LONG-TERM STRATEGY FOR THE STATISTICAL DESIGN OF A FOREST HEALTH MONITORING SYSTEM*

HANS T. SCHREUDER and RAYMOND L. CZAPLEWSKI
Mathematical Statisticians, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, 240 West Prospect Road, Fort Collins, Colorado 80526–2098, U.S.A.

(Received: February 1992; revised: July 1992)

Abstract. A conceptual framework is given for a broad-scale survey of forest health that accomplishes three objectives: generate descriptive statistics; detect changes in such statistics; and simplify analytical inferences that identify, and possibly establish cause–effect relationships. Our paper discusses the development of sampling schemes to satisfy these three objectives, but without any design restrictions implied by existing sample surveys. A general vision of a desirable future system will increase chances that short-term decisions will lead to better environmental monitoring systems in the long term.

1. Introduction

Surveys have been traditionally planned to provide estimates for current objectives. Over time, additional objectives can become important. For example, the original objective of the USDA (U.S. Department of Agriculture) Forest Service, Forest Inventory and Analysis (FIA) units, was to estimate merchantable wood volume by tree species and diameter class, and area of forest land by type, stand size, ownership, site quality, and stocking. Over time, additional information was desired, such as growth, mortality, timber removals, and regeneration rates.

Analytical inference using forest survey data is becoming more important. The 1982 and earlier FIA surveys of the States of Georgia and Alabama, USA, are striking examples. Average growth rates in tree diameters decreased (Bechtold *et al.*, 1991; Ruark *et al.*, 1991), but the cause of this decline is not known. Different groups propose various hypotheses, including: regional changes in average stand structure of forests, anthropogenic stressors such as atmospheric pollutants, past agricultural land use policies, and weather. As a result of the ensuing controversy, efforts were made to determine whether survey data could identify cause–effect relationships, and how survey data might more effectively serve analytical uses.

Use of survey data to establish cause–effect relationships is controversial (Schreuder and Thomas, 1992), although survey data have been used to identify possible cause–effect relationships in epidemiology (Feinstein, 1988). Schreuder and McClure (1991) suggest modifications of FIA to improve change detection and identify possible causes of change. The basic question is: how far can existing FIA surveys be changed from observational studies to completely controlled

* Invited paper originally presented at the International Union of Forest Research Organizations Conference on Optimal Design of Forest Experiments and Forest Surveys, London, September 10–13, 1991.

Environmental Monitoring and Assessment **27:** 81–94, 1993.

experiments within the budgetary constraints, without compromising traditional FIA objectives? They conclude that additional new sample plots paired with all or a subset of the existing FIA plots could achieve these new objectives, although traditional objectives could be jeopardized. Cause–effect relationships could be explored with existing FIA plots, and the additional plots used to test cause–effect hypotheses. Sample size of additional plots would be smaller than the existing FIA sample size so that more intensive and more frequent measurements could be made, including destructive sampling. However, cause–effect relationships are difficult to establish, and other methods might be required, depending on circumstances. The modifications to FIA proposed by Schreuder and McClure (1991) allow for several broad objectives: generate descriptive statistics; detect changes in such statistics; and analyze the data to identify, and possibly establish, cause–effect relationships. But even with these modifications, FIA surveys would be biased towards efficient descriptive statistics.

Our paper discusses the development of sampling alternatives to satisfy these same three objectives, but without design restrictions imposed by existing sample surveys. A general vision of a desirable future system will increase the chances that short-term decisions will lead to better systems in the long term. This is timely because the United States and other countries are implementing systems to monitor forest health, with objectives very different from the traditional forest inventories such as FIA. However, these infant forest health monitoring systems build on the traditions and infrastructure of existing forest inventory systems. New funding for forest health monitoring, plus new technologies that have been developed, may make new sampling strategies highly desirable and cost-effective in the future.

2. Review of the Literature

We will briefly review the topics of survey design, cause–effect, and possibly causes of tree growth declines and increased tree mortality.

2.1 Possible Causes of Forest Decline

To show how difficult it is to establish cause–effect for forest ecosystems in general, and specifically with survey data, we will briefly discuss several examples of possible causes for growth decline and increased mortality. Prinz (1987) considers climate change, and increases in ozone and acid deposition, as the most probable explanations for tree growth declines in Europe. Weather is a triggering or synchronizing factor in the appearance and recovery of short-term forest damage. Elevated ozone concentrations and continuous loss of soil nutrients by acid deposition cause long-term forest damage. Ozone concentration increases with altitude, and there are natural differences in soil nutrients; these confounded factors can affect the spatial distribution of forest decline. At the ecosystem level, air pollutants may influence nutrient cycling, population dynamics of arthropods or microbial species, succession, species composition, and biomass production.

Schulze (1989) hypothesizes that spruce decline in a German study area is caused by deposition of sulfur, nitrate, and ammonium, which significantly modify plant nutrition and soil chemistry. Spruce roots may be taking up ammonium rather than nitrate or magnesium. Soil acidification is accelerated and calcium/aluminum and magnesium/aluminum ratios decrease in the soil solution owing to the leaching of nitrate and sulfate. Root development and water and nutrient uptake are affected by soil solution chemistry. Canopy uptake of atmospheric nitrogen coupled with root uptake stimulated growth, caused a nitrogen-to-cation imbalance resulting in decline symptoms ranging from needle yellowing and loss to tree and stand mortality.

2.2 CAUSE–EFFECT

What are the conditions necessary to establish cause–effect, and how difficult is it to achieve them? Mosteller and Tukey (1977) note that two of three criteria–consistency, responsiveness, and mechanism – must be satisfied to infer cause–effect relationships. Consistency implies that the presence and magnitude of the effect (y) is always associated with a minimal level of the suspected causal agent (x). Responsiveness is established by experimental exposures to the suspected causal agent and reproducing the symptoms. Mechanism is established by demonstrating a biological or ecological process that causes the observed effect. Only consistency can be confirmed by observation alone. In addition, Feinstein (1988) advocates the following scientific principles: stipulate a research hypothesis prior to analysis; study a well-specified cohort having a statistical factor in common (e.g., undisturbed, naturally regenerated pine stands in the eastern coastal plain); collect high-quality data; study possible explanations; and avoid detection bias.

Smith and Sugden (1988) give statistical conditions necessary to establish cause–effect in experiments and surveys for analytical inference. With simple random sampling, sampling mechanism can be ignored for inference. The treatment assignment can be ignored if assignment is independent of the response variable (y). One can establish causal inference if the true population being sampled is similar to a population consisting of all possible responses to all possible treatments measured on each unit. Clearly, these are different populations; the first is a finite, actual population, and the second is a superpopulation that does not exist.

If each of N units in the population have two treatments, T_1 and T_2, we assume the realized response values are $y_i^{(j)}$ for the random variables $Y_i^{(j)}$ when treatment j $(j = 1, 2)$ is applied to unit i. The $N \times 2$ matrix of all potential responses defines the hypothetical population $Y = (Y^{(1)}, Y^{(2)})$, which are the $N \times 1$ response vectors if all units in the population receive both treatments $j = 1, 2$. Then the true causal effect of treatment 2 versus 1 for the ith population unit is the unobservable difference $Y_i^{(2)} - Y_i^{(1)}$, and the true population causal effect is the finite population mean difference:

$$\bar{D} = \bar{Y}^{(2)} - \bar{Y}^{(1)} . \tag{1}$$

If U denotes the population and $U^{(j)}$ is the set of population units assigned to treatment j, then the apparent population causal effect of treatments 1 and 2 is the difference between the two subpopulation means:

$$\bar{D}_a = \bar{Y}^{(2)}_{U_{(2)}} - \bar{Y}^{(1)}_{U_{(1)}} \tag{2}$$

which may be different from D in (1) depending on the treatment assignment, regardless of the sampling method used. Surveys for analytical inference followed by experiments or vice versa are the ideal ways to document cause–effect because both mechanisms can be ignored (Smith and Sugden 1988).

2.3 SURVEY DESIGN

It is clear from the literature (e.g., Kish, 1987; Skinner *et al.*, 1989) that designs involving simple random sampling within strata are preferred. The design should not be modified over time. With adequate statistical oversight, simple and stable designs allow for noncontroversial estimates of change, building of valid regression models, fitting calibration models for measurement errors, and valid testing for significant changes.

Scott (1990) and Scott and Alegria (1990) conclude that fixed area plots are more efficient than variable radius plots for estimating number of trees per unit area (mortality, net change, and current values – but no removals); tree regeneration and mortality (number of trees, basal area, and volume); and new growth and change in basal area. Variable radius plots samples are more efficient for current basal area; net volume (accretion, change, and current values – but not regeneration and mortality); and removals (number of trees, basal area, and volume). Generally, variable radius plots are more efficient for attributes associated with the large trees or current basal area (e.g., fiber volume); fixed area plots are more efficient for attributes associated with small trees (e.g., mortality and regeneration). Other variables, such as soil chemistry and biodiversity, might not be well correlated with tree size. Fixed area plots are more readily treated in experiments, which can be important to infer cause–effect relationships (Smith and Sugden, 1988).

Environmental monitoring requires routine, perhaps yearly, estimates for each geographic region. However, monitoring large regions can be prohibitively expensive; annual remeasurements by field crews can inadvertently disturb sample plots; and frequent remeasurements can strain landowners' tolerance of government agents who measure plots on private lands. Remeasurement of an interpenetrating subsample of plots can mitigate many of these problems. An interpenetrating subsample should be well distributed in space across the population (Messer *et al.*, 1991). For example, consider a systematic sample of 300 plots on a triangular grid. An interpenetrating subsample might consist of measuring 100 plots during the first year, where the first plot in each contiguous group of 3 plots would be measured. Another interpenetrating subsample of 100 plots could be measured during the second year, consisting of the second plot in each group of 3. The remaining plots would be measured in the third year. The interpenetrating cycle could then be repeated, beginning in the fourth year, by remeasuring the plots that were originally measured during the first year.

3. Sampling Strategy Considerations

Survey criteria and design and estimation (including updating estimation) are important considerations in formulating sampling strategies.

3.1 SURVEY CRITERIA

Credibility of the results from a monitoring program is of paramount importance. This requires stringent criteria in the survey. Based on our objectives, the following criteria should be satisfied.

1. Only variables that have negligible mesaurement errors, or can be efficiently calibrated with reference data with negligible measurement errors, should be used. Subsampling might be used for subjective variables, such as classification of a site as productive forest land or ocular estimates of canopy condition. However, subjective variables should not be emphasized in a survey unless they can be reliably calibrated across time and different observers using objective reference measurements. Subjective measurements have high rates of measurement error and unpredictable biases that can compromise their utility, and the cost of more objective and accurate measurements can be justified.

2. Data on certain stressor variables, such as low soil moisture or exposure to pollution, are needed for each plot. This requires several visits to each plot over the season; inexpensive recording instrumentation system placed on the plots; or reliable geostatistical interpolation of plot measurements from off-plot stations (e.g., kriging of atmospheric pollution data). The method of data collection can vary by stressor. For example, kriging might be used for estimating annual exposure of each plot to pollutants that are expensive to repeatedly measure over time, while inexpensive recording devices can more precisely measure a soil moisture index for each growing season on each plot.

3. Destructive sampling cannot be allowed to affect sample plots because remeasurement over time is essential for efficiency; off-plot destructive sampling might be restricted to the immediate vicinity of the plot. Exact locations of sample plots need to be kept secret to avoid biased treatment by landowners.

4. Redefinition (or reemphasis) of some variables is needed so that they can be measured accurately with remote sensing, rather than field sampling. Remote sensing can be more efficient than field data for some variables, while avoiding the inadvertent disturbance of plots by field crews and denial of access to plots by landowners. Truth defined in the field is different from truth defined by low-altitude photography, and it is not always clear which is better.

5. Simplicity in design.
6. Consistency of design over time.
7. Flexibility to address new environmental or management issues while maintaining design consistency.
8. Flexibility to incorporate new measurement technologies while maintaining design consistency.
9. Ability to treat each sample unit as a population. This means, for example, no missing data for a sample unit because of the design used.
10. Use interpenetrating sampling (Messer *et al.*, 1991) or similar method so sampling intensity can be readily increased in time and space as needed.
11. Provide flexibility to accommodate replacement of plots to deal with damage caused by the measurement process (e.g., trampling or destructive sampling) or denial of access to plots by private landowners; examples include sampling with partial replacement or rotating panel design.
12. Ability to handle systematic missing data, such as data not available because certain types of landowners may deny access to plots.
13. Implement a strong quality assurance program so that true changes in sample plots over time will not be confounded with changes in measurement error or subtle details in measurement protocol.

The basic design should be tailored to the specific population of interest. The Southern United States (13 states) is used as our example because we are familiar with that area; however, the basic design should have utility in other areas of the world with suitable modifications. The estimators used should be generally applicable and robust. Complexity in estimation is not as serious a problem as complexity in design, so we favor use of complex estimators if this results in significantly improved estimation.

3.2 Design

Desirable criteria for the sampling design are simplicity, flexibility to incorporate new technologies, and opportunity for retrospective data collection and analyses. The basic design we visualize is stratification, double sampling, and subsampling within the strata. The first level of sampling is to divide the population of interest into broad contiguous strata or subpopulations of interest. Forest subpopulations in the Southern United States might be coastal plain, Piedmont (low hilly areas, which usually are the most urbanized), and mountain areas. These areas differ in forest types and cause–effect situations (e.g., growth decline due to pollution would most likely be observed in the urban areas of the Piedmont), and different analyses

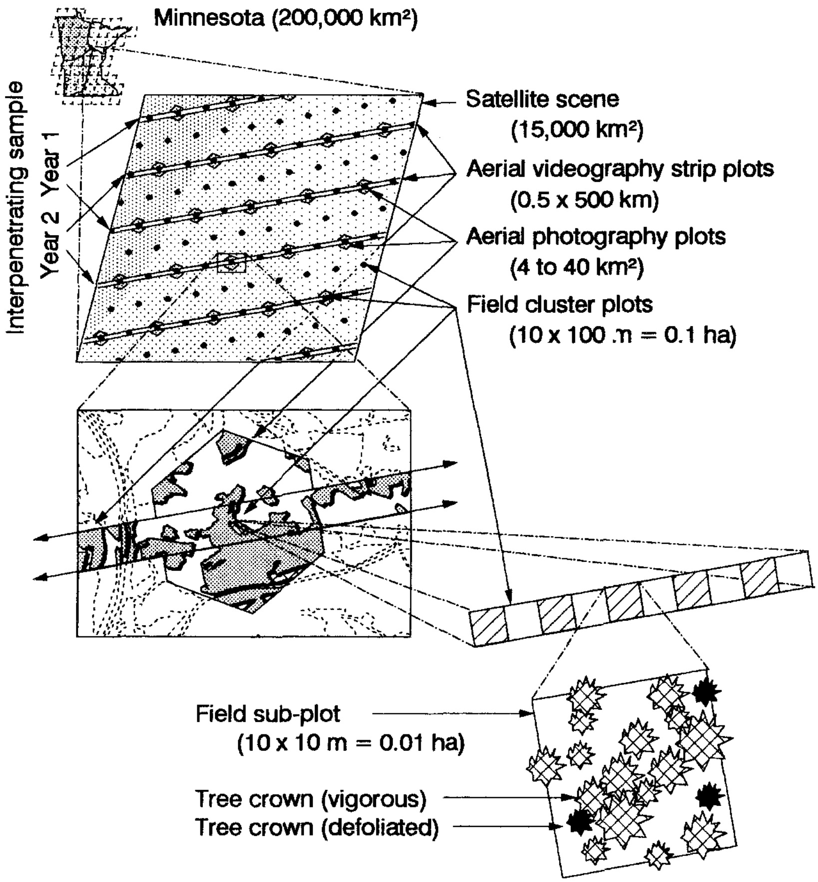

Fig. 1. Simplified example of a multi-level sampling frame for three broad objectives: generate descriptive statistics, detect changes in such statistics, and analyze the data to identify, and possibly establish cause–effect relationships. The coarsest level is a map of the population that is made with satellite data. This would monitor forest fragmentation and habitat corridors, and precisely estimate proportion of the population that is forest. New satellite imagery is interpreted every 1 to 4 years. Aerial photography and videography are used with strip plots and large cluster plots to monitor disturbances and changes in areal extent of more detailed categorizations of forest cover and land use; new imagery is interpreted every 1 or 2 years. Field plots are 0.1 ha clusters of 10 plots that are 10×10 meters in size. Some variables for a subsample of field plots are measured with the aerial photography or videography every 1 to 4 years; other variables are measured in the field every 4 to 16 years. The geometry of the triangular sampling grid permits incremental changes in sampling intensity, while maintaining even spacing of sample plots (Messer *et al.*, 1991).

might be of interest (e.g., the coastal plain is more intensively managed for wood products than mountain areas).

The next level of sampling would consist of an intensive grid of large, permanent, rectangular strip plots or transects (Figure 1). Strip plots can be used to efficiently sample small patches of forest disturbance caused by arthropods or microbial species. Low-level, high-resolution aerial videography would be acquired for strip plots that are 0.25 to 1.0 km wide by 100 to 500 km long, arranged in parallel every 4 to 32 km apart (i.e., 1.5% to 25% sample by area). (This assumes engineering solutions to problems associated with imaging long permanent transects.) This imagery should provide sufficient resolution to identify continguous areas of major disturbances to forest cover, such as landclearing, partial harvesting, tree-plantation establishment, and extreme tree mortality or defoliation. Aerial strip plots would be used primarily as an intermediate stage to estimate extent for different types of forest cover and land use, changes in extent over time, and extent and rate of change in disturbances. Changes would be measured by a single photointerpreter with paired images of the same sites that are acquired at two points in time; this minimizes variations in misclassifications errors among different interpreters. To maximize variability within each strip plot, plot length would match the scale of predominant landscape features, and plots would be oriented across the most prevalent landscape pattern (e.g., east–west if rivers oriented north–south). This orientation could also maximize image acquisition for broad-scale elevational ecotones, where ambient stress is more common and cumulative effects might first appear. Remotely sensed data would be acquired for each strip plot every 1 to 4 years. Estimates for each year could be produced with an interpenetrating subsample of plots. For example, assume half of the strip plots are imaged each year: every other plot could be sampled in years 1, 3, 5, . . ., and the remaining plots could be sampled in years, 2, 4, 6, . . . (Figure 1). Some remotely sensed variables would have minimal measurement error (e.g., area of mature forest); other such variables would be used as covariates with field measurements. This requires a subsample of accurately registered field plots in a regression or calibration framework.

At certain intervals along each strip plot, higher resolution aerial photography would be simultaneously acquired; the photography would image a subsample of the permanent field plots and their surrounding landscapes (Figure 1). Photoscales of 1:2000 to 1:12000 would be sufficient to interpret some of the following indicators: individual-tree mortality, crown discoloration, defoliation, regeneration, crown competition, and gap dynamics. The landscape immediately surrounding each field plot would be characterized with the same imagery; associations between landscape attributes and field measurements could suggest certain cause–effect relationships (e.g., forest fragmentation might be associated with increased defoliation caused by insect activity). Aerial photography would be useful for retrospective studies, where variables are measured with remote sensing techniques and historical imagery to provide estimates of past conditions.

A further subsample of permanent plots would be measured in the field (Figure 1). For purposes of discussion, we propose a plot 10 × 100 m (0.1 ha). These small plots would be permanently monumented for future remeasurement in the field. A subsample of field plots would fall within the remotely sensed imagery. An interpenetrating subsample of plots would be measured each year (e.g., 25% of plots measured first year, . . . 25% measured fourth year, 25% remeasured fifth year, ...). A field plot would be wide enough to accurately register to remotely sensed images and relocate in the field. This plot could be considered a cluster of ten 10 × 10 m subplots; where some variables are measured on all subplots and others on a subset. Some subplots could be set aside for analytical uses (such as more intensive measurements or destructive sampling).

Some descriptors of forest condition require coarse classification and mapping of all forest lands within the population. Examples include forest fragmentation and habitat corridors between patches of forested habitat within a nonforested landscape. For large regions, this requires satellite imagery (Figure 1). Entire satellite images would be completely interpreted every 4 to 12 years on an interpenetrating schedule (e.g., 1 satellite scene in every 2 by 2 cluster of scenes would be interpreted each year in a 4-year cycle). This would use digital techniques with a simple classification system; for example, forest might be classified into several general forest types (pines, hardwoods, mixed), subdivided by several greenness categories (i.e., classifications associated with high, medium, and low levels of biomass density or tree stocking). A simple classification system at this level avoids high probabilities of misclassification errors that are associated with more detailed classification systems. Multidate satellite imagery would be used to identify those field plots that were likely disturbed since the previous field observations (e.g., change in general forest type or stocking levels); these interpretations can allocate field sampling to those strata that have experienced similar levels of change or disturbance. However, such an allocation procedure can violate the simplicity criterion. Satellite data could improve precision of the estimated proportion of forest cover in the population, assuming accurate measurements for precisely registered sample plots are used as reference data in calibration, regression, or double-sampling estimators.

Variables with significant measurement error require calibration or regression estimators. Variables that are expensive to measure require double sampling. Models might be required for variables with high temporal or spatial variability. Destructive sampling requires special consideration in plot design. These considerations divide variables into several categories.

1. Variables that can be accurately measured on aerial photographs so that field sampling is not necessary. For example, change in area of mature forest over time can be measured accurately on high-resolution aerial photographs, but change in commercially suitable forest cannot. The latter classification is even subjective with field measurements.

2. Variables that can be measured easily in the field. Examples are tree mortality and number of trees. Such variables may also often be correlated with variables measured on aerial photographs.

3. Variables that are difficult or expensive to measure in the field. Examples are tree volume, tree crown condition, and horizontal and vertical vegetation profiles. Such variables might be weakly correlated with remotely sensed

measurements. Surrogates for these measurements should be sought (e.g., tree diameter and height measurements are surrogates for tree biomass or volume); measurement error might be substantial, and calibration estimators would be required.

4. Variables for which a within-growing season distribution may be desired. Examples are rainfall amounts, ozone concentrations, chemical composition of tree components, and symptoms of arthropod or microbial effects on trees.

5. Variables for which destructive sampling is required. Examples are soil and needle samples for chemical composition and tree cores for growth and dendrochronological studies.

3.3 SAMPLING INTENSITY

Sample size should be large enough to detect important changes and to allow for detailed *post*-stratification into subpopulations of interest. Sufficient power to detect change might mean a probability of at least 80 percent of rejecting the null hypothesis of no change when the true change is 5 percent and the level of significance used to test the null hypothesis is 5 percent. The null hypothesis might be defined by some disconcerting magnitude of change, such as a decrease in growth rate that exceeds 20 percent over a 10-year interval. After initial data analyses, different responses might be discovered in certain forest subpopulations. Natural variation in species diversity microsite and ecological conditions might also require *post*-stratification. In either case, *pre*-stratification might not be possible, and large sample sizes are desirable. This is analogous to the large data bases used to study human health and the complex interactions between people and their environments. Because cost cannot be ignored, prudent compromises are required among sample size, measurement detail, remeasurement frequency, and remotely sensed or field measurement protocols.

3.4 STATUS ESTIMATION

Desirable criteria for estimating status of a population are efficiency, consistency and robustness of estimators, and reliability of variance estimators. We recommend consideration of the general regression estimator $\hat{Y}_{gr}$ (Särndal 1982), where

$$\hat{Y}_{gr} = \sum_{i=1}^{n} y_i/\pi_i + a_s\Big(N - \sum_{i=1}^{n} 1/\pi_i) + b_s(X - \sum_{i=1}^{n} x_i/\pi_i\Big) \tag{3}$$

with y_i = the value of interest on unit i; x_i = the covariate value on unit i; π_i = probability of selecting unit i;

$$X = \sum_{i=1}^{N} x_i \; ,$$

n = sample size; N = population size; a_s and b_s = estimated regression coefficients:

$$a_s = \left[\sum_{i=1}^{n} y_i/(\pi_i v_i) - b_s \sum_{i=1}^{n} x_i/(\pi_i v_i)\right] / \sum_{i=1}^{n} 1/(\pi_i v_i) \,, \tag{4}$$

$$b_s = \frac{\{\sum_{i=1}^{n} (x_i - \tilde{x}_s)\, y_i/(\pi_i v_i)\}}{\{\sum_{i=1}^{n} (x_i - \tilde{x}_s)^2 \,/\, (\pi_i v_i)\}} \,. \tag{5}$$

v_i = variance weight for unit i.

The estimator is general and includes the widely used Horvitz–Thompson estimator as well as many regression estimators as special cases. $\hat{Y}_{gr}$ is asymptotically unbiased (Robinson and Särndal, 1983). A bootstrap variance estimator was consistently reliable in a simulation study (Schreuder and Ouyang, 1992). Additivity of estimates is desirable but not required. Reliability (precision) of estimates should be a higher priority. Estimation strategy should have the flexibility to utilize new sources of data as technology evolves.

3.5 Change Estimation

Desirable criteria for change estimation are the same as for status estimation. In addition, power to detect significant change is required. Plots should be permanent, and all plots should be remeasured. Sampling with partial replacement is not recommended because it causes considerable analytical complexity, particularly for subpopulations, but such sampling may be required in view of possible needs for destructive sampling and not exhausting landowner tolerance. Also, individual sampling units can change subpopulations over time (e.g., succession of pine stands to hardwood stands). If estimates at times t and $t-1$ are $\hat{Y}_t$ and $\hat{Y}_{t-1}$, however those may have been obtained, we prefer estimating ΔY_t as

$$\Delta\hat{Y}_t = \hat{Y}_t - \hat{Y}_{t-1} \,. \tag{6}$$

3.6 Updating Estimates

Desirable criteria for updating estimates are credibility with user groups, robustness, efficiency, and verification and updating of models. We distinguish this from change estimation because updating may be more model-dependent, and hence less credible than change estimation. In updating, we may use a model to predict change in some variables (e.g., growing stock volume per unit area for undisturbed strata) without actually measuring the change in those variables since the last field measurements, then use change estimation procedures to directly estimate current state for less predictable variables (e.g., growing stock volume per unit area for the disturbed stratum). The sum of the two types of estimates will produce estimates of population totals (e.g., total growing-stock volume of forest). Model parameters will be updated with new field data, and prediction residuals must be inspected for biases and changes in variance of prediction errors. This proposal is untested.

3.7 Analysis with Models

Desirable criteria for analyses are simplicity, credibility, opportunity for retrospective analysis, biologically realistic mechanisms, and well-defined comparisons. The design being proposed is simple and there may be no need to weight observations in building models. The models should be built on the basis of data collected and biological knowledge, and their predicted values for variables should be compared to new observations on these variables as they become available. This comparison allows verification and modification of the models. Models are collections of hypotheses regarding ecosystem response, which can be scientifically evaluated with monitoring data.

3.8 Identifying Changes and Potential Causes

Desirable criteria for identifying and establishing cause of substantial changes are: maximize probability of establishing the first two Tukey-Mosteller postulates, minimize required assumptions, and provide opportunity for follow-up analyses. Substantial changes are those that are both statistically significant and out of the ordinary. For example, if ΔY_{t_1} and ΔY_{t_2} are statistically different, this could be a substantial change if the magnitude of the difference is judged to be large (e.g., 20 percent change over 10 years). If a substantial change is observed, all potential causes for such change need to be listed and examined. Experiments could test hypotheses that are developed with monitoring data, but use off-frame experimental units (e.g., long-term ecological research sites). Or a subset of field plots set aside for analytical purposes can be used for experimental treatments. Interpenetrating subsampling would allow earlier remeasurements of existing plots if needed to quickly evaluate new hypotheses. Historical aerial imagery for plots can allow for some retrospective studies; high-resolution aerial photography could be used to quantify past disturbance height-growth and mortality, and tree cores used to quantify past tree-growth on plots not measured in the past.

As the forest on a permanent plot matures, growth mortality and regeneration rates change, even in the absence of abnormal stressors. Substantial changes in forest condition might be caused by regional demographics of forest lands, which are effected by regional land use and land management practices. Expected effects of individual stand characteristics and silvical factors on growth mortality and regeneration must be quantified before hypotheses regarding exogenous factors are tested. Analysis of residual differences between model predictions and field observations might improve the power to test hypothesized causes of regional change that are not expected through normal silvics. Models can be considered a collection of hypotheses that explain forest responses, and alternative models (i.e., alternative hypotheses) can be compared based on their relative fit to field observations from forest health monitoring plots. If one or more dynamic models have explicit processes that describe response to environmental stressors, then cause–effect hypotheses might be evaluated directly with models. Process-oriented models might provide one link between extensive sample surveys for forest health, and more intensively-studied long-term ecological research sites.

4. Summary and Conclusions

A very large institutionalized program is required to monitor forest health in a meaningful fashion. This is justified based on the growing concerns about our environment in general, and our forests in particular. The cost of environmental policies are high, and we need to scientifically evaluate their effectiveness. This can help societies better allocate their limited resources to those actions that produce the greatest environmental benefits. A forest health monitoring system could approach or surpass the size of national forest inventory systems. If designed properly, we could achieve forest health monitoring objectives and, at the same time, provide benefits to current forest inventory systems. For example, the proposed 500-km aerial transects could follow portions of the existing FIA plot grid, and a subsample of existing FIA photo and field plots could be monitored each year with forest health monitoring procedures. A long-term vision of a desirable future forest health monitoring system will increase the chances that short-term design decisions will ultimately lead to better systems in the long-term.

References

Bechtold, W.A., Ruark, G.A., and Lloyd, F.T.: 1991, 'Changing Stand Structure and Regional Growth Reductions in Georgia's Natural Pine Stands', *For. Sci.* **37**, 703–717.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., and Bailey, R.L.: 1983, *Timber Management: A Quantitative Approach.* J. Wiley and Sons, New York, 333 pp.

Feinstein, A.R.: 1988, 'Scientific Standards in Epidemiologic Studies of the Menace of Daily Life', *Science* **242**, 1257–1263.

Kish, L.: 1987, *Statistical Design for Research.* J. Wiley and Sons, New York, 267 pp.

Messer, J.J., Linthurst, R.A., and Overton, W.S.: 1991, 'An EPA Program for Monitoring Ecological Status and Trends', *Environ. Monit. Assess.* **17**, 67–78.

Mosteller, F. and Tukey, J.W.: 1977, *Data Analysis and Regression.* Addison-Wesley Publishing Co., Reading, Massachusetts.

Prinz, B.: 1987, 'Causes of Forest Change in Europe', *Environment* **29**, 10–36.

Robinson, P.M. and Särndal, C.E.: 1983, 'Asymptotic Properties of the Generalized Regression Estimator in Probability Sampling', *Sankhya* **45**, 240–248.

Ruark, G.A., Thomas, C.E., Bechtold, W.A., and May, D.M.: 1991, 'Growth Reductions in Naturally Regenerated Southern Pine Stands in Alabama and Georgia', *South. J. Appl. For.* **15**, 73–79.

Särndal, C.E.: 1982, 'Implications of Survey Design for Generalized Regression Estimation of Linear Functions', *J. Stat. Plan. Inform.* **8**, 155–170.

Schreuder, H.T. and McClure, J.P.: 1991, 'Modifying Forest Survey Procedures to Establish Cause–Effect, Should it be Done?', in: *Proc. IUFRO World Forestry Congress*, Paris, France, Sept.

Schreuder, H.T. and Ouyang, Z.: 1992, 'Optimal Sampling Strategies for Weighted Linear Regression Estimation', *Can. J. For. Res.* **22**, 239–247.

Schreuder, H.T. and Thomas, C.E.: 1992, 'Establishing Cause–Effect Relationships Using Forest Survey Data', *For. Sci.* **37**, 1497–1525.

Schulze, E.D.: 1989, 'Air Pollution and Forest Decline in a Spruce *(Picea abies)* Forest', *Science* **244**, 776–783.

Scott, C.T.: 1990, 'An Overview of Fixed Versus Variable-Radius Plots for Successive Inventories', in: *Proc. Symposium, State-of-the-art Methodology of Forest Inventory.* USDA Forest Service, Pacific Northwest Research Station General Technical Report PNW–GTW–263, pp. 97–104.
Scott, C.T. and Alegria, J.: 1990, 'Fixed Versus Variable-Radius Plots for Change Estimation', in: *Proc. Symposium, State-of-the-art Methodology of Forest Inventory.* USDA Forest Service, Pacific Northwest Research Station General Technical Report PNW–GTW–263, pp. 126–1323.
Skinner, C.J., Holt, D., and Smith, T.M.F. (Eds.): 1989, *Analysis of Complex Surveys.* J. Wiley and Sons, New York, 309 pp.
Smith, T.M.F. and Sugden, R.A.: 1988, 'Sampling and Assignment Mechanisms in Experiments, Surveys, and Observational Studies', *Int. Stat. Rev.* **56**, 165–180.