CONTINUOUS INVENTORIES AND THE COMPONENTS OF CHANGE

FRANCIS A. ROESCH

USDA Forest Service

Southern Research Station, P.O. Box 2680, Asheville, NC, 28802-2680, USA

ABSTRACT. The consequences of conducting a continuous inventory that utilizes measurements on overlapping temporal intervals of varying length on compatible estimation systems for the components of growth are explored. The time interpenetrating sample design of the USDA Forest Service Forest Inventory and Analysis Program is used as an example. I show why estimation of the traditional components of growth, as presented by Meyer (1953, *Forest Mensuration.*), is less useful than it was in previous inventory designs and give a discrete analog to the time invariant redefinition of the components of change that were first given by Eriksson (1995, *For. Sci.* 41(4):796-822).

KEYWORDS: forest inventory, FIA, sample design, estimation, forest growth

1 INTRODUCTION

Traditionally, the components of forest growth have been expressed as:

$$V_2 - V_1 = S + I - M - C, \tag{1}$$

where:

V_i = the total value at time i, i=1,2,
S = survivor growth,
I = ingrowth,
M = mortality, and
C = cut.

These components of growth have commanded a long-standing interest in forest inventories. Equation (1) was presented by Meyer (1953), and used in related contexts by Beers and Miller (1964), Martin (1982), Roesch (1988), and Roesch et al. (1989), to name a few. In this paper, I will show why estimation of the traditional components of growth is less useful than it was in previous inventory designs and revisit the time invariant redefinition of the components of change that were first given by Eriksson (1995). Subsequently, I will propose a discrete analog of Eriksson's components of change for use by the USDA Forest Service's Forest Inventory and Analysis program (FIA) and suggest options for estimation.

In equation (1), the growth components describe changes between *measurement* times 1 and 2, and therein lies the problem with the traditional definitions. Times 1 and 2 refer explicitly to measurement periods, rendering the components dependent upon the time that the sample is executed. These components implicitly assume that these two points in time are of special interest. This is seldom the case. What is truly of interest is the continuous monitoring of the population of trees as its members (1) enter into merchantability categories, (2) progress through those categories and (3) exit from each category. Given this, we must conclude that the weakness of the traditional definitions of the components of growth lies in their inherent dependence on the time of observation and the length of the measurement interval. By these definitions, these components are not strictly population parameters to be estimated. Rather, they are a convenient marriage of a population and the sample design used to observe the population. Unfortunately, the marriage only remains convenient as long as the sample design doesn't change in any significant way.

An inseparable problem concerns a lack of clarity in what is actually being estimated by "average annual growth". For the historic periodic inventories of FIA, the average annual growth was estimated over a relatively long remeasurement interval and the growth was "annualized" by simply dividing by the length of the measurement interval for each plot, which usually varied between 7 and 12 years. These estimates of average annual growth, calculated over different growth intervals, are actually measures of different population parameters in each case, and are not readily comparable. To demonstrate this, let's look from the population perspective, and consider a single condition class. For discussion purposes, I will use a "typical" yield curve for cubic foot volume, specifically the one given in Avery and Burkhart (1983, p277):

$$Y = e^{10-32A^{-1}} \tag{2}$$

where A= age of the stand. The graph for equation (2) is given in Figure 1.

When we calculate the average annual cubic foot accretion from a 1-year lag, a 2-year lag, a 5-year lag, a 10-year lag, and a 20-year lag, using the model in equation (2), we obtain the five curves in Figure 2. If our sample design had a 2-year measurement interval (a 2-year lag), we would be selecting our estimates of average annual cubic foot accretion from the curve weighted the most toward the actual annual cubic foot accretion curve (the 1-year lag curve). If we used a 5-year measurement interval, we would be selecting our annual cubic foot accretion estimates from the curve weighted more toward the right. These curves are the distributions from which we sample growth given a particular measurement interval. This, coupled with the fact that we sample many forest types concurrently, all following different initial growth curves, shows us that the usual indifference assumption with respect to measurement interval length is ill advised. Estimators should estimate population attributes and not be subject to the whims of a sample design.

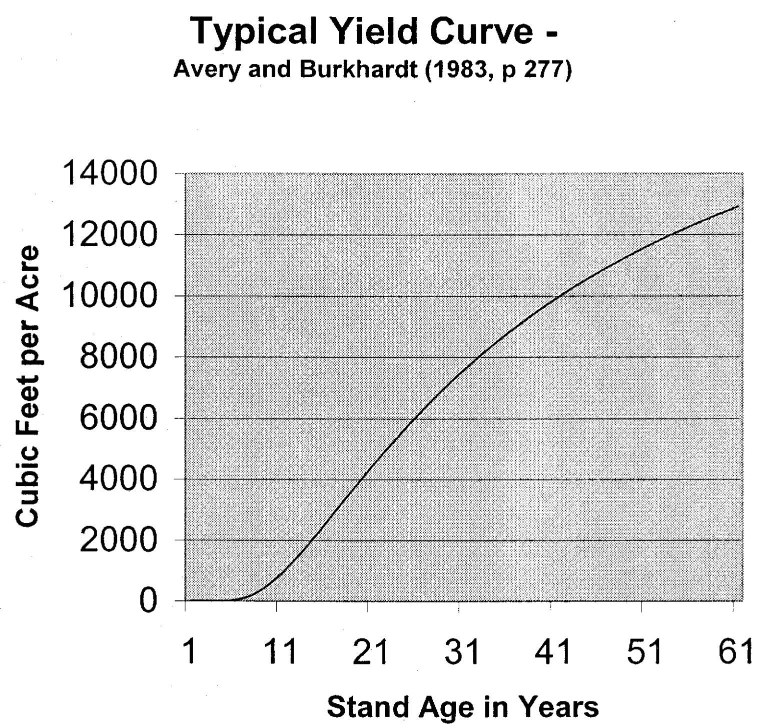

Figure 1: The graph of a "typical" yield curve ($Y = e^{10-32A^{-1}}$) for cubic foot volume, given in Avery and Burkhart (1983, p277).

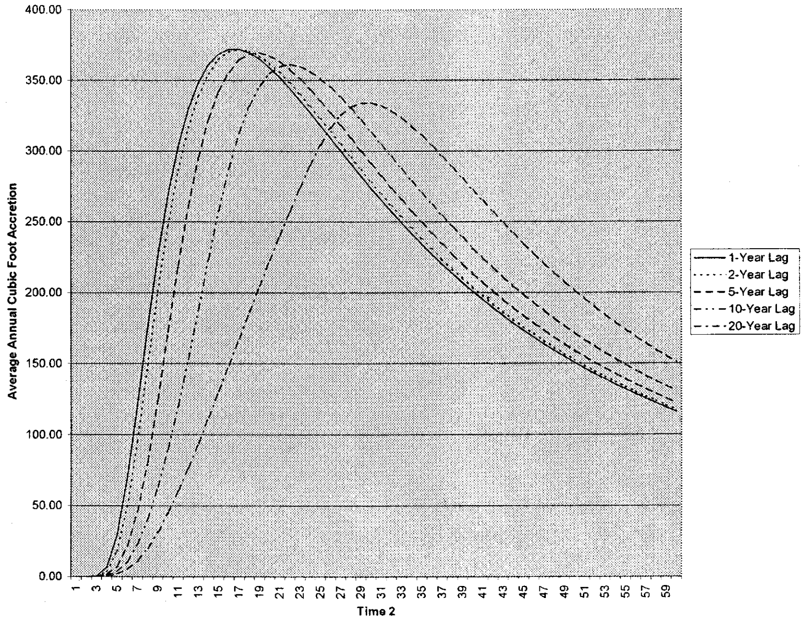

Figure 2: The average annual cubic foot accretion from a 1-year lag, a 2-year lag, a 5-year lag, a 10-year lag, and a 20-year lag, using the model in Figure 1.

2 THE COMPONENTS OF CHANGE

2.1 Definitions. Eriksson (1995) recommended a new set of definitions and labeled them the components of change, applicable over a temporal continuum as opposed to the traditional sample-based definitions, defined over discrete periods. Eriksson took the position that the traditional growth component definitions do not allow estimators that are time-additive over multiple period lengths. For example, if a ten-year period were of interest and all plots were measured in years 0, 5, and 10, the sums of the expected values of the estimators of each component over the two intervals (years 0-5 and 5-10) would not equal the expected value of the estimators sans the year 5 measurement. Non-additivity is a valid concern due to a fundamental flaw in the original definition of the components of growth. Additionally, the redefinitions become extremely compelling in the realm of annual inventories and at least suggest that the original definitions are inadequate for time-interpenetrating sample designs, such as the FIA's rotating panel design.

The traditional component of ingrowth consists of (1) the value of the ingrowth trees when they reach the minimum merchantability limit and (2) the value growth subsequent to attaining the minimum merchantability limit. Eriksson (1995) identified this later component as a component of survivor growth, using the argument that the tree is a survivor once it has passed the minimum merchantability threshold, regardless of whether or not the graduation was observed. Paraphrasing the work in Eriksson (1995), live tree growth is the growth in value that occurs on trees after the minimum merchantability limit has been achieved. Entry is the value of trees as they attain the minimum merchantability limit. Mortality is the value of trees as they die, while Cut is the value of trees as they are harvested. These definitions were purposefully continuous, however, Roesch (in review) used the discrete analog with a small (1 year) interval length. This was done under the assumption that 1 year is about the minimum interval length, in most forest conditions, required for the growth signal to overpower measurement error, as well as to allow some reasonable temporal partitioning of the observations. Additionally, the discrete intervals allow the definition of a set of indicator matrices, one for each component, having one row for each tree in the population during the forest inventory. For example, the indicator matrix for the entry component:

$$\begin{array}{c c}
\text{time} = \;\; \text{P} \;\; \text{P-1} \;\; \cdots \;\; 1 & \text{tree} \\
\mathbf{I_E} = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \ddots & 0 \\ 0 & 1 & \ddots & 0 \\ 1 & 0 & \cdots & 0 \end{bmatrix} & \begin{array}{c} 1 \\ \vdots \\ \vdots \\ \text{N} \end{array}
\end{array}.$$

In $\mathbf{I_E}$, the columns represent the year of entry, the first column (P=Present) being the most recent year of the inventory, and each successive column one year prior to the previous column. Each row corresponds to an individual tree in the population. Analogous indicator matrices for tree mortality, $\mathbf{I_M}$, and tree harvest, $\mathbf{I_C}$, are similarly defined. The indicator matrix for the live category contains a 1 for each year that a tree is alive subsequent to its entry year and prior to its year of harvest or death, and a 0 otherwise:

$$\begin{array}{c c}
\text{time} = \;\; \text{P} \;\; \text{P-1} \;\; \cdots \;\; 1 & \text{tree} \\
\mathbf{I_L} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & 1 \\ 0 & 1 & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix} & \begin{array}{c} 1 \\ \vdots \\ \vdots \\ \text{N} \end{array}
\end{array}$$

The four indicator matrices are of equal dimension and sum to the population indicator matrix, $\mathbf{I_P} = \mathbf{I_E} + \mathbf{I_L} + \mathbf{I_M} + \mathbf{I_C}$. Likewise, in the true value matrix, row i represents tree i, and column j represents time j in reverse annual order from the present year back:

$$\mathbf{v} = \begin{bmatrix} v_{1,P} & v_{1,P-1} & \cdots & v_{1,1} \\ v_{2,P} & v_{2,P-1} & & \\ \vdots & & \ddots & \vdots \\ & & & \\ v_{N,P} & & \vdots & v_{N,1} \end{bmatrix}, \text{ while the entry value vector is: } \mathbf{v^E} = \begin{bmatrix} v_1^E \\ v_2^E \\ \vdots \\ \vdots \\ v_N^E \end{bmatrix}$$

Usually a specific period of length (t) beginning in year h will be of interest. Define a column vector $\mathbf{Y_{h+t,h}}$ in which rows represent time in reverse annual order. A row contains a 1 for a year of interest and a 0 otherwise.

Additionally, we define the first difference matrix with P-1 columns and P rows, such that the ordered pair $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ appears once and only once in each column and all other entries are zero. The ordered pair occupies the first two positions in the first column, and moves one position down in each subsequent column:

$$\mathbf{d}^1 = \begin{array}{cc} & \begin{array}{ccccc} P\text{-}1 & P\text{-}2 & \cdots & \cdots & 1 \end{array} \\ & \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 \\ 0 & -1 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix} \begin{array}{c} P \\ P\text{-}1 \\ \vdots \\ \vdots \\ 1 \end{array} \end{array}$$

The change components to be estimated can now be represented as:

Entry: $\mathbf{E}_{\mathbf{h,h+t}} = \left(\mathbf{I}_{\mathrm{E}}{}'\mathbf{v}^{\mathrm{E}}\right)' \mathbf{Y}_{\mathbf{h+t,h+1}}$,

Live growth: $\mathbf{L}_{\mathbf{h,h+t}} = \left(\mathbf{I}_{\mathrm{L}}{}'\left(\mathbf{vd}^1\right)\right)' \mathbf{Y}_{\mathbf{h+t,h+1}} + \left[\mathbf{I}_{\mathrm{E}}{}'\left(\mathbf{v} - \mathbf{v}^{\mathrm{E}}\right)\right]' \mathbf{Y}_{\mathbf{h+t,h+1}}$,

Merchantable Mortality: $\mathbf{M}_{\mathbf{h,h+t}} = \left(\mathbf{I}_{\mathrm{M}}{}'\mathbf{v}\right)' \mathbf{Y}_{\mathbf{h+t-1,h}}$, and

Merchantable Cut: $\mathbf{C}_{\mathbf{h,h+t}} = \left(\mathbf{I}_{\mathrm{C}}{}'\mathbf{v}\right)' \mathbf{Y}_{\mathbf{h+t-1,h}}$.

We also estimate Merchantable Volume at times h+t to h: $\mathbf{V}_{\mathbf{h+t,h}} = \left(\mathbf{I}_{\mathbf{P}}{}'\mathbf{v}\right)' \mathbf{Y}_{\mathbf{h+t,h}}$.

2.2 Exponentially Weighted Difference (EWD) Estimator. There are many potential schemes to combine the information from temporally overlapping panels in order to gain strength for the interval estimates. One possible strategy that I've dubbed an Exponentially Weighted Difference (EWD) estimator (Roesch, in review), is similar in concept to the exponentially weighted moving average (EWMA) estimator common in the quality control literature (i.e. Chandra, 2000) and the econometrics literature (i.e. West and Harrison, 1989, p. 55). In the EWD estimator, a series of differences between a measurement and the one preceding it for all successive measurements of the same panel are calculated. The difference is attributed to the center of the time interval and combined with the m-1 adjoining interval differences. The supporting panels are down-weighted exponentially with each step away from the central panel. For example, let:

$\bar{d}_{h,h+t} =$ the mean of a remeasured panel difference, such as $t^{-1}\hat{L}_{h,h+t}$

$r = (m-1)/2$, m is odd and is the number of remeasured panels used in the estimator, and

$$\alpha = \left(\frac{r}{r+1}\right).$$

The EWD estimator, applied to year k is:

$$EWD_k = \sum_{i=-r}^{r} \frac{(1-\alpha)}{\left(1+\alpha-2\alpha^{r+1}\right)} \alpha^{|i|} \bar{d}_{k-r+i,k+r+i} \, .$$

The EWD is more responsive to local variation than if equal weights were used. One (albeit inefficient) method of assuring compatibility would be to estimate the change components separately, and then use the sum across time t to estimate the value vector V.

2.3 Mixed Estimator. The components of change definitions are superior to the traditional components of growth for a number of reasons, the time additivity advantage pointed out by Eriksson (1995) being the most obvious. However, that does not necessarily translate into a stronger argument for forcing compatibility of the estimates of the components. Suppose that our sample design consists

of five overlapping continuously remeasured panels, one panel measured each year and then remeasured five years hence. The strongest signal for value during any particular year will come from the panel actually measured in that year, while the strongest signal for the live growth component will come from the panel with a remeasurement interval centered on that year. A mixed estimator would seem to be a good way to balance the desire for the "best" estimate for each component with a desire for compatible estimates. Like the EWD Estimator, the mixed estimator also draws strength from overlapping panels. It's a generalized least squares estimator in which model constraints are appended to the data matrices. Van Deusen (1996, 1999, 2000) showed mixed estimators for successive annual estimates. Roesch (2001) tested mixed estimators using both real and simulated data, finding the mixed estimators to perform quite well relative alternative techniques. Roesch (in review) argues for an approach that involves building compatibility constraints for the components of change into a mixed estimator. There it is noted that compatibility requires, for any i and any t, the strict equality, $\hat{V}_{i+t} - \hat{V}_i = \hat{L}_{i,i+t} + \hat{E}_{i,i+t} - \hat{M}_{i,i+t} - \hat{C}_{i,i+t}$, which leads to the variable constraint model:
$V_t - V_{t-1} - L_{t-1,t} - E_{t-1,t} + M_{t-1,t} + C_{t-1,t} = s_t; s_t$ iid $(0, p\sigma_t^2/n_t)$, $t = 2,\ldots T.$.

However, the FIA sample design (with k panels) is more conducive to an alternative model in which the observed midpoint values are used to constrain the component estimates. For k, an integer, and $t \geq k+1$, let

$$\delta_t^k = \begin{cases} \left(V_{t-(k-\frac{1}{2})} - V_{t-(k+\frac{1}{2})}\right), & \text{if } k \text{ is odd;} \\ \left(V_{t-(k-\frac{2}{2})} - V_{t-(k+\frac{2}{2})}\right), & \text{if } k \text{ is even;} \end{cases}$$

and form a four-column row vector of components such that:

$$\chi_t^k = \begin{cases} \dfrac{L_{t-k,t}}{k} \,\Big|\, \dfrac{E_{t-k,t}}{k} \,\Big|\, \dfrac{-M_{t-k,t}}{k} \,\Big|\, \dfrac{-C_{t-k,t}}{k}, & \text{if } k \text{ is odd;} \\ \dfrac{2L_{t-k,t}}{k} \,\Big|\, \dfrac{2E_{t-k,t}}{k} \,\Big|\, \dfrac{-2M_{t-k,t}}{k} \,\Big|\, \dfrac{-2C_{t-k,t}}{k}, & \text{if } k \text{ is even;} \end{cases}$$

Assume an observation model at each time t:

$$\delta_t^k = \boldsymbol{\chi}_t^k \boldsymbol{\beta}_t + e_t, \tag{3}$$

where $\boldsymbol{\beta}_t$ is a vector of coefficients with a row for each component and e_t is iid $(0, \sigma_t^2/m_t)$, and combine it with a reasonably constrained transition model. Roesch (in review) suggests:

$$\mathbf{1}\boldsymbol{\beta}_t - 2\mathbf{1}\boldsymbol{\beta}_{t-1} + \mathbf{1}\boldsymbol{\beta}_{t-2} = \varepsilon_t \; ; \tag{4}$$

where $\mathbf{1}$ is a 1x4 vector of ones and ε_t is iid $(0, p\sigma_t^2 / m_t)$.

Form a vector from the δ_t^k's, $\boldsymbol{\Delta} = \left[\delta_{k+1}^k, \ldots, \delta_T^k\right]'$, a matrix $\mathbf{X}$, from the χ_t^k's, having ((T-k)*4) columns. The vector $\boldsymbol{\chi}_i^k$ is placed in row i beginning in column ((i-1)*4)+1). The rest of the elements in the row are zeros. Concatenate successive elements of the column vectors $\boldsymbol{\beta}_t$ into the column vector

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_{k+1} \\ \vdots \\ \boldsymbol{\beta}_T \end{bmatrix},$$ having ((T-k)*4) rows. Form vectors from the error terms

$\mathbf{e} = \left[e_{k+1}, \ldots, e_T \right]'$ and $\boldsymbol{\varepsilon} = \left[\varepsilon_{k+1}, \ldots, \varepsilon_T \right]'$. Represent equation (3) with:

$$\boldsymbol{\Delta} = \chi\boldsymbol{\beta} + \mathbf{e} \tag{5}$$

Represent the covariance matrix of $\boldsymbol{\Delta}$ with $\boldsymbol{\Sigma}$. The temporal constraints can be re-expressed as:

$$\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\varepsilon} \tag{6}$$

where **R** is the appropriately sized matrix of constraints for the transition model.

Combining the observation model with the transition model

$$\begin{bmatrix} \boldsymbol{\Delta} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{R} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{e} \\ \boldsymbol{\varepsilon} \end{bmatrix} \tag{7}$$

Theil's mixed estimator (1963, 1971), is:

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \frac{1}{p}\mathbf{R}'\boldsymbol{\Omega}^{-1}\mathbf{R} \right]^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}$$

and the covariance matrix estimator:

$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}\right) = \left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \frac{1}{p}\mathbf{R}'\boldsymbol{\Omega}^{-1}\mathbf{R} \right]^{-1}.$$

The error vectors **e** and $\boldsymbol{\varepsilon}$ are assumed to follow independent multivariate normal distributions. $\boldsymbol{\varepsilon}$ represents random deviations applied to the beta coefficients, which should be independent of the sampling errors represented by **e**. For now, the transition covariance matrix $\boldsymbol{\Omega}$ is assumed to be a scaled submatrix of $\boldsymbol{\Sigma}$. Van Deusen (1999) gave a maximum likelihood estimator of the parameter *p*, which determines the strictness of the constraints that would apply in this case after a minor modification.

REFERENCES

Avery, T.E. and H.E. Burkhart. 1983. *Forest Measurements*. McGraw-Hill. NY. 331 pp.

Beers, T.W. and C.I. Miller. 1964. Point sampling: Research results, theory, and applications. *Purdue Univ. Research Bull* 786. 56 p. and insert.

Chandra, M.J. 2000. *Statistical Quality Control.* CRC Press. NY. 284 pp.

Eriksson, M. 1995. Compatible and time-additive change component estimators for horizontal-point-sampled data. *Forest Sci.* 41(4):796-822.

Martin, G.L. 1982. A method of estimating ingrowth on permanent horizontal sample points. *Forest Sci.* 28(1):110-114.

Meyer, H.A. 1953. *Forest Mensuration*. Penn Valley Publishers, Inc. State College, Pa. 357 pp.

Roesch, F.A. Jr. 1988. Compatible estimators for the components of growth and instantaneous volume calculated from remeasured horizontal point samples. Unpublished Ph.D. Dissertation, Rutgers University, New Brunswick, NJ.

Roesch F. A. 2001. Mixed Estimation for a Forest Survey Sample Design. In *1999 Proceedings of the Section on Statistics and the Environment: 1-6*, American Statistical Association, Presented at the Joint Statistical Meetings Baltimore, MD August 8-12, 1999.

Roesch, F.A. Jr., E.J. Green, and C.T. Scott. 1989. New compatible estimators for survivor growth and ingrowth from remeasured horizontal point samples. *Forest Sci.* 35(2):281-293.

Roesch, F. A. (in review). Compatible Estimators of the Components of Growth for an Annual Forest Inventory Design.

Theil, H. 1963. On the use of incomplete prior information in regression analysis. *JASA* 58:401-414.

Theil, H. 1971. *Principles of Econometrics.* John Wiley & Sons. New York.

Van Deusen, P. C. 1996. Incorporating predictions into an annual forest inventory. *Can. J. For. Res.* 26: 1709-1713.

Van Deusen, P. C. 1999. Modeling trends with annual survey data. *Can. J. For. Res.* 29(12): 1824-1828.

Van Deusen, P.C. 2000. .Alternative sampling designs and estimators for annual surveys. In: *Integrated Tools For Natural Resources Inventories In The 21st Century,* USDA Forest Service, North Central Research Station GTR NC-212, Mark Hansen & Tom Burk editors.

West, M., and J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag. New York. xxi + 704 pp.