AN EFFICIENT ESTIMATOR TO MONITOR RAPIDLY CHANGING FOREST CONDITIONS

Raymond L. Czaplewski, Michael T. Thompson, and Gretchen G. Moisen[1]

Abstract.—Extensive expanses of forest often change at a slow pace. In this common situation, FIA produces informative estimates of current status with the Moving Average (MA) method and post-stratification with a remotely sensed map of forest-nonforest cover. However, MA "smoothes out" estimates over time, which confounds analyses of temporal trends; and post-stratification limits gains from remote sensing. Time-series estimators, like the Kalman Filter (KF), better detect and analyze unexpected or rapid changes in dynamic forests. KF is a recursive multivariate model-based estimator that separates complex time-series of panel estimates and multi-sensor remotely sensed data into a sequence of smaller and more manageable components. Population-level results are disaggregated into expansion factors that assure additivity and simplify small area and small domain estimation. Other statistics gauge fit of alternative models to annual FIA panel data, which permits quantitative rankings among alternative cause-effect hypotheses.

INTRODUCTION

The 1998 Report by the Blue Ribbon Panel on Forest Inventory and Analysis (FIA) motivated the comprehensive redesign of the FIA program (Bechtold and Patterson 2005). FIA replaced decadal periodic surveys with annual panel surveys to produce more timely analyses for every State. However, a single annual panel uses only 10 to 20 percent of the field plots available to a periodic survey. To improve precision, FIA uses a simple Moving Average (MA) of five or more annual panels. While design-unbiased as an estimator of the average conditions among multiple panels, MA is biased for time-series of annual estimators (Bechtold and Patterson 2005). This bias is acceptably small whenever net change is relatively minor, but not when landscapes are affected by unusually rapid changes. Concerns with the MA for annual trend analyses are escalating. In 2012, the National FIA User Group recommended "renewed efforts to investigate alternatives to the simple moving average for improved trend detection and estimation, … including short term projections (5 to 10 years)." We describe an estimator designed to satisfy this and previous recommendations from the National FIA User Group.

[1] Research Mathematical Statistician (RLC), U.S. Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84404; Forester (MTT) and Research Forester (GGM), U.S. Forest Service, Rocky Mountain Research Station. RLC is corresponding author: to contact, call 970- 218-5907 or email at rlczaplewski02@fs.fed.us.

MONITORING ANALYSES AND THE NATURE OF CHANGE

To some degree, every acre of every forest changes every year through predictable processes of stand dynamics, ambient disturbances, timber management, and socioeconomic forces. In a static landscape, changes in forest conditions and land use are nearly at equilibrium, and MA is an acceptable statistical estimator. However, other landscapes change more rapidly.

Unexpected change can be subtle, spatially ubiquitous, and undetected during early onset. An example is the growth decline of southern pines during the 1970s. Causes might have been changes in air pollution, climate, land use, and/or distribution of stand conditions (Gadbury and Schreuder 2004). Such changes are not well observed with remote sensing, although geospatial data on stressors (e.g., air pollution) contribute valuable circumstantial information. Sufficiently precise estimators require a large sample of FIA field plots, which implies analyses over large geographic areas (e.g., a multi-state ecoregion) and long time intervals (e.g., 5 to 20 years). Detailed analyses might include the small subsample of "Phase 3" Forest Health Monitoring plots (Bechtold and Patterson 2005). Cause-effect analyses might use model-based inference to compare alternative cause-effect hypotheses (Gadbury and Schreuder 2004).

Other changes are episodic disturbances, which are often apparent with the naked eye. An example is severe mortality of western pines caused by outbreaks of mountain pine beetles. Other examples include changes in the extent of wildfires, timber harvest and management treatments, conversion and reversion among agricultural fields and forestlands, and development within the wildland–urban interface. Though locally intense, they might affect only 1 to 5 percent of a forested landscape per year, and they are observed with a correspondingly small number of FIA plots. Annual remote sensing provides indicators that are well correlated with the extent, intensity and location of such changes. Remeasurements of FIA field plots at 5- to 10-year intervals monitor detailed tree-level consequences of stand-level changes.

Models of population processes are arguably essential for detailed monitoring. Model-based estimators increase precision with small sample sizes. A model can capture an analyst's hypotheses regarding expected behavior of a forest population. Residual differences between model predictions and panel estimates detect deviations from expectations. Different models represent alternative sets of cause-effect hypotheses, and analyses of residuals compare the fit of each alternative to FIA field data. Models forecast future conditions based on past processes.

Sensitivity of a monitoring program depends upon a sufficiently large sample size within each annual panel. Furthermore, numerous field plots are required for statistical methods that empirically compensate for systematic measurement errors with remotely sensed data. Hence, the target population must cover large geographic expanses, perhaps spanning several states. However, certain changes tend to "average out" as the extent of the population increases, and many monitoring questions involve small subpopulations. A partial solution is multivariate small area estimation, which uses diverse sets of full-coverage geospatial data (e.g., Landsat and MODIS) as predictors of field observations (e.g., Czaplewski 2010).

A MODEL-BASED TIME-SERIES ESTIMATOR

The sample-survey literature covers a diverse collection of estimators for individual pieces of a statistical monitoring system. However, the multivariate "Kalman filter" (KF) estimator (Bar-Shalom et al. 2001) can integrate all pieces into a single cohesive structure. The senior author is using a matrix language to develop software that implements the following.

KF is a time-series technique. It combines a design-based panel estimator with a model-based estimator for expected changes in population parameters (Czaplewski and Thompson 2009). KF is a sequential recursive method. It starts at time t=1 with the first panel of field data and FIA's design-based estimator. The resulting vector estimate of population parameters (i.e., "state-vector") serves as initial conditions in a model for changes in the population as posited by the analyst. This multivariate linear model predicts the state-vector at time t=2, including a covariance matrix for random errors propagated from time t=1 plus estimated prediction errors between times t=1 and t=2. The second panel of field plots provides an independent design-based estimate of corresponding population parameters for t=2. KF uses the multivariate composite estimator to "optimally" combine these competing model-based and design-based estimates. The result is a single, more precise estimator at time t=2. This "best" estimate at time t=2 serves as initial conditions for the transition model that predicts the state vector at time t=3. This sequential recursive technique proceeds for the entire time-series.

Analyses of residual differences between model-based predictions and design-based panel estimates help improve the estimated covariance matrix for model prediction errors, thereby mitigating bias in the model-based portion of the KF estimator. The model represents analysts' understanding of population-level processes (Czaplewski and Thompson 2012) or a population-level aggregation of plot-level processes (see for example Healey et al., these proceedings). The model can forecast future conditions and the associated covariance matrix for random errors.

The state vector has partitions for each year. This autoregressive structure improves estimates for many time periods with each FIA panel. Its covariance matrix provides variance estimates for changes between 5- or 10-year intervals, which KF uses with corresponding design-based estimates from plot remeasurements to improve annual estimates of status and changes.

The state vector may have hundreds of variables, and digital "round-off" errors can degrade numerical results. However, the engineering literature abounds with solutions that use the square root of a covariance matrix (e.g., Bar-Shalom et al. 2001). Covariance matrices are typically rank-deficient, and feasible estimates require thoughtful pivots of state-space.

KF computes a vector of "optimal" weights that combines each model- and design-based vector estimate at each time-step. Restrictions on those weights can impose inequality constraints. For example, the estimated annual mortality rate of insect-infected live trees can exceed 100 percent if sampling errors in two independent annual panels are large. Inequality constraints force the estimated rate between 0 and 100 percent. Those same population-level weights may be stored in the FIA plot-level database as time-series of multivariate expansion factors, one for each state variable, at the condition and tree levels (Czaplewski 2010). This step assures additivity across statistical tables, facilitates certain types of small domain and small area estimation, and potentially reduces certain difficulties in analyses with mixed-condition plots.

Insufficient sample size causes sampling zeros, which can produce implausible estimates for "rare" variables. We collapse classification systems so that each category has at least 50 plots within each annual panel. Assuming no cross-classifications, annual sample size within Colorado's forests would support only seven forest type groups, seven ownership categories, and seven tree species groups. KF expansion factors permit more detailed estimates within the FIA database, but the statistical efficiencies of those detailed estimates are suboptimal.

Czaplewski (2010) developed KF structures for multiple sources of multivariate remotely sensed and other geospatial data. Unlike post-stratification, geospatial variables may be continuous or categorical, with or without cross-classification. KF uses full-coverage Landsat data or sample surveys with LIDAR or high-resolution aerial photography. This KF structure accommodates time-series of remotely sensed data, including annual change detection. Czaplewski (2010) illustrates compatible methods that use geospatial data for small-area estimates for special studies, which improves the compromise between large sample sizes and small analysis areas.

DISCUSSION AND CONCLUSIONS

FIA analysts and user community require defensible estimates of trends in forest resources. Estimates at one point in time, such as forest area or amount of standing live biomass, have limited value. Detailed assessments of insect epidemics, such as the mountain pine beetle in the West, require reliable estimates of annual tree mortality over long timespans. Monitoring broad-scale trends in tree growth helps better understand effects of climate change. Before making major capital investments, a forest products company must know trends in timber volume, and their causes, within a modestly sized geographic area. Trends in tree removals are a substantial component of economic assessments, such as the effect of recessions on the forest product sector. To serve these analysis requirements, FIA requires an easily understood statistical estimator that supports diverse analyses of time-series with panel data.

Although the MA estimator is easily understood, it can have substantial lag-bias. On the other hand, the purely design-based estimator for each independent panel is unbiased for annual monitoring. Unfortunately, the latter is limited by small sample sizes. Annual trends can be estimated only through differences among estimates from independent panels (e.g., independent estimates of lodgepole pine in Montana for 2010, 2011, and 2012). Sampling error can exceed net annual change, however, obscuring major changes in a rapidly changing population, or producing statistical estimates of change that misleadingly appear large for a truly static population. Furthermore, independent panels limit the ability to understand the causes of annual change. Remeasurements of individual FIA plots at 5- or 10-year intervals help better understand long-term changes at the plot and tree scales, but this protracted remeasurement interval obscures annual trends. Regardless, the design-based approach, by itself, restricts an analyst's ability to quantify and interpret trends at the annual time scale.

The multivariate Kalman filter is a relatively simple alternative in an annual monitoring program. It fully utilizes all available remotely sensed data. It stores results as condition- and tree-level expansion factors, which simplifies analyses. Its structure helps detect unexpected changes and rank competing sets of cause-effect hypotheses. This model-based approach is inherently multidisciplinary, however, and success requires teamwork among analysts, modelers, remote sensing technologists, computer scientists, and statisticians.

ACKNOWLEDGMENTS

The authors thank reviewers Scott Baggett, John Stanovick, and Erkki Tomppo. The authors ultimately accept full responsibility for the published content.

LITERATURE CITED

Bar-Shalom, Y.; Li, X.-R.; Kirubarajan, T. 2001. **Estimation with applications to tracking and navigation.** New York, NY: John Wiley and Sons. 584 p.

Bechtold, W.A.; Patterson, P.L. 2005. **The enhanced Forest Inventory and Analysis program - national sampling design and estimation procedures.** Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Czaplewski, R.L. 2010. **Complex sample survey estimation in static state-space.** Gen. Tech. Rep. RMRS-GTR-239. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 124 p.

Czaplewski, R.L.; Thompson, M.T. 2009. **Opportunities to improve monitoring of temporal trends with FIA panel data.** In: McWilliams, W.; Moisen, G.; Czaplewski, R., comps. Forest Inventory and Analysis (FIA) Symposium 2008; 2008 October 21-23; Park City, UT. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 55 p.

Czaplewski, R.L.; Thompson, M.T. 2012. **Model-based time-series analysis of FIA panel data absent re-measurements.** Res. Paper. RMRS-RP-xx. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Gadbury, G.L.; Williams, M.S.; Schreuder, H.T. 2004. **Revisiting the southern pine growth decline: Where are we 10 years later?** Gen. Tech. Rep. RMRS-GTR-124. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p.