A Review of Methods for Updating Forest Monitoring System Estimates

Hector Franco-Lopez, Alan R. Ek, and Andrew P. Robinson

Abstract.—Intensifying interest in forests and the development of new monitoring technologies have induced major changes in forest monitoring systems in the last few years, including major revisions in the methods used for updating. This paper describes the methods available for projecting stand- and plot-level information, emphasizing advantages and disadvantages, and the subsequent use of the updated information in constructing new estimates. Research directions that might lead to improving the updating process are also discussed.

OVERVIEW

Lund (1993) defined forest monitoring as the periodic measurements or observations of selected physical, chemical, and biological parameters for establishing baselines for detecting and quantifying changes in the forest over time, and to evaluate management and silvicultural alternatives (including harvesting). In practice, a forest monitoring system (FMS) is typically a strategic survey designed to check on or to keep track of the condition of the forest in a region, providing up-to-date information for the development of policies and programs for protection, management, and forest utilization.

The usefulness of broad-scale forest inventory information is determined by its frequency, accuracy, and subject matter (Sheffield 1987). The most generally requested improvements associated with these factors are more frequent inventories (up-to-date information), greater spatial sampling intensities to provide accurate local estimation, improved subject matter coverage (incorporation of new variables and improved measures of change), and increased information availability. Another concern with the information provided by FMS's is the reduction in length of the inventory reporting cycle. There is substantial demand for more current data (4- to 6-year cycles instead of the typical 10- to 15-year cycle).

Smith (1990) defines the term "update" as an estimate of the current forest condition derived by modeling the dynamic change in a forest from a known time and condition in the past. Update estimates are developed because full-scale inventories are too expensive to be conducted as information is needed; by updating, information needs often can be met at substantial cost savings over remeasurement. In addition, updating can enhance the quality of regional and national assessments since inventories varying in age can be projected to a common year.

Several authors (e.g., Ek 1990 and Smith 1990) define the following as desirable attributes of an updating technique:

1. The update should include revision of all plot and tree variables that might have changed since the last measurement;
2. Input and output of an updating system would necessarily need to be consistent and compatible with inventory data definitions and availability. This ensures that the same software can be used to summarize updated and remeasured records;
3. The level of resolution of the updating system needs to be capable of long-term projections as well as short-term updates;
4. The updating system should be calibrated for all stand conditions, and calibration data should be representative of the type of stands to which the data will be applied. This condition is essential to get unbiased updates; and
5. The system should meet the requirements of a range of users, from sophisticated users that require individual tree records to those seeking only aggregated inventory statistics.

Since updating can be defined as projecting forest inventory information to an specific date, we have two basic questions to answer: (1) How do we project tree and/or plot information to a specific date? and (2) How do we use the updated information to construct estimates? The following sections review and summarize the available methods for addressing these two questions.

PLOT INFORMATION PROJECTION

Two groups of plot-level variables are of concern in updating. The first group is comprised of the categorical variables, such as land use, cover type, stand size and density class, and disturbance history. The second group includes continuous variables, such as volume, basal area, and number of trees per acre.

Research Assistant and CONACyT fellow, Professor, and Research Assistant, respectively, Department of Forest Resources, University of Minnesota, St. Paul, MN, USA.

Updating Categorical Variables

In large-area FMS's, there may be substantial land-use and forest-type changes over the monitoring cycle. Classification or stratification update is the basic ingredient for area estimation. Teuber (1990) stated that classification updating has been practically restricted to the identification and quantification, through remote sensing, of important area changes resulting from timber harvesting, regeneration, land-use conversion, and natural disturbance. There are two basic approaches to the use of remote sensing for updating. The first compares two independent estimates of area (without an identification process in which you can say which specific plots have changed). The second approach consists of registering imagery obtained at two or more points in time and comparing them to determine specific plots or areas that have changed (digital change detection). Poso (1993) proposed a postclassification approach based on field measurements, avoiding all preclassifications and using measurable tree and stand variables. An advantage of his approach is that a classification function based on field measurements is constructed and can be used to update tree, plot, and stand information. However, this approach assumes a much more intensive inventory than those normally conducted in North America.

Other approaches to updating classifications include the use of econometric techniques to update area statistics by forest type, and the Kalman filter (Czaplewski 1990). In the first case, exogenous data on harvesting and planting are used to model classification changes. In the second, the Kalman filter approach combines a series of forest inventories, updates, or monitoring estimates, using a model of expected change in forest condition over time, to predict cover changes from land-use conversions, regeneration, and harvest.

The seemingly straightforward approach of estimating area by class from remote sensing leaves little room for updating detailed statistics, because classification is imprecise and even small changes in remote sensing classification affect area estimations in a very significant way. Thus, even though classification or stratification is the basic ingredient for estimation of totals (volume, growth, and removals), the lack of improvements in the precision of the approaches we use to update remote sensing classification has stalled this approach.

As a more direct solution to classification and updates, considerable progress has been made in Northern Europe. Tomppo (1991) and Tokola *et al.* (1996) applied the k-nearest neighbors (kNN) method for classification problems and used it to produce localized estimates and maps in the national forest inventory of Finland. This "pixelwise" approach of kNN is used to propagate field sample plot classification and volume information to all pixels in a region. The method is based upon the spectral distance between the pixels having known field data and the pixels to be classified. A distance function (e.g., Euclidean) $d_{(i)p}$ is computed in the feature space from the pixel p to each pixel i (on satellite imagery) whose ground truth is known from sample plots. The estimate of the ground variable m for the subject pixel is then given by:

$$\hat{m}_p = \sum_{j=1}^{k} w_{(j),p} m_{(j),p}$$

where $w_{(j),p}$ are related distance calculated weights and $m_{(j),p}$ are the values of the variable m in the k closest pixels in the spectral space to the pixel p. Classifications are imputed as the modal or "nearest" class. Variations of this approach have also been tested with success in the Swedish national forest inventory (Nilsson 1997).

Updating Continuous Variables

Burkhart (1993) described how inventories are commonly updated through the application of growth and yield models. Many different approaches to modeling growth and yield have been taken, ranging from techniques that use individual trees as the basic projection unit to methods that use only overall plot or stand summary data.

Stand-Level Models

In plot- or stand-level models, variables such as plot age, site index, basal area and/or number of trees per unit of area are used to update aggregate stand basal area, number of trees, and volume. Since volume distribution by tree species and size class is difficult to project directly, disaggregation or allocation models may be needed to provide that detail.

Individual Tree Models

Updating using this kind of model involves simulation of the growth of each tree and then aggregation to provide estimates of stand growth and yield (i.e., we obtain the estimates by summing individual tree characteristics and multiplying by plot, and other expansion factors). Such models typically consist of three basic tree submodels: (1) diameter growth, (2) height growth, and (3) mortality.

Individual tree models can be divided into distance independent and distant dependent types (Munro 1974). Distance independent models do not consider individual tree locations on plots in developing the basic submodel estimates of tree growth and mortality. Further, they typically use a plot level measure of competition (such as basal area) in estimating tree level growth. Distance dependent models typically use a competition measure based on the size and distance of competitors.

When individual tree models satisfy the desirable attributes of an updating technique, they provide the following advantages over other methods:

1. They provide updated information about stand dynamics and structure, including the distribution of tree and stand variables by classes. Through the analysis of these characteristics, the user obtains information about more complex issues (cover type change, succession, biodiversity, etc.).
2. They can be used for updating tree categorical variables (i.e., log grade, health, or condition, etc.).
3. An updated plot has the same information as a field measured plot, so the same database management system and software can be used for inventory analysis and compilation as well as update.
4. They supply a wide range of information for analysis by sophisticated and non-traditional users.

Individual tree models have been the most widely used technique for forest inventory updating. Individual tree models such as STEMS (Stand and Tree Evaluation and Modeling System) and PROGNOSIS emerged in the mid-1970's and have become the choice for many Forest Inventory and Analysis (FIA) applications. The use of STEMS to project undisturbed plots in the FIA unit of the USDA Forest Service's North Central Research Station is a typical example of the use of individual tree models for updating FMS's (Birdsey and Schreuder 1992).

As interest in updating FMS statistics through the use of models has increased, the following improvements have been commonly requested:

- **Variance information.** Most projection systems provide predictions of different forest attributes, but usually they do not provide estimates of the variance of the predictions. This information is crucial in the updating procedure to assess the precision of predictions, to evaluate the general performance of the models, and to calculate confidence intervals for the estimates. Given the usual complexity of the models used in forestry, analytical calculations of prediction variances are impossible to get. Mowrer (1988) used Monte Carlo simulation to provide a relative estimate of the errors propagated over repeated measurements by two models. This is a relatively straightforward procedure that propagates only errors in regression predictor variables and their cross products. Under Mowrer's approach, the variance associated with the regression coefficient errors was ignored. Gertner (1987) used the error propagation method as an alternative to obtain estimates of precision of predictions made with a forest model. He concluded that although the estimates were slightly biased, this method provides a mean for calculating the variance in a computationally efficient manner.

- **Accuracy and detail.** The estimation for early stages of stand development is commonly weak. Improvements are needed in the form of including ingrowth and/or full regeneration estimation. Only a few models contain components that will predict regeneration conditions after harvesting. The principal factor limiting the widespread development of regeneration models is the imprecision in estimating postharvest conditions. Ek *et al.* (1997) suggested that a regeneration model should predict stand conditions for a given stand at some time during the first two decades after harvest. It is important to improve early-stage estimation for credible long-term projection and to achieve credibility among users with ecological interests (Ek 1990).

- **Sensitivity to local conditions.** It is a very frequent practice to extend the results or to use a model outside of the region in which it was initially calibrated. But doing so typically introduces bias in estimations. Although a re-calibration of the model (calculating new regional coefficients) to compensate for local conditions is in order, the use of model adjustments is frequently found in the literature (Smith 1983). Adjustment factors attempt to correct for the difference between a predicted value and the real field measured value. This approach is popular because of its simplicity, but its usefulness does not last long. Also, because it pools all local factors together, there is no way to get a good explanation of their separate influence on the predictions. There are two possible solutions to improve local sensitivity. The first solution is to identify the factors driving the local responses (weather, disturbances, etc.), assess the specific weight of each factor among the dependent variables, and then develop new models using them as predictors. The second possibility is to improve the model re-calibration procedure for models.

Alternative methods to the use of regression models for updating FMS's are beginning to appear in the forestry literature. Simple and multiple imputation, as proposed by Rubin (1987), is a statistical technique designed to handle missing data. This technique allows statistical analysis using the same software used for complete databases. Simple imputation consists of filling in a single value for each missing value from an underlying model and database. A model in this context can be as simple as the tabular representation of the data corresponding to different strata. The single value imputed does not reflect sampling variability about the actual value or additional uncertainty when more than one underlying model is being entertained. However, it does retain the variability in the data and simplifies data processing. Multiple imputation consists of filling each value *m* times with randomly selected data from an underlying model and database. The resulting m data sets are analyzed separately using the method for complete data. These *m* intermediate results are then pooled into a final result.

Van Deusen (1997) discussed alternative methods for the analysis of annual, interpenetrating, systematic samples. He concluded that any analysis appropriate for a complete data set is appropriate for imputed, simulated-complete data sets, whether the imputed values might come from regression models or by matching measured with unmeasured sample units.

The few examples of the use of imputation in forestry can all be catalogued as simple imputation. Poso (1978) described the use of the grouping method in the Finnish national forest inventory. In this method, photo plots were interpreted for all the important attributes that show sufficient correlation between photo and field estimates. The photo plots were then grouped, and from each group one photo plot was randomly drawn and measured in the field. The information from this plot was then propagated to all photo plots belonging to the respective homogeneous group. Since a single value from an underlying model (the strata) was imputed for each one of the photo plots, this constitutes a single imputation example. Moeur and Stage (1995) briefly described another single imputation example in the Swedish FMS. In this case, an unsampled (in the field) first-phase unit is paired with a randomly selected second-phase (field sampled) unit from the same stratum. Ek *et al.* (1997) developed and tested imputation models to estimate postharvest forest stand characteristics. The imputation models were developed by tabular analysis of different stand regeneration conditions. They recommended cover type, age, and site index based models that provided realistic and detailed species and size class descriptions for regenerated stands.

"Plotwise" and "treewise" kNN methods have also been proposed as an alternative to regression techniques for updating continuous variables. In this approach, kNN is used to propagate plot or tree field information to unmeasured units. Haara *et al.* (1997) compared the kNN with Weibull parameter models for basal area diameter distribution estimation in a "plotwise" approach. They determined that the accuracy of the Weibull function estimation was better than that of kNN, but the latter reproduced the covariance structure of the attributes best.

Korhonen and Kangas (1997) used the "treewise" kNN approach for generalizing sample tree data. They found diameter, location (y and x coordinates), mean age of growing stock, basal area of growing stock, and mean diameter of growing stock to be good predictors for stem volume and sample tree characteristics (height and age).

Moeur and Stage (1995) described the "most similar neighbor" inference procedure, which chooses the most similar plot from the set of plots detailed to act as stand-in. The stand-in is chosen on the basis of similarity measures that summarize the multivariate relationships between a set of low resolution indicator attributes and detailed second-phase sample attributes.

ESTIMATION

Estimation involves use of the available data to actually construct forest monitoring systems estimates. In the cases described below, current observations, updates of earlier observations, and older observations that have not been updated are all considered as available data.

Estimation "as is"

This is the simplest estimation case. It consists of using the updated plot and/or tree data directly, relying entirely on the projection technique employed (i.e., using updates "as is," as field data). However, because the techniques we employ to project information may not be calibrated for all conditions, or simply because they are biased or imprecise, we typically seek more refinement in estimation.

The estimators described next can be considered members of a family of estimators commonly called "composite estimators." The general approach consists of combining two or more alternative estimators of a single quantity, i.e., weighting them to find a good estimator in some well-defined sense (Burk *et al.* 1982). Examples include the following:

Moving Average

This method is biased and usually considered an *ad hoc* procedure. It consists simply of taking the last n-year's observations and averaging them. The underlying assumption is that no major changes occur during those n-years. Optional weighting (inverse to time since last measurement or inverse to the variance) is sometimes considered. This estimator does not use projected plot information, but it is attractive for FMS's where some percentage of plots are measured each year. If area estimates, including stratification by area or condition, are precise and accurate and conducted in a consistent manner over time, then for many stratum variables, a moving average of stratum means can be simple and effective.

Double Sampling for Regression

The main assumption behind the standard double sampling with regression estimators is that a linear relationship exists between the variable of interest and a covariate. Examples of viable covariates are previous observations, or projected or imputed values substituted for observations. When projections are used, ratio or regression models describing the observed versus predicted relationship may then be used to adjust model outputs directly. These same adjustments may be used to adjust outputs for projections into the future. The same observed versus predicted data may also be used to recalibrate the model or to adjust certain parameters. However, with complex non-linear models, iterative derivation of parameter adjustments may be necessary to reach agreement between observed and predicted values. The North Central FIA unit uses the growth model enhanced, two-phase sample design described by Hansen (1990), using model projections as a covariate. In this particular case, the standard error for the projection model is unknown and the differences between the projected and measured values are used to compute model adjustment factors.

Gertner (1987) indicated that this method of calculating the variances is limited when projected observations are used as covariates. The disadvantage is that an independent set of remeasured plots is needed for calibration purposes, and this independent set must match the characteristics of the plots being projected in order for the variances to be correctly estimated.

The utility of this estimator depends on the strength of the relationship between the variate and covariate. An advantage of this estimator is that its statistical properties are well known; i.e., it provides valid error estimates. It is also a versatile method since several different attributes can be used as covariates in the same inventory. For example, consider adjusting the projected observations using ancillary data (e.g., local weather disturbances during the updating period). With this approach, many useful variables not usually considered in the projection models can be introduced very easily.

The Kalman Filter

This is a composite estimator that combines two independent estimates at a time, each of which is weighted inversely proportional to its variance (Czaplewski 1990). One of the independent estimates is a current estimate or monitoring measurement, and the other is a previous estimate that is updated for expected changes over time using a deterministic prediction model. The variance for this updated estimate includes the effects of (1) errors in the previous forest inventory that are propagated over time and (2) model prediction errors between previous and current estimates. Errors in this composite estimator are typically less than errors in either of the prior estimates alone. However, as with many other update techniques, there is a need for accurate models providing prediction errors. Descriptions of this technique can be found in the forestry literature (Czaplewski 1990, Dixon and Howitt 1979).

FUTURE DIRECTIONS

The above discussion describes the difficulties in the seemingly simple task of updating forest inventories. This review further highlights the areas needing research if updating is to be improved. The most important research needs are summarized below.

- Explore the use of imputation (single and multiple), especially the kNN method, as an alternative to regression in updating large-scale FMS's. In particular, research should consider improving local area updates, including those for classifications, growth estimation, and mapping of diverse forest inventory variables.

- Develop a set of procedures to generalize the use of computationally intensive methods for calculating the variance of the estimations from complex projection systems.

- Revisit re-calibration for local adjustment. This is an everyday problem in updating. This research would include identifying the factors driving local responses in models (weather, disturbances, etc.). This kind of analysis would be helpful in the variable selection procedure for constructing new models, for re-calibration for local adjustment, and for better understanding of the factors behind forest growth prediction.

- Given the expansion in methods for updating tree and plot information (single and multiple imputation), revisit the estimation procedures and evaluate them under these new circumstances.

ACKNOWLEDGMENT

The authors gratefully acknowledge helpful reviews by Thomas E. Burk, Professor, University of Minnesota, St. Paul, MN, USA; and Mark H. Hansen, Research Forester, USDA Forest Service, North Central Research Station, St. Paul, MN, USA. Research supported by the College of Natural Resources and the Agricultural Experiment Station, University of Minnesota.

LITERATURE CITED

Birdsey, R.A.; Schreuder, H.T. 1992. An overview of forest inventory and analysis estimation procedures in the Eastern United States: with an emphasis on the components of change. Gen. Tech. Rep. RM-214. Fort

Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 11 p.

Burk, T.E.; Hansen, M.H.; Ek, A.R. 1982. Combining sources of information for improved in-place inventory statistics. In: In-place resource inventories: principles and practices, proceedings of a national workshop; 1981 August 9-14; Orono, ME. SAF Publ. 82-02. Bethesda, MD: Society of American Foresters: 413-442.

Burkhart, H.E. 1993. Tree and stand models in forest inventory. In: Proceedings of Ilvessalo symposium on national forest inventories; 1992 August 17-21; Helsinki, Finland. Res. Pap. 444. Helsinki, Finland: The Finnish Forest Research Institute: 165-170.

Czaplewski, R.L. 1990. Combining inventories of land cover and forest resources with prediction models and remotely sensed data. In: Global natural resource monitoring and assessment: preparing for the 21st century: proceedings of the international conference and workshop; 1989 September 24-30; Venice, Italy. American Society for Photogrammetry and Remote Sensing. 3: 1079-1089.

Czaplewski, R.L. 1990. Kalman filter to update forest cover estimates. In: State of the art methodology of forest inventory: a symposium proceedings; 1989 July 30-August 15; Syracuse, NY. Gen. Tech. Rep. PNW-263. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: 457-465.

Dixon, B.L.; Howitt, R.E. 1979. Continuous forest inventory using a linear filter. Forest Science. 25(4): 675-689.

Ek, A.R. 1990. Interfacing forest inventory with growth and yield models. In: State of the art methodology of forest inventory: a symposium proceedings; 1989 July 30-August 15; Syracuse, NY. Gen. Tech. Rep. PNW-263. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: 415-419.

Ek, A.R.; Robinson, A.P.; Radtke, P.J.; Walters, D.K. 1997. Development and testing of regeneration imputation models for forests in Minnesota. Forest Ecology and Management. 94(1-3): 129-140.

Gertner, G. 1987. Approximating precision in simulation projections: an efficient alternative to Monte Carlo methods. Forest Science. 33(1): 230-239.

Haara, A.; Maltamo, M.; Tokola, T. 1997. The k-nearest neighbour method for estimating basal-area diameter distribution. Scandinavian Journal of Forest Research. 12: 200-208.

Hansen, M.H. 1990. A comprehensive sampling system for forest inventory based on a individual tree growth model. St. Paul, MN: Department of Forest Resources, University of Minnesota. 256 p. Ph.D. dissertation.

Korhonen, K.T.; Kangas, A. 1997. Application of nearest-neighbour regression for generalizing sample tree information. Scandinavian Journal of Forest Research. 12(1): 97-101.

Lund, H.G. 1993. Sampling designs for national forest inventories. In: Proceedings of Ilvessalo symposium on national forest inventories; 1992 August 17-21; Helsinki, Finland. Res. Pap. 444. Helsinki, Finland: The Finnish Forest Research Institute: 17-25.

Moeur, M.; Stage, A.R. 1995. Most similar neighbor - an improved sampling inference procedure for natural resource planning. Forest Science. 41(2): 337-359.

Mowrer, H.T. 1988. A Monte Carlo comparison of propagated error for two types of growth models. In: Forest growth modeling and prediction: proceedings of the IUFRO conference; 1987 August 23-27; Minneapolis, MN. SAF Publ. 87.12. Bethesda, MD: Society of American Foresters. 2: 778-785.

Munro, D.D. 1974. Forest growth models - a prognosis. In: Growth models for tree and stand simulation: proceedings of 1973 IUFRO meeting, working party S4.01-4; Stockholm, Sweden: Royal College of Forestry Research Notes. 30: 7-21 .

Nilsson, M. 1997. Estimation of forest variables using satellite image data and airborne lidar. Umeå, Sweden: Swedish University of Agricultural Sciences. 31 p. Ph.D. dissertation.

Poso, S. 1978. National forest inventory in northern Finland. In: National forest inventory joint meeting of IUFRO groups: S4.02 forest resource inventory and S4.04 forest management planning and managerial economics; 1978 June 18-26; Bucharest, Romania. Bucharest, Romania: IUFRO: 378-385.

Poso, S. 1993. Forest classification based on sampling and measurements. In: Proceedings of Ilvessalo symposium on national forest inventories; 1992 August 17-21; Helsinki, Finland. Res. Pap. 444. Helsinki, Finland: The Finnish Forest Research Institute: 255-259.

Rubin, D.B. 1987. Multiple imputation for nonresponse in surveys. New York: John Wiley and Sons. 258 p.

Sheffield, R.M. 1987. Advancing the usefulness of forest surveys. In: Forests, the world, and the profession: proceedings of the 1986 Society of American Foresters national convention; 1986 October 5-8; Birmingham, AL. SAF Publ. 87.02. Bethesda, MD: Society of American Foresters: 321-324.

Smith, W.B. 1983. Adjusting the STEMS regional forest growth model to improve local predictions. Res. Note NC-297. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 5 p.

Smith, W.B. 1990. Inventory updates: overview of the North Central experience. In: State of the art methodology of forest inventory: a symposium proceedings; 1989 July 30-August 15; Syracuse, NY. Gen. Tech. Rep. PNW-263. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: 442-448.

Teuber, K.B. 1990. Remote sensing techniques for updating forest inventories. In: State of the art methodology of forest inventory: a symposium proceedings; 1989 July 30-August 15; Syracuse, NY. Gen. Tech. Rep. PNW-263. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: 436-441.

Tokola, T.; Pitkänen, J.; Partinen, S.; Muinonen, E. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing. 17(12): 2333-2351.

Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. International Archives of Photogrammetry and Remote Sensing. 28(7-1): 419-424.

Van Deusen, P.C. 1997. Annual forest inventory statistical concepts with emphasis on multiple imputation. Canadian Journal of Forest Research. 27(3): 379-384.