Editorial Advisory Board. He has served in both elected and appointed capacities of the Eastern North American Region of TIBS. He is an elected Fellow of the American Statistical Association. He has been awarded the Distinguished Achievement Medal by the ASA's Section on Statistics and the Environment; and he was honored with the year 2000 award in Forest Science by the Society of American Foresters.

Gregoire has co-authored and edited *Sampling Methods for Multiresource Forest Inventory* (Wiley, 1992); *Modelling Longitudinal and Spatially Correlated Data* (Springer-Verlag, 1997); and articles and chapters in *Biometrics, The American Statistician, Journal of Agricultural, Biological, and Environmental Statistics, Environmental and Ecological Statistics, Journal of Statistical Computation and Simulation, Journal of Applied Statistics, Survey Methodology, Communications in Statistics, The Statistician, Ecology, Forest Science, Forest Ecology and Management, Canadian Journal of Forest Research, Photogrammetric Engineering and Remote Sensing, Remote Sensing of Environment, Handbook in Statistics, Volume 12: Environmental Statistics* (Elsevier, 1994), *The Productivity and Sustainability of Southern Forest Ecosystems in a Changing Environment* (Springer-Verlag, 1997).

He has served as associate editor of Biometrics, *Environmetrics, Forest Science, Silva Fennica*, and as section editor of *Encyclopedia of Environmetrics* (Wiley).