APPLICATION OF A BAYESIAN TECHNIQUE FOR INCREASING THE PRECISION OF GROWTH PROJECTION ESTIMATES

George Z. Gertner

Department of Forestry
110 Mumford Hall
University of Illinois
Champaign, Illinois, 61801
U.S.A.

ABSTRACT

A procedure is presented for merging in a Bayesian framework growth projection estimates from a nonlinear diameter growth model with monitored growth from survey plots. Essentially, the procedure is a feedback system which provides a means for increasing the precision of growth estimates. The procedure was implemented and evaluated using growth data from the Pacific Northwest, U.S.A.

Keywords: Bayesian, feedback, growth projection system, sequential estimation.

INTRODUCTION

Contemporary forest growth projection systems for analyzing and updating forest survey plots (i. e., Lanford and Cunia 1977, and Hahn et al. 1979) are often developed for large geographic regions. These forest growth projection systems vary in their construction, but in general they consist of a series of models for projecting, on either an individual tree or stand basis, the forest dynamics of growth, mortality, regeneration, and removal. Often these models are highly nonlinear. In the development of these models, the steps of model specification, parameter estimation, verification, and revision are usually based on data collected over a large geographic region for a number of years. Developed in this way, a forest growth projection system will only furnish broad-scale regional estimates, and will not necessarily provide adequate estimates for sub-regions within the overall region, i. e., counties, forest districts, stands, etc. The primary reason for this is that at the current state of development, regional growth models do not fully account for subregional site quality and stocking variability, genotypic variability, interaction between trees, local climatic fluctua-tions, etc. (Turnbull 1977). For regional estimates, these unexplained factors will usually average out, but for subregional estimates, they may not, hence resulting in subregional estimates that may be biased.

Operational fall-down is also another important source of unexplained variability (Bruce 1977) which can lead to bias projections. Many growth project systems have been calibrated with data acquired from research plots that are located in uniform stands that are undamaged and of very high quality. When these systems are used to project stands that are not maintained under the same optimal conditions as reseach plots, predicted growth is commonly found to be higher than observed growth. To compensate for operational fall-down, often rudimentary approaches are used to adjust the growth projection estimates so they will not over-predict growth.

In time, region growth models will be formulated with the capabilities to explicitly account for these different factors. In the meantime, a means for implicitly accounting for these sources of unexplained variability is to monitor the different forest components in a subregion and then adjust or localize the regional estimates to the subregion.

There are a number of approaches that can be used to adjust projection estimates (i.e. Stage 1973, and Smith 1981). One possible approach is to use a Bayesian procedure which provides a feedback function between the model and the forest. A feedback function is a function which tends to keep a certain relationship between a predicted attribute and an observed attribute. By using the difference between growth projection estimates and observed growth dynamics as a control, monitored deviations can be fed back into the system such that the difference approaches zero by appropriately adjusting the regional parameters of a growth projection model. With the Bayesian procedure, the degree of adjustment of the regional parameters will depend not only on the difference between the predicted attribute and observed attribute from the subregion, but also on both the relative quality and quality of new information from the subregion in comparison to the prior information from the overall region. As would be intutitive, if the difference between the observed and predicted attribute is large, the adjustment will be large only if the number of observations from the survey of the subregion is relatively large in comparison to the number of observations used to parameterized the regional model, and/or if the variance of the estimated attribute based on the survey of the subregion is relatively small in comparison to the variance of the predicted attribute from the regional model. Even if the difference is large, the adjustment will be relatively small if the number of observations from the survey is relatively small, and/or if the variance of estimated attribute based on the survey is relatively large.

Conceptionally, there is a very appealing aspect to using Bayesian methods for analyzing and updating forest resource information. In a well-planned forest survey, one objective, that is often aimed for, is to minimize the relative cost per unit of information. Information about the growth and yield has been obtained for the most part from very costly on-the-ground-survey plots. Amassed in a forest growth projection system is a large amount of past or prior information which is relatively inexpensive to access. The Bayesian procedure provides a statistical foundation for combining prior estimates from a forest growth projection system with estimates obtained from survey plots. It is possible to demonstrate that the combined estimates, usually referred to as posterior estimates, will be of equal or greater precision than the estimates based solely on survey data (Meditch 1969). With the increase in precision of the estimates, the number of survey plots needed to be measured and maintained can be decreased if the objective is to minimize cost for specified level of precision. As more information is incorporated into a forest growth projection system through model refinement and data base enlargement, the tracting ability of the growth projection system should continually improve. The need for survey plots should gradually lessen, for there should be a progressive decrease in extractable new information that can be obtained from the survey plots.

METHODOLOGY

Consider the case where it is of interest to adjust a linear model for only one time period. Suppose for the linear case regional observations Y_{01}, $Y_{02}, \ldots, Y_{0m}$ were postulated as having arisen from a model such

that $Y_0 = X_0 B_0 + \varepsilon_0$ (1)

where $Y_0' = (Y_{01}, Y_{02}, \ldots Y_{0m})$ is a vector of m regional observations,

$B_0' = (B_{01}, B_{02}, \ldots, B_{0u}, \ldots, B_{0q})$ is a vector of q regional parameters, u = 1, 2, q,

$X_0(mxq)$ is a matrix of independent variables,

$\varepsilon_0' = (\varepsilon_{01}, \varepsilon_{02}, \ldots, \varepsilon_{0m})$ vector of errors.

If ordinary least squares is used to estimate the regional parameters B_0, the standard assumption usually made are

$\varepsilon_0 \sim N\ (0, I\sigma^2)$ $B_0 \sim N\ (\beta, V)$

$COV\ (B_0, \varepsilon_0) = 0$

where I(mxm) is an identity matrix, σ^2 is the constant variance of the ε_0, V(qxq) is the covariance of B_0, β is the expected value of B_0, and both B_0 and ε_0 are normally distributed.

Suppose after estimating the regional model a random sample is taken from the subregion, $Y_1' = [Y_{11}, Y_{12}, \quad Y_{1m}]$, where the number of observations, m, from the subregion need not be the same as from the overall region. If the regional parameter estimates can be assumed to be random, then Baye's theorem can be used to combine prior parameter estimates from the region with estimates from the subregion. Baye's theorem is of the form

$$f(B_0/Y_1) = \frac{f(Y_1/B_0)f(B_0)}{f(B_0)} = f(Y_1/B_0)f(B_0)$$

where $f(B_0)$ is the prior probability density function of random parameter B_0 obtained from fitting the growth function over the entire region,

$f(Y_1/B_0)$ is the conditional probability density function of observations taken from the subregion given the parameter B_0,

$f(B_0/Y_1)$ is the posterior distribution of parameter B_0 given Y_1. This function contains information from the entire region $f(B_0)$, as well as subregion $f(Y_1/B_0)$.

If measurement error of the new information has a mean of zero and a constant variance, and is additive, independent and normally distributed, then the adjusted or posterior estimate b_1 can be obtained by maximizing $f(B_0/Y_1)$ by taking the derivative of $f(B_0/Y_1)$ in respect to B_0. The resulting equation for obtaining the posterior estimate is then

$$b_1 = B_0 + P_1 X_1'(Y_1 - X_1 B_0)\sigma_1^{-2} \quad (2a)$$

where $P_1 = [((X_1 X_1)^{-1}\sigma_1^2)^{-1} + P_0^{-1}]^{-1}$ (2b)

and b_1 is a posterior estimate of B_0,

P_0 is the estimate of the covariance B_0,

P_1 is the estimate of the covariance of B_1, and

X_1(mxp) is a matrix of independent variables from the subregion.

In practice, b_0, the regional estimate is used in the place of B_0. The revised parameter values can be substituted back into the original model, Eq. 1, to obtained adjusted predictions.

The size of the adjustment of the regional parameter estimate will depend on the size of 1) the residual or difference between the subregional observations and predicted observations based on the regional parameter estimates, $e = (Y_1 - X_1 B_0)$, 2) the covariance of the estimate of b_1 based solely on the subregional data, $(X_1'X_1)^{-1}\sigma_1^2$, and 3) the covariance of B_0, P_0. The adjustment will be large if e is large and the relative size of $(X_1'X_1)^{-1}\sigma_1^2$ is large in comparison to P_0. The magnitude of P_0 will depend roughly on the size of the sample used to parameterize the regional model and the predictive power of the model based on the regional parameter estimates. P_0 will usually be large when sample size is small or the quality of the model is poor. Similarly, the same holds true for $(X_1'X_1)^{-1}\sigma_1^2$. The size of the covariance will depend roughly on the size of the sample from the subregion and the fit of the model if only the subregional data is to be used to parameterize the model. Even if e is large, the adjustment will be small if $(X_1'X_1)^{-1}\sigma_1^2$ is large in comparison to P_0. This is because little weight is given to the subregional data because of the lack of quantity or quality of the new information relative to the old information.

In its present form, Eq. 2 can be only used to adjust the parameters of a linear model for only one period in time. Eq. 2 can be easily extended to adjust sequentially for one or more time periods the parameters of a nonlinear model (Gertner 1982). In short, this can be done by linearizing the non-linear model with a Taylor series and generalizing Eq. 2 such that at time t, prior information at time t-1 would be information from the overall region as well as from the subregion in past periods. The sequential equation can be expressed as

$$b_t=b_{t-1} + P_t X_t'(Y_t-X_t b_{t-1})\sigma_t^{-2}$$

where $P_t=[(X_t'X_t)^{-1}\sigma_t^2)^{-1}+P_{t-1}^{-1}]^{-1}$

and t is a subscript for time.

APPLICATION

In a recently completed study, the Bayesian procedure was employed to sequentically adjust over four time periods the regional parameters of the nonlinear diameter increment model used in the Stand and Tree Evaluation and Modeling System - STEMS (USDA - FS, 1979). STEMS is a distance independent individual tree based growth projection system calibrated for the North Central and Northwest regions of the U.S.A. The diameter increment model used for the study was of the form:

$$ADG =[b_1(1-e^{b_2 CR})e^{b_3 DBH}][b_4 SI][1 \quad (e^{-b_5 N^{b_6} DBH^{b_7} BA^{-b_8}})]$$

where

ADG = annual diameter growth,

b_1, $b_2,\ldots,b_8$ = regional parameter estimates,

BA = stand basal area,

CR = crown ratio (crown length/total height),

DBH = diameter at bresent height,

N = number of trees per unit area,

SI = site index

The model consists of three components. Enclosed within the first brackets after the equals sign is the potential diameter function, within the second brackets is the site reduction function, and within the third brackets is the competition reduction function.

The regional parameters used for the diameter increment model were for unmanaged second growth Douglas-fir (Pseudotsuga menziess [Mirb.] Franco) located in western Oregon (Shifley and Fairweather 1981). A total of 2762 growth observations were used to calibrate the regional model. Data to localize the model came from a permanent plot measured at two-year intervals for ten years, resulting in four growth measurements for each tree[1].

Figure 1 shows the movement of the parameters for each time period. The regional parameter estimates correspond to period 0. The parameters of the potential and the site reduction function, b_1 through b_4, show relatively little movement, while the parameters of the competition reduction function, b_5 through b_8, changed significantly. Notice that each of the last four parameters tends to reach an asymptote, indicating that additional observations in time would not provide any more information about the stand. Presented in Figure 2 are the residual mean and residual mean square for the unadjusted and adjusted models. The residual means for the unadjusted model were all negative for each of the periods, indicating that the unadjusted model was over estimating diameter growth, while the residual means of the adjusted model were all close to zero for each period, indicating little bias. By using the sequential method, the residual mean squares were nearly halved in comparison to the unadjusted estimates.

CONCLUSION

The Bayesian procedure is not only a means for localizing a regional model, but also is a way to incorporate information from a growth projection model with conventional inventory information. A large amount of historical information is incorporated within a forest growth projection system. This information is relative inexpensive in comparison to information obtained from conventional inventories. By merging growth projection estimates with conventional inventory estimates, it is possible to obtain estimates that are more precise than if estimates are based solely

---

[1]The data was kindly provided by the Regional Forest Nutrition Project which is sponsored by the College of Forest Resources, University of Washington.

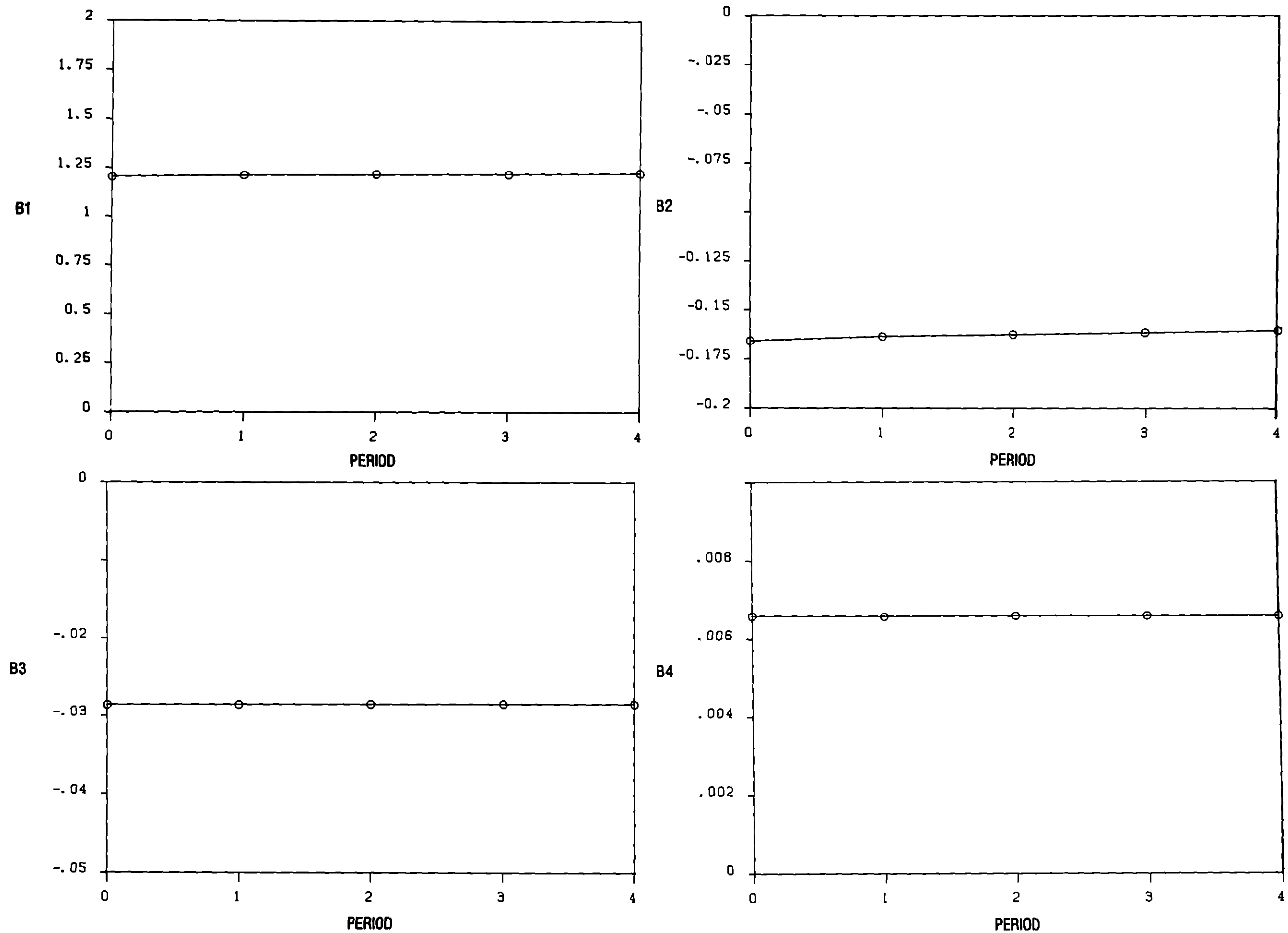


Figure 1. Adjusted parameter estimates for periods 1 to 4.

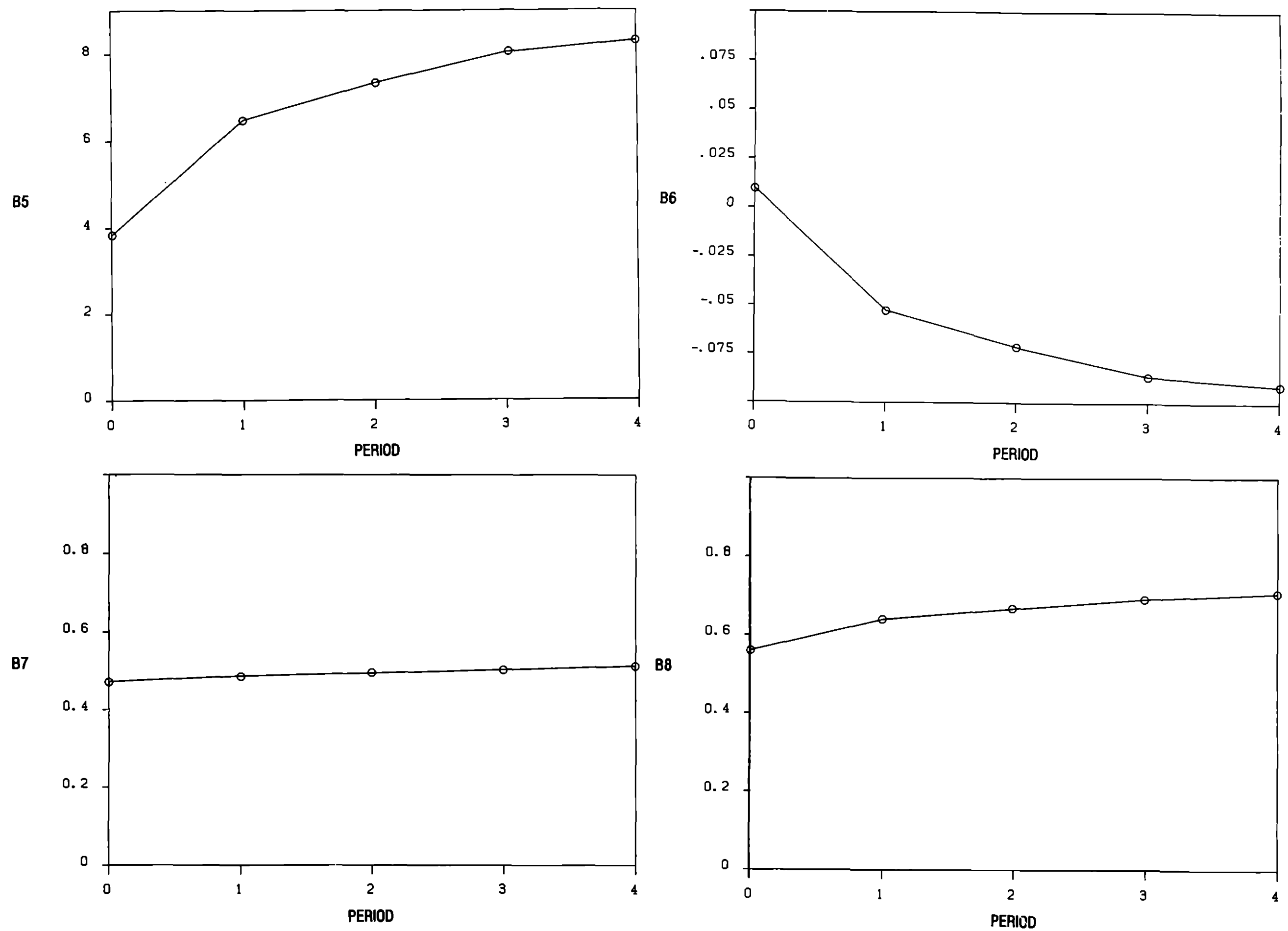
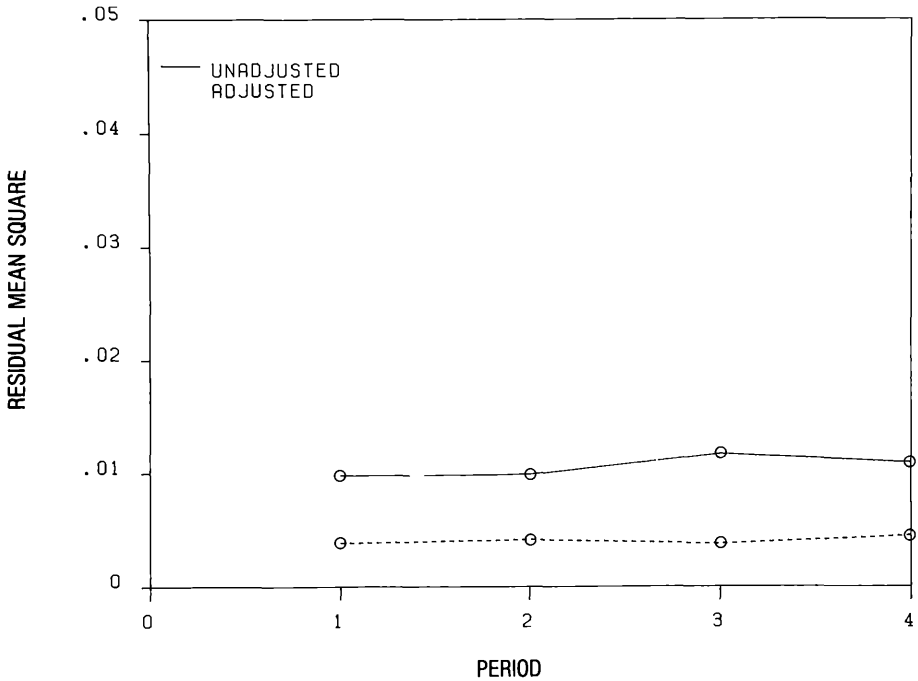


Figure 1. Continued.



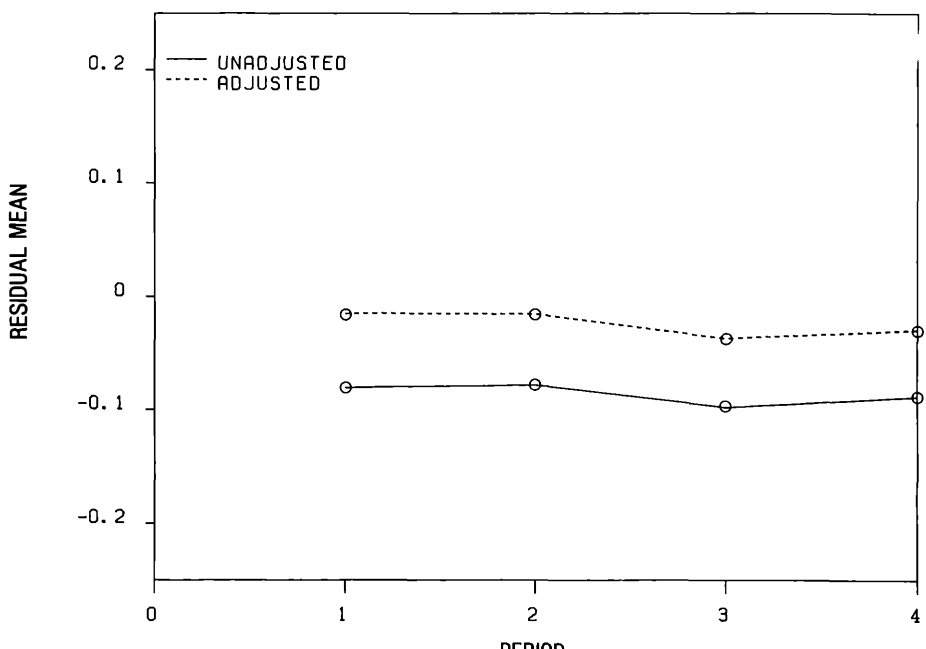


Figure 2. Residual means and residual mean squares of the unadjusted and adjusted model for each period.

on growth projections or conventional inventories. Essentially, the use of the Bayesian procedure in conjunction with a computer growth projection system can be considered to be the first stage in the development of a system that can learn through experience.

BIBLIOGRAPHY

Beck, J. and K. Arnold. 1977. Parameter Estimation in Engineering and Science. John Wiley and Sons, Inc., N.Y. 501 pp.

Bruce, D. 1977. Yield differences between research plots and managed forests. Jour. of Forestry, p. 14-22.

Gertner, G. 1982. Localizing a diameter increment model with a sequential Bayesian procedure. Submitted to Forest Science.

Hahn, J. T., D. M. Belcher, M. R. Holdaway, G. J. Brand, and S. R. Shifley. 1979. FREP78. The updated tree growth projection system. In: Forest Resource Inventories Workshop Proceedings. Colorado State University 1: 211-222.

Lanford, B. L., and T. Cunia. 1977. Stand projection by simulation using CFI data. Swiss Federal Institute of Forestry Research, Report 171, p. 29-52.

Meditch, J. S. 1969. Stochastic Optimal Linear Estimation and Control. McGraw-Hill Book Co., New York. 394 pp.

Mendel, J. 1973. Discrete Techniques of Parameter Estimation. Marcel Dekker, Inc., New York. Vl, 385 pp.

Shifley, S., and S. Fairweather. 1981. Adaption of the FREP growth projection system to the forest species and cover types of the Pacific Northwest. Report submitted for publication to Canadian Jour. of Forest Research.

Smith, W. B. 1981. Adjusting a regional forest growth model to improve local predictive power. Report submitted for publication to Canadian Jour. of Forest Research.

Stage, A. 1973. Prognosis model for stand development. USFS Research Paper INT-137, 32 pp.

Turnbull, K. J. 1977. Long-term yield forecasting models; validation and iterative estimation. In: Growth Models for Long-term Timber Yield. Fries et al., Virginia Polytechnic and State University, FWS 1-78.

USDA - FS. 1979. The user's guide to STEMS stand and tree evaluation and modeling system. USDA Forest Service North Central Forest Experiment Station Ten. Tech. Rep. NC-70.