Inventory Implications of Using Sampling Variances in Estimation of Growth Model Coefficients

Albert R. Stage and William R. Wykoff

Abstract.—Variables based on stand densities or stocking have sampling errors that depend on the relation of tree size to plot size and on the spatial structure of the population. Ignoring the sampling errors of such variables, which include most measures of competition used in both distance-dependent and distance-independent growth models, can bias the predictions obtained from regression analysis.

The bias occurs if the sampling errors differ between the calibration data and the prediction data. Differences arise from several sources. For example, plot designs often differ between two major applications of growth models: updating inventory statistics, and simulating effects of proposed management actions. Even in a single analysis, not all data come from the same plot size, especially if plot designs vary by diameter classes. In addition, in growth modeling, plot-size to tree-size relations change as the trees grow and management modifies spatial structure. However, if the sampling errors of the variables are used in estimating model coefficients, and again when the model is applied, the same model can be used with different plot designs. Local variation in stocking can be measured by individual subplots in the context of a distibution of plots within the stand. When imbedded in a (semi-) distance-independent, individual-tree model for stand development, the Structural-Based-Prediction method increases model sensitivity to competition compared to ordinary least squares estimates. This model has particular application to spatially irregular or patchy stands.

Models of many ecological processes depend on variables that are defined by spacing, density, and size of individual trees or of aggregates of trees at a location. Examples range from the size of big-game home-ranges (Pauley *et al.* 1993) to estimates of sediment deposition (Elliot and Hall 1997), from estimates of Net Primary Productivity to detailed models of stand development, and you can supply many more. Once developed, application of these models by resource analysts requires inventory data to provide initial values of density and stocking. Unfortunately, both modelers and inventory analysts usually overlook the effects on their analysis of predictor variables that are sample-based, and hence include a random error. Ignoring the error can lead to underestimating the significance of variables and to biases in the predictions (Fuller 1987, Jaakkola 1967).

Although statistical solutions for errors-in-variables problems have been finding their way into forestry applications (Goelz and Burk 1996, Kangas 1998, Lappi 1991), Stage and Wykoff (1993, 1998) appear to be the first to directly solve the problem posed by sampling errors of the competition variables in growth models.

We address two sources of sampling error in competition variables—within a single plot or subplot, and between the several subplots that may be used to characterize the spatial variation within the stand or inventory cluster.

Figure 1 illustrates the sampling error in just the tree density component of competition as measured on a single plot applicable to a single subject tree. Here, the sampling variance of density about the subject tree is the variance among all possible random sampling points that would include the subject tree. This variation depends on the size and shape of plot and on the spatial arrangement of the trees. For stocking measures such as basal area per unit of land area, tree size distribution is a component of the variation. Tree-centered measures of competition are also subject to sampling variation, although from different sources (Schreuder and Williams 1995).

The within-plot source of variation is seldom measured, so we estimate it from a model of spatial variation. The second source of variation—between the subplots within an inventory cluster at a location or within a stand examination—is readily computed from the inventory data. However, if the inventory measures only a single plot at each location, the same variance model used for the within-plot variation can be invoked.

Principal Mensurationist (retired) and Research Forester, respectively, USDA Forest Service, Intermountain Research Station, Forestry Sciences Laboratory, 1221 S. Main Street, Moscow, ID 83843, USA; (208) 883-2345; FAX: (208) 883-2318; astage@moscow.com; wykoff@moscow.com.

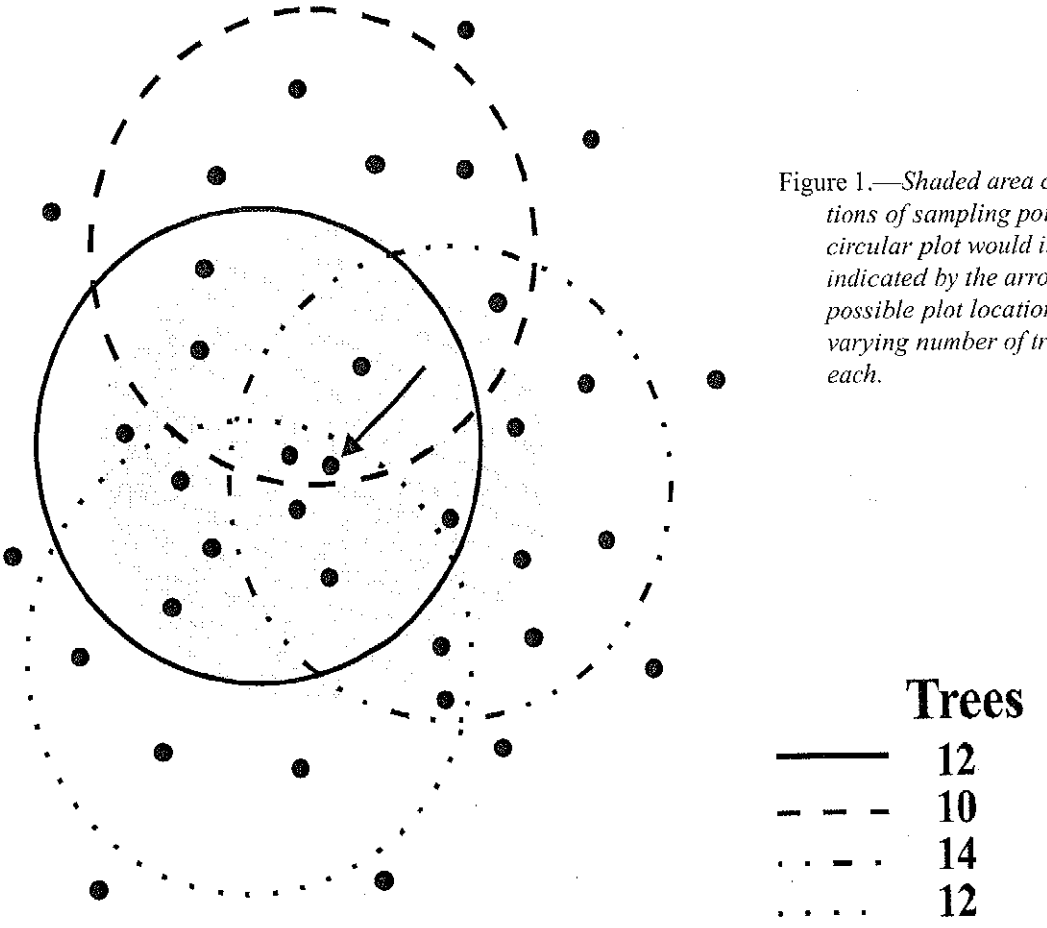


Figure 1.—*Shaded area covers possible locations of sampling points about which a circular plot would include the subject tree indicated by the arrow. Four of these possible plot locations are shown, with the varying number of trees to be tallied in each.*

INTRODUCING GROWTH MODELS TO SAMPLING ERROR

Models based on ordinary least squares estimation using inventory data can avoid the bias part of the problem if they are, in turn, applied to inventory data having the same error properties (Stage 1977). This criterion is violated when plot sizes (or number of plots per stand) differ between the calibration data set and the application data set. It is further violated in iterations of growth models beyond the first cycle of prediction if plots of differing sizes are used to sample different size classes of trees. During a projection, tree sizes change, but plot size is fixed at the time of the inventory. Likewise, outputs from growth models used as input to models of other ecosystem components are vulnerable to the same problems.

Our solution to the problem of using different plot sizes and numbers of subplots per stand or location has three steps:

1. Estimate the sampling errors applicable to the calibration data set. Depending on the definition of the competition variables, this variance may depend on the standard error of the mean between subplots. If the competition variable is defined separately for each subplot, develop a model of the variance within the subplot.

2. Subtract the sampling error covariances characterizing the calibration data set from the covariances of the observed variables. Save the coefficients of this structural model and relevant parts of the inverse of its covariance matrix to be used at prediction time.

3. When calculating predictions, the error covariances applicable to the new data set are added to the reduced covariance matrix from step 2. The result is a new covariance matrix of "observed" variables. From its inverse and the coefficients of the structural model, new coefficients are calculated for prediction. Fortunately, there are useful shortcuts to the matrix algebra that cut this task to a reasonable size. (This step is the major difference between our procedure and the one used by Kangas (1998)). We have called this three-step procedure "Structural Based Prediction" (SBP).

Admirers of Robert Service may have the sense that this procedure bears certain similarities to Chewed-Ear Jenkins' solution to the syndicate holding bets on the outcome of his hair restoring efforts. While his wife poured on the hair-restoring elixir, Chewed-Ear secretly shaved off the fuzz. However, we show that effects can be substantial when either plot configurations differ, or when management action such as thinning changes the variation within or between plots.

APPLYING SBP IN A SEMI-DISTANCE-INDEPENDENT MODEL OF FOREST GROWTH

The model we developed was motivated by the need to provide a predictive model of stand development with explicit representation of within-stand (or within-location) variation in stocking. In particular, we wanted a model that could represent stocking variation at a broader scale than seems feasible with distance-dependent formulations. Therefore, the model was designed to be used either with a stand examination consisting of a scatter of subplots through the delineated stand or with an inventory of clusters of subplots. An inventory in which the cluster is reduced to a single plot at widely scattered locations is just a special case, for which the model could be used with some loss of resolution.

Competition Measures

To meet our objective of increasing sensitivity to spatial variation, we provided for two scales of competition effects: effects dependent on the local intensity of stocking as measured by a single, rather small plot; and effects that depend on the general context of stocking in which the plot is located.

Background competition is represented by the mean of stand basal area per hectare (BA). Its sampling variation is simply the standard error of the mean basal area of the subplots in the stand. The measure of competition for trees within a subplot is the product of basal area per unit area at that particular subplot multiplied by the percentile of the subject tree in the diameter distribution of all trees in the stand. Because this index is analogous to the basal area in larger trees introduced by Wykoff *et al.* (1982) and Wykoff (1990) and used in many increment models, we continue to refer to it as BAL. Because the single plot does not provide a sample-based estimate of its basal-area variance, we derive a first approximation to the sampling error of BAL as a function of the basal area variance in a Poisson (random) field of stem locations, the percentile, and the total stem count in all subplots (Stage and Wykoff 1998).

Comparison to Ordinary Least Squares (OLS) Estimation

Differences between the structural model and an ordinary least squares model of some northern Idaho inventory data are shown in several figures; figure 2 shows remarkable changes in the effect imputed to crown ratio (including the term that depends on dbh). Note the complete reversal of the trend with dbh. Differences in predicted diameter increment are illustrated in figures 3-5 (Wykoff 1997). Both BA and BAL were competition variables for which sampling variances were estimated. Because BAL is computed from basal area at the sample point occupied by the subject tree and BA is a stand average, allowing for their relative sampling variation greatly increases the sensitivity of the model to local competition effect (fig. 3). But differences are not limited to effects of variables measured with sampling error. The effect of dbh also changes markedly (fig. 4). In this, as in many similar models, crown ratio can be the alias for direct effects of competition. Explicitly recognizing the error variance of the competition variables shifts competition effects from the crown ratio term to the competition terms (fig. 5). In analyses of thinning effects, the immediate response will be greater because crown ratio takes some time to regain its natural correlation with stocking.

We compared the effect of two different sampling designs on the relative behavior of the SBP-based model. The "stand" for which the simulations are calculated was generated by the Regeneration submodel of the Inland Empire variant of the Forest Vegetation Simulator (Prognosis). From this stand projection, a "patchy" stand was created having equal area of each of three age classes. Means of 20 simulations of two scenarios—thinned, and

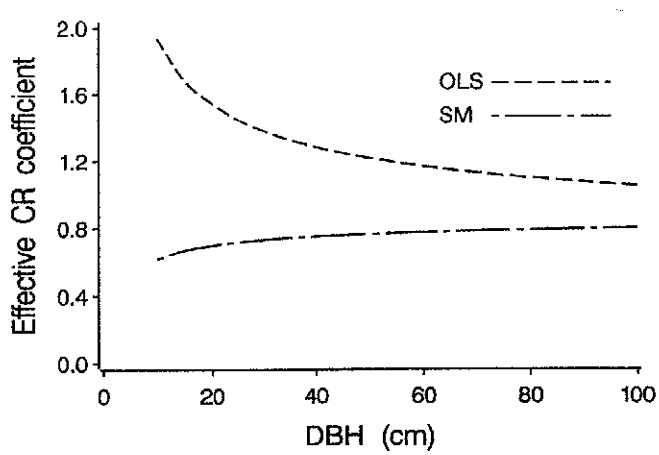
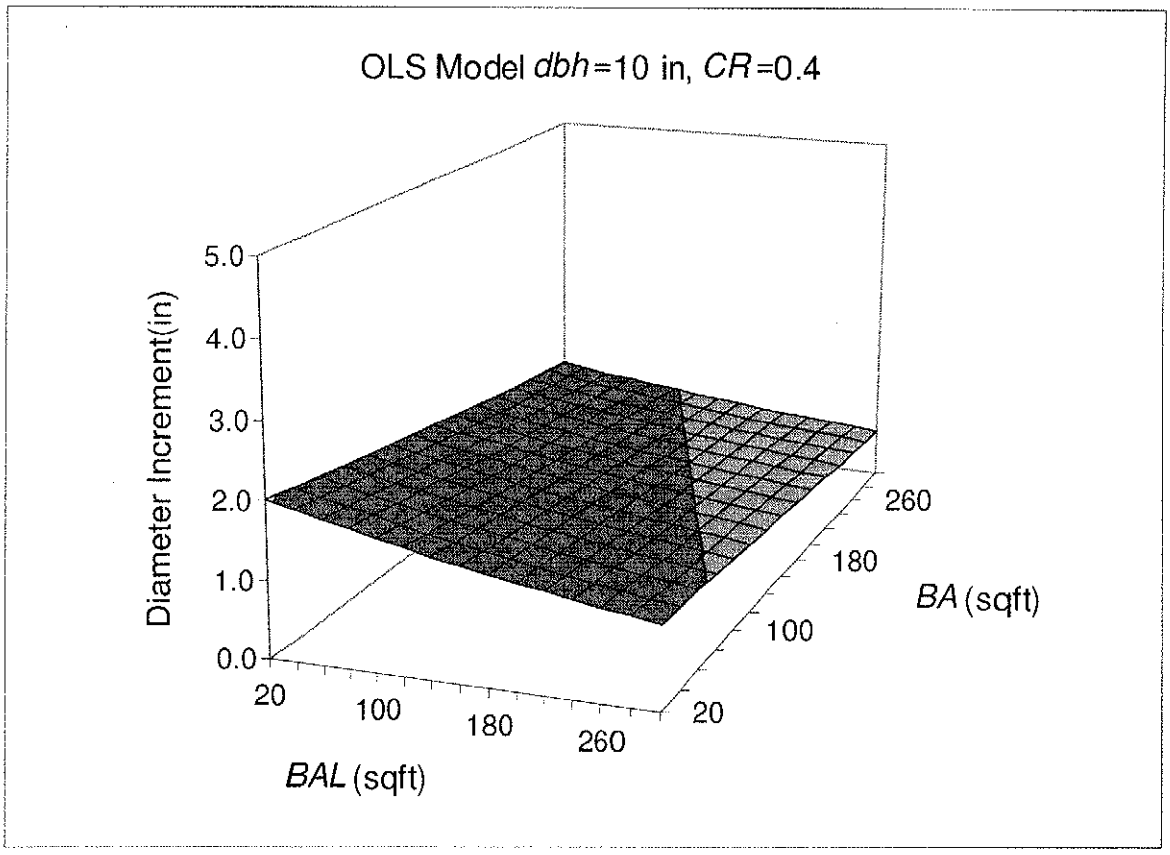


Figure 2.—*Effect of crown ratio on prediction of ln(DDS) in the structural model is very different from its effect in the ordinary least squares model because of correlation with the competition variables.*



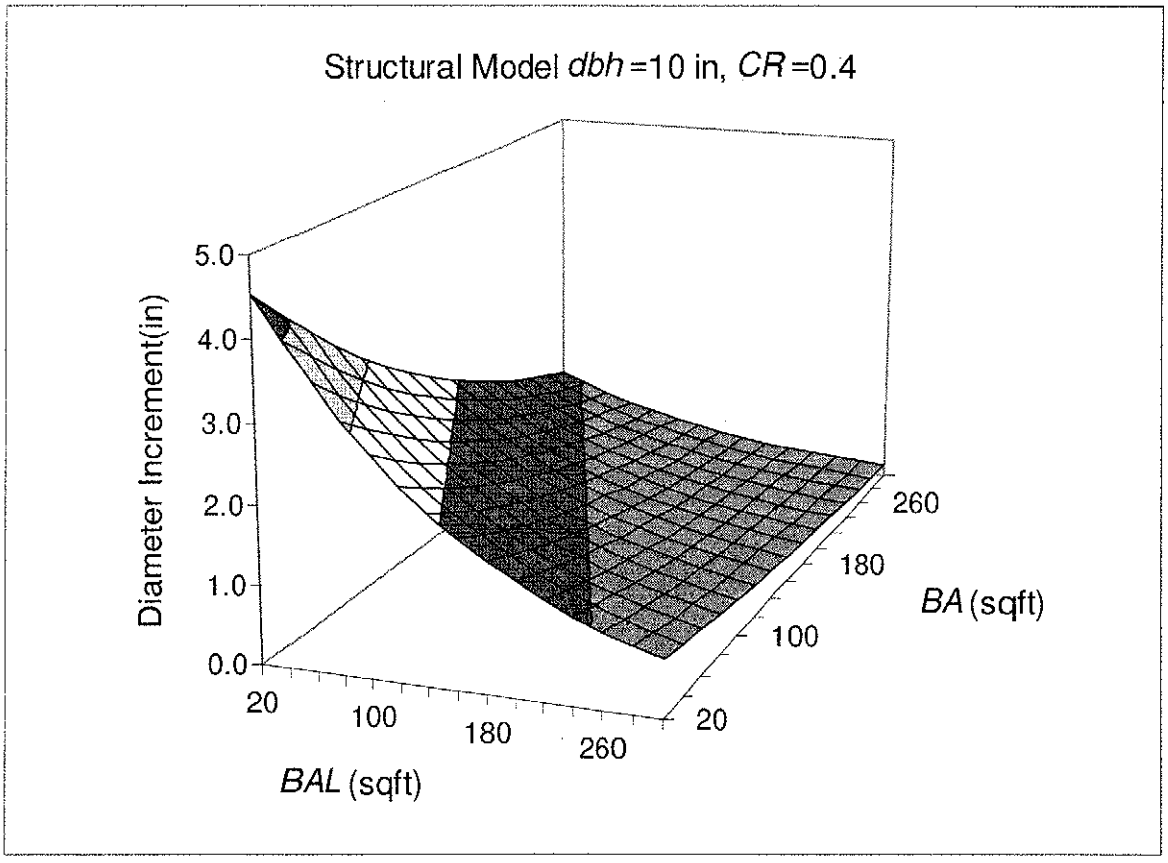
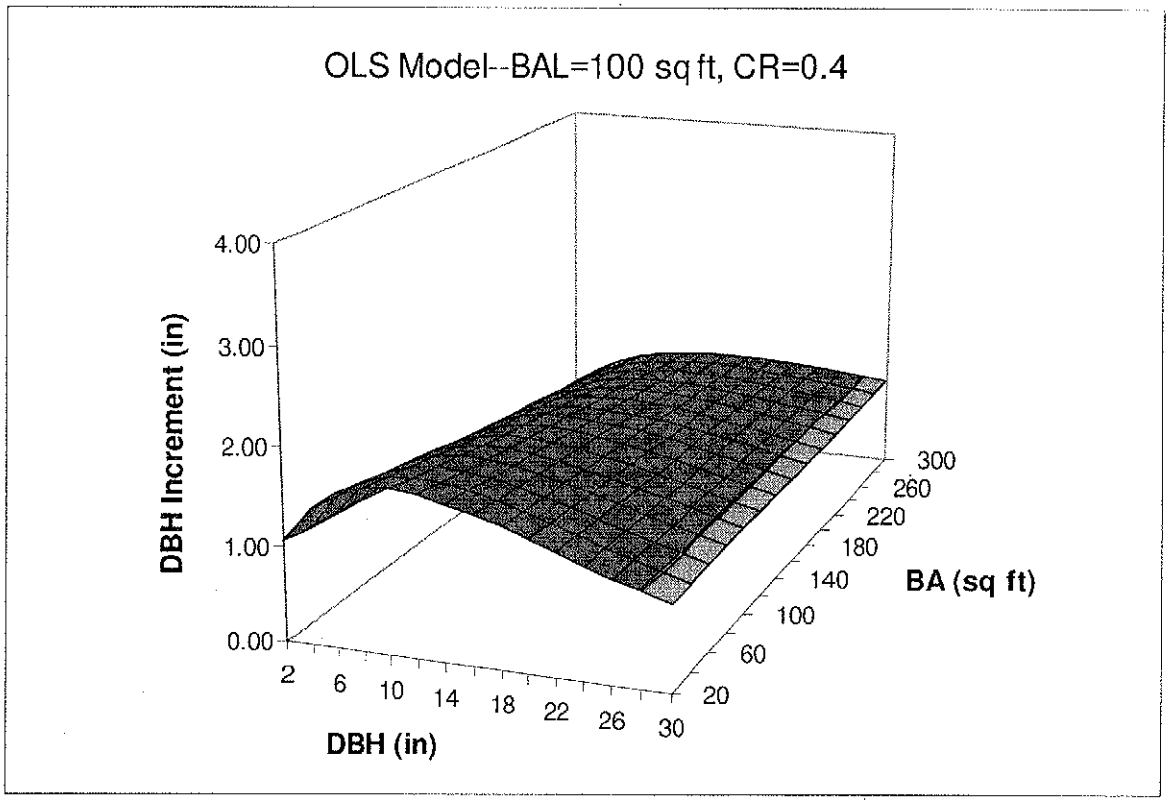


Figure 3.—*Differences in response surface between ordinary least squares model (OLS) and structural model (SM)—varying BAL and BA.*



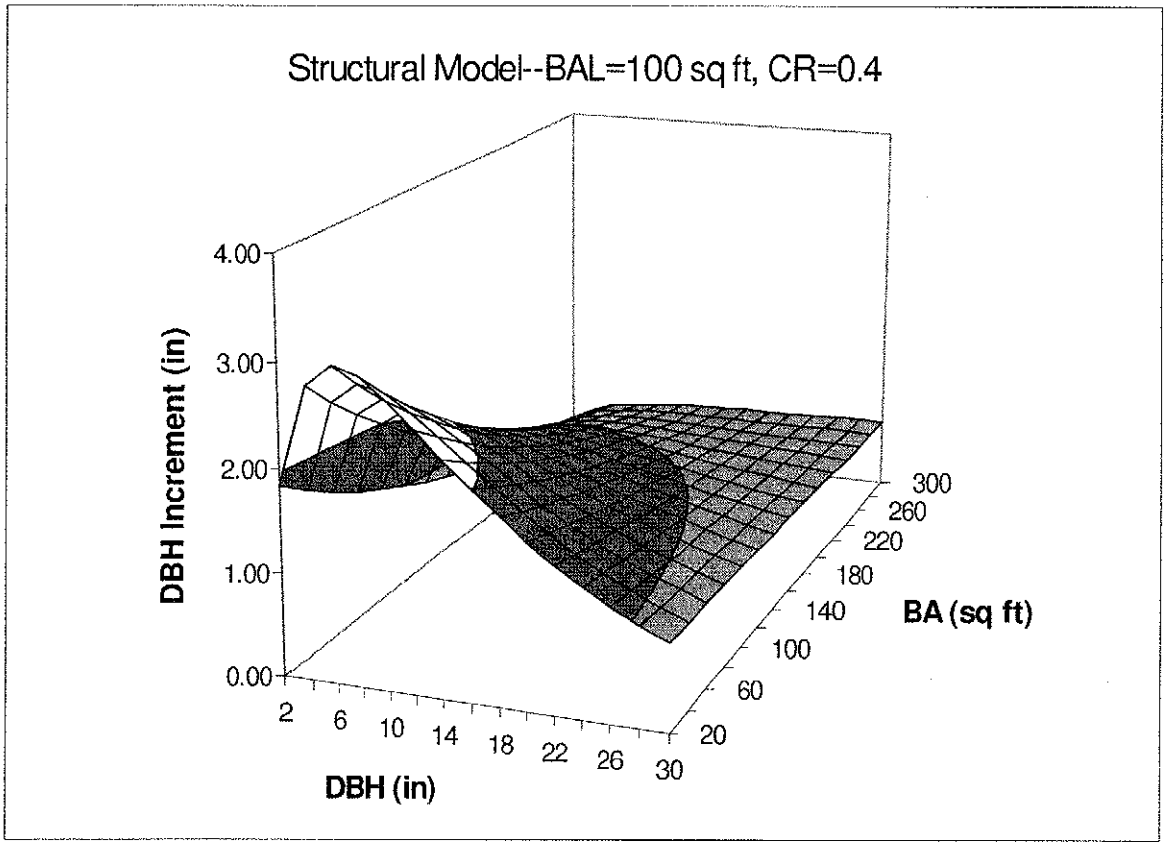
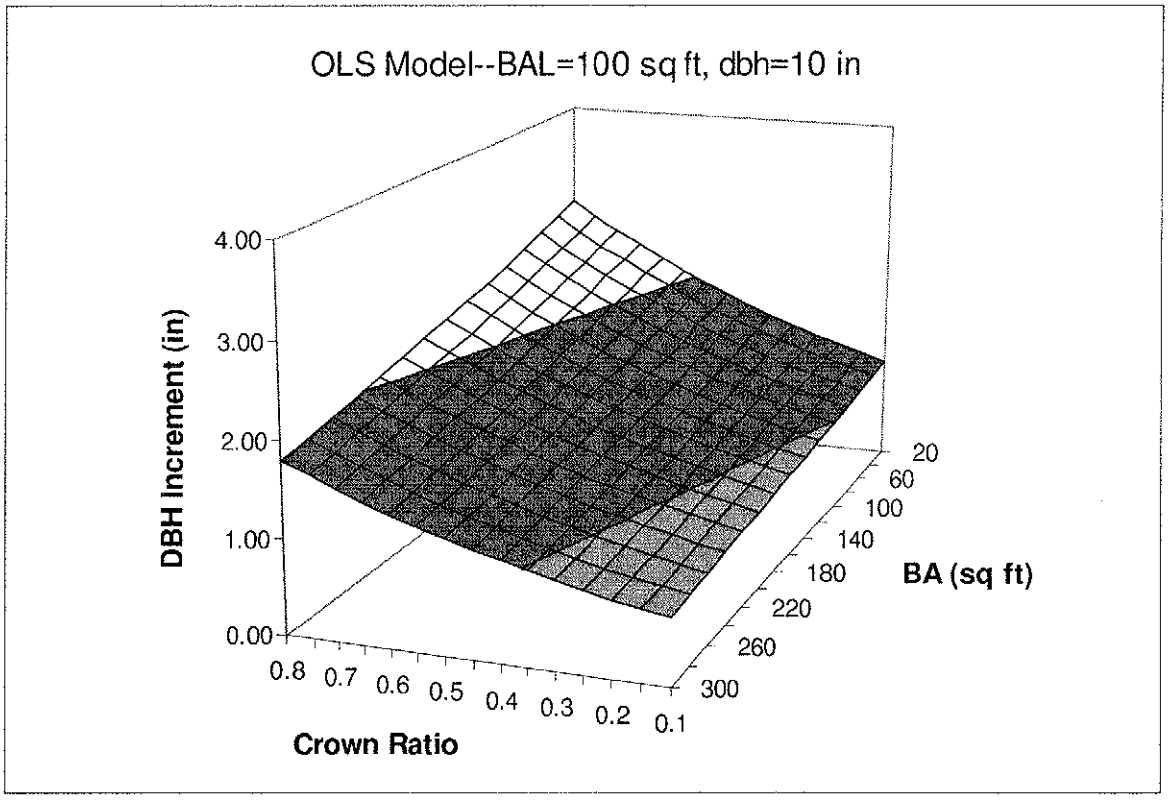


Figure 4.—*Differences in response surface between ordinary least squares model (OLS) and structural model (SM)—varying DBH and BA.*



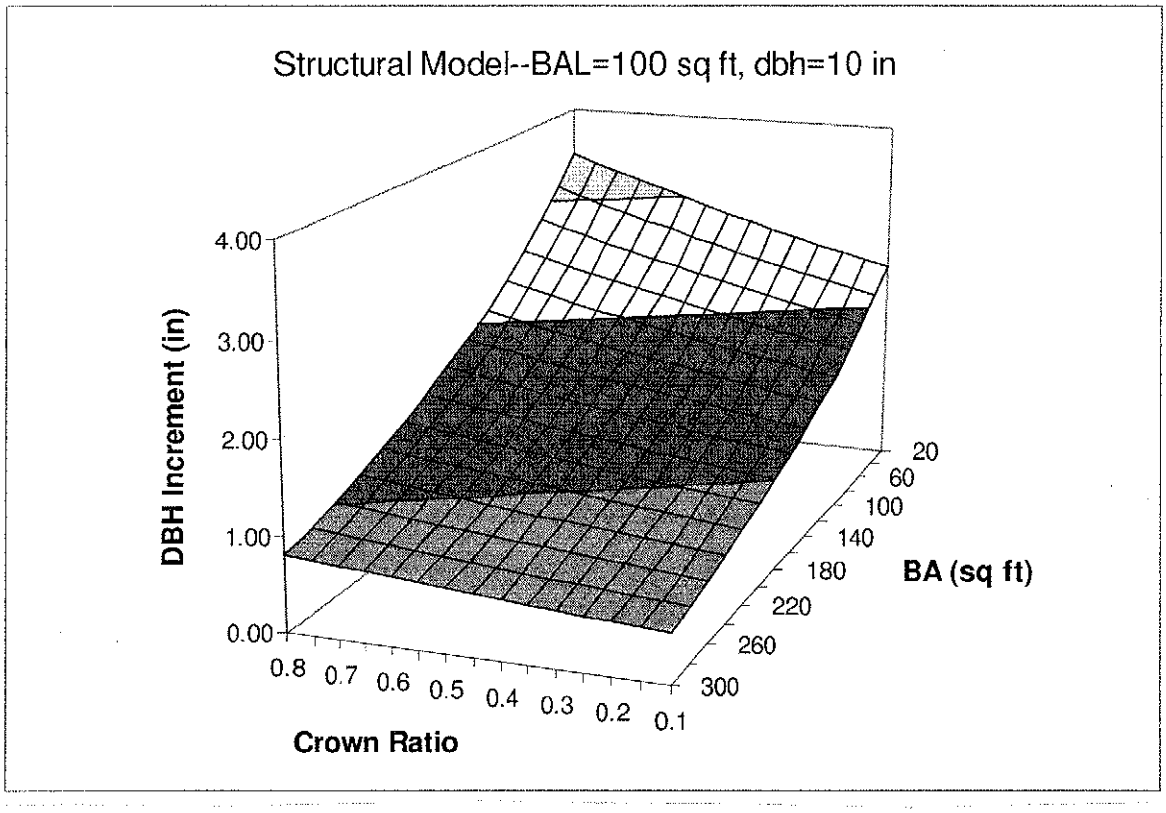


Figure 5.—*Differences in response surface between ordinary least squares model (OLS) and structural model (SM)—varying CR and BA.*

unthinned—are compared. The differences between the two inventory methods are the result of simulating the stand as if initially "inventoried" by 24 plots—each 1/300 acre (0.00135 ha), or as if inventoried by a single plot of 1 acre (2.47 ha.). The model calibration data were actually inventoried by variable-radius plots invariably larger than 1/300 acre (0.00135 ha).

For the more uniform stand left after thinning, projections are nearly coincident (fig. 6). It could be inventoried either way. However, the single-plot projections of the unthinned stand are almost parallel to the thinned projection, indicating little response to thinning. The 24-plot projection of the unthinned stand, by comparison, increases less rapidly, which indicates that there was an immediate response to the thinning.

IMPLICATIONS FOR INVENTORY DESIGN

This development in growth modeling raises new questions about optimal inventory design. Because we make explicit use of statistical effects of varying plot size as part of the increment estimation process, we can evaluate the tradeoff between optimizing the statistical precision and optimizing the resolution with which the biological processes can be represented. Now relations between subplot size and number and the diameters of trees can be varied to capture the biological structure of the stand, thereby sharpening the portrayal of future development. In turn, thinning algorithms used in simulations can base thinning decisions on local conditions in much the same way that a silviculturist would treat different parts of the same stand.

Sampling errors used in SBP for the two competition variables, BA and BAL, depend on size and number of the subplots. Size of plots can be varied, but they should be small enough that all trees in a single plot are close enough to interact through, say, no more than one intermediary. Representation of the biological structure of the stand can be improved by a design using plot sizes that increase with tree diameter—either smoothly as in "prism plots," or in several steps as in nested plots. Circular plots are preferred over rectangular because they have less edge. The BAL variable is favored in two respects by sampling with variable-sized plots. First,

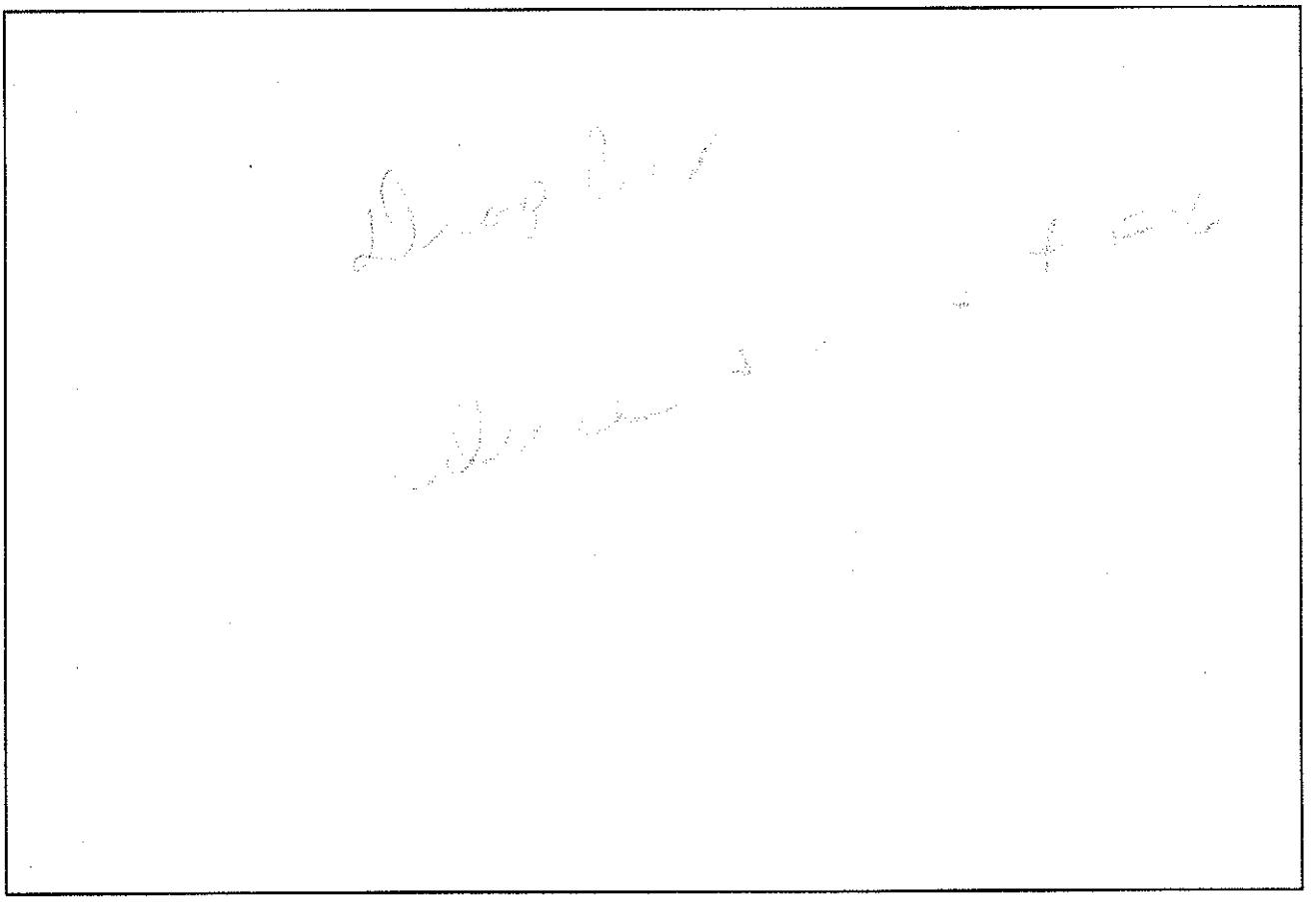


Figure 6.—*Differences in projections of a patchy stand attributable to the sampling design describing the stand to be projected. Two scenarios are compared: thinned vs. unthinned.*

trees larger than the subject tree are sampled on larger plots so that fewer of its competing trees would be "off plot." And second, efficiency with which the basal area component is estimated is increased in comparison to fixed-area plots.

Here to fore, the only use of variance among plots in a cluster or stand has been to optimize estimates of population means or totals. While at least two subplots per location provide an estimate of sampling variance of mean BA, the variance of the variance is high for two to three plots per location. For example, in an infinite population, the variance of the sample estimate of the variance is:

$$V(s^2)=\frac{2\,\sigma^4}{n-1}\left[1+\frac{n-1}{2n}G_2\right]$$

where G2 is the Fisher measure of kurtosis.
Note, however, that the biases in the predictions with SBP now depend on variances of variances, rather than on just the variance as with OLS.

CONCLUSIONS

In this paper, we addressed the question of how inventory design affects use of inventory data for analyzing alternative management scenarios. The false assumption that competition variables are known without sampling error was shown to have substantial effects on coefficients of models. The procedure we call Structural Based Prediction was demonstrated using a semi-distance-independent growth model as example. Implications of SBP for inventory design were discussed in the context of using inventory data and a growth model to project future stand attributes and to evaluate silvicultural alternatives.

ACKNOWLEDGMENTS

The following people reviewed this paper: Dr. Kelsey Milner, University of Montana, Missoula, MT, USA, and Dean Stuck, Potlatch Corp., Lewiston, ID, USA.

LITERATURE CITED

Elliot, W.J.; Hall, D.E. 1997. Water Erosion Prediction Project (WEPP) forest applications. Gen. Tech. Rep. INT-365. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 11 p.

Fuller, W.A. 1987. Measurement error models. New York, NY: John Wiley and Sons. 440 + xxiii p.

Goelz , J.C.G.; Burk, T.E. 1996. Measurement error causes bias in site index equations. Canadian Journal of Forest Research. 26: 1585-1593.

Jaakkola, S. 1967. On the use of variable size plots for increment research. In: Anonymous, ed. Vol VI, Proceedings XIV International Union of Forestry Research Organizations Congress Sect 25; Munich, Germany: 371-378

Kangas, A.S. 1998. Effect of errors-in-variables on coefficients of a growth model and on prediction of growth. Forest Ecology and Management. 102: 203-212.

Lappi, J. 1991. Estimating the distribution of a variable measured with error: stand densities in a forest inventory. Canadian Journal of Forest Research. 21: 469-473.

Pauley, G.R.; Peek, J.M.; Zager, P. 1993. Predicting white-tailed deer habitat use in northern Idaho. Journal of Wildlife Management. 57: 904-913.

Schreuder, H.T.; Williams, M.S. 1995. Assessing measures of tree competition using fixed-area, variable-radius, and horizontal-line sampling. Canadian Journal of Forest Research. 25: 1578-1582.

Stage, A.R. 1977. Forest inventory data and construction of growth models. In: Schmid-Haas, P., ed. Inventories on successive occasions. Papers presented at the XVIth IUFRO World Congress in Oslo, June 20-July 2, 1976. Berichte Nr. 171, Eidgenössische Anstalt für das forstliche Versuchswesen CH8903 Birmensdorf.

Stage, A.R.; Wykoff, W.R. 1993. Procedure for representing spatial variability within distance-independent stand growth models: an application of measurement error models. In: Burkhart, *et al.*, eds. Modeling stand response to silvicultural practices. Proceedings of the IUFRO S4.01 conference; 1993 September 27-October 1; Blacksburg, VA. Pub. FWS-1-93. Blacksburg, VA: College of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University: 131-144.

Stage, A.R.; Wykoff, W.R. 1998. Adapting distance-independent forest growth models to represent spatial variability: effects of sampling design on model coefficients. Forest Science. 44(2): 224-238.

Wykoff, W.R.; Crookston, N.L.; Stage, A.R. 1982. User's guide to the stand prognosis model. Gen. Tech. Rep. INT-133. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 112 p.

Wykoff, W.R. 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. Forest Science. 36: 1077-1104.

Wykoff, W.R. 1997. Predicting basal area increment in irregular stands: a progress report. In: Teck, *et al.*, comps. Proceedings Forest vegetation simulator conference; 1997 February 3-7; Fort Collins, CO. Gen. Tech Rep. INT-373. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station: 199-205.