Forest Ecology and Management 483 (2021) 118777

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Updating annual state- and county-level forest inventory estimates with data assimilation and FIA data

Zhengyang Hou [a,*], Grant M. Domke [b], Matthew B. Russell [c], John W. Coulston [d], Mark D. Nelson [b], Qing Xu [e,**], Ronald E. McRoberts [c]

[a] College of Forestry, Beijing Forestry University, Beijing 100083, PR China
[b] Northern Research Station, U.S. Forest Service, Saint Paul, MN, United States
[c] Department of Forest Resources, University of Minnesota, Saint Paul, MN, United States
[d] Southern Research Station, U.S. Forest Service, Blacksburg, VA, United States
[e] Key Laboratory on the Science and Technology of Bamboo and Rattan, International Centre for Bamboo and Rattan, Beijing 100102, PR China

ARTICLE INFO

Keywords:
Data assimilation
Updating procedure
Annual estimation
Mixed estimator
Unbiasedness
Forest inventory and analysis
Forestland area estimation

ABSTRACT

The United Nations Framework Convention on Climate Change requires annual estimates for forestry and ecological indicators to monitor the change in forest resources, the sustainability of forest management, and the emission and sink of forest carbon. It is particularly important to update estimates of forestland area in a timely fashion and at flexible geographical scales, not only for its value in monitoring biological diversity at the ecosystem scale, but also because of its close association with other indicators such as forest biomass and carbon. However, in the US, the Forest Survey Handbook advises that the sampling error should not exceed 3% per 404686 ha (one million acres) of forestland area, a demanding standard barely met by pooling the Forest Inventory and Analysis (FIA) panel data measured in an inventory cycle of 5–10 years. Consequently, this study aims to propose and illustrate an updating procedure using data assimilation that integrates a design-based estimator with a model-based mixed estimator for updating annual estimates at two population levels, the state- and county-levels. The three states in the USA, Minnesota (MN), Georgia (GA) and California (CA), representing the Northern, the Southern and the Pacific Northwest FIA programs, constitute the study areas. FIA data collected were based on a 5-year inventory cycle for MN (2006–2010) and GA (2005–2009), and a 10-year cycle for CA (2001–2010). The total number of sample plots was 17764 for MN, 6323 for GA, and 16740 for CA. Distinguishing features attribute to this procedure include: (1) unbiasedness: the integration of design-based estimates into the mixed estimator introduces a favorable property – unbiasedness, which could be the property national forest inventories concern the most; (2) efficiency: considerable improvements in estimation precision greater than 55%, achieving sampling errors as small as those relying on using 5–10 years pooled FIA data; (3) time: compared with the temporal trends reflected by design-based estimates, the updated trends were of much smoother trend lines and narrower confidence intervals that would better depict temporal changes for a population at flexible spatial scales; (4) space: this procedure is scale-invariant, meaning its efficiency is not affected by an inventory employing either a large- or small-area estimation, which was demonstrated at the two population levels; and (5) generalizability: this procedure is unbiased and efficient, 100% compatible with the FIA database which is readily available to the public, and thus suitable for various official reporting instruments.

1. Introduction

Intergovernmental organization and processes including the United Nations Framework Convention on Climate Change require annual estimates for forestry and ecological indicators to monitor the change in forest resources, the sustainability of forest management, and forest carbon emissions and sinks. In the United States, the National Report on Sustainable Forests provides a new opportunity to produce and distribute these updates (NRSF, 2019). Among the indicators, forestland area is particularly important because of its immediate association with

* Corresponding author.
** Corresponding author.
E-mail addresses: houzhengyang@bjfu.edu.cn (Z. Hou), qing.xu@icbr.ac.cn (Q. Xu).

https://doi.org/10.1016/j.foreco.2020.118777
Received 10 January 2020; Received in revised form 15 October 2020; Accepted 4 November 2020
Available online 17 November 2020

others such as forest biomass and carbon. For example, with the gain-loss approach, emissions and sinks for a land use change class are estimated as the product of the class area estimate and the per unit area estimate of carbon change for the class (GFOI, 2016; IPCC, 2006). Therefore, the national Greenhouse Gas Reporting Program has a similar, yet stronger need for updating estimates of forestland area in the U.S., not only in a timely fashion but at flexible geographical scales (GHRRP, 2019).

The Forest Inventory and Analysis (FIA) national program of the Forest Service, U.S. Department of Agriculture, conducts annual forest inventories in all states. The U.S. Agricultural Research Extension and Education Reform Act of 1998 requires the measurement of 10–20% of the permanent sample plots in each state each year to construct panel data spanning a measurement cycle of 5–10 years. Sample rates vary between and within FIA regions, states, and measurement years due to budgetary and operational factors (FIA, 2019). Permanent plots are subsequently remeasured on a 5- or 10-year interval. Forestland area is estimated from measurements in the sample plots, different with many other countries where the estimation is made from maps or satellite information. These data are stored in an extensive FIA database, publicly available for producing estimates with associated measures of uncertainty. The U.S. Forest Service Handbook (FSH 4809.11) advises that the uncertainty measure using sampling error should not exceed 3% per 404 686 ha (one million acres) of productive forestland area, resulting in a base sample intensity of one plot per 2400 ha. This demanding accuracy standard is barely achievable by pooling all panel data of respective years on a 5- or 10-year rotating basis (Bechtold and Patterson, 2005). To further reduce the sampling error, some states and National Forests choose to intensify their sample at e.g., two times or greater rates than the base sample intensity.

For producing the annual estimates from FIA data, there have been at least three approaches considered (McRoberts, 1999, 2001). The most straightforward approach simply uses the panel data to obtain estimates in the current year for reflecting current conditions. A drawback to this approach is that the inferential precision may be unacceptable for most forest attributes and ecological indicators due to the small annual sample size based on 10–20% of plots measured in any year. The second approach is a moving average estimator currently adopted by FIA (Patterson and Reams, 2005; Edgar et al., 2019). The precision of this approach is increased because data for all sample plots are used for estimation, but the downside is that the estimates reflect a moving average of conditions over the measurement cycle and trail current conditions in the presence of temporal trends (Van Deusen, 2002). A third approach is to update to the current year the initial estimates obtained from an estimator that can be design- or model-based. If the updating procedure is unbiased and precise, this approach can provide nearly the same precision as using pooled data without suffering from the adverse effects of using outdated information. Today, updating procedures of this kind are categorized as data assimilation.

Data assimilation (DA) is an umbrella term for a broad category of mathematical procedures that update existing parameter estimates for the purpose of estimating as precisely as possible the state of a system through the integration of multisource information (Fletcher, 2017; Lahoz et al. 2010). DA problems can be solved using various updating procedures rooted in estimation theory including the mixed estimator, Kalman filter and Bayesian statistics (Blayo et al., 2014; Katzfuss et al., 2016; Theil and Goldberger, 1961). Apart from forestry, DA has already been used in many domains such as biogeography and meteorology typically entailing estimation of parameters, calibration of observation networks, and prediction of missing values (Lahoz and Schneider, 2014; Robinson, 1991).

Distinct from Bayesian statistics, the best linear unbiased predictor (BLUP) is a frequentist counterpart formula for estimating the conditional mean for multivariate normal vectors (Henderson, 1975; Hou et al., 2019; Robinson, 1991). BLUP essentially adjusts the parameters of the marginal distribution of a random variable to the parameters of its conditional distribution given the jointly distributed variables observed. Although BLUP is mostly used for predicting random effects for mixed models, it is the equation from which the Kalman filter was mathematically derived (Czaplewski, 1990; Kalman 1960; Robinson, 1991).

Like the Kalman filter, the mixed estimator (ME) is also derived from BLUP and is regarded as a compromise between frequentist and Bayesian approaches (Kangas, 1991; Theil 1971). Mathematically, the form of ME is the same as the BLUP for predicting the random effects when there is no fixed effect (section 5.2 in Robinson, 1991). In theory, ME is equivalent to the Kalman filter under specific assumptions (Dixon and Howitt, 1979); in use, ME was more efficient for shorter time series compared to the Kalman filter, because ME does not require unknown starting values that adversely influence short-term Kalman filter results (Van Deusen, 1991, 1999, 2002), a feature advantageous to updating annual forest inventories typically ranging from five to 10 years.

Consequently, the objectives of this study were threefold: (1) to compare two design-based estimators, the expansion estimator and the post-stratified estimator, for the annual estimation of forestland area at both the state- and county-levels; (2) to propose and illustrate a DA-based updating procedure that integrates the design-based estimators with the model-based mixed estimator for updating annual estimates at the two population levels; and (3) to compare the inferential properties before and after applying the updating procedures for the multiple scenarios. A highlight of this updating procedure resides in its combination between the first and third approaches introduced above for producing the annual estimates, thus ensuring unbiasedness of the procedure as illustrated below.

2. Materials

2.1. Populations

The three states in the USA, Minnesota (MN), Georgia (GA) and California (CA), representing the Northern, the Southern and the Pacific Northwest FIA programs, constitute the study areas. In accordance with the national Greenhouse Gas Reporting Program on conducting inventories at flexible geographical scales, the population was defined for two-levels, the state-level and the county-level. At the state-level, each state is a population; and at the county-level, each county is a population. In this study, the inference is about the total of forestland area in a population.

2.2. FIA data

For MN and GA, FIA measures one panel of sample plots per year that comprises a base sampling intensity of 20% of all plots in each state within a complete inventory cycle of five-years. California conducts FIA annual inventory on a ten-years cycle, making a panel per year comprised of 10% of all plots.

The Forest Service has established field plot centers in permanent locations using a quasi-systematic sampling design that is regarded as producing an equal probability sample (McRoberts et al., 2010). Each sample plot consists of four 7.32 m radius circular subplots, configured as a central subplot and three peripheral subplots with centers located at 36.58 m and azimuths of 0°, 120°, and 240° from the center of the central subplot. Exhaustive information on measuring trees, conditions, and plots is described in Bechtold and Patterson (2005) and FIAFG (2019).

Data used in the present study were based on a 5-year inventory cycle for MN (2006–2010) and GA (2005–2009), and a 10-year cycle for CA (2001–2010). The total number of sample plots was 17764 for MN, 6323 for GA, and 16740 for CA. GA and CA plots were sampled at the base intensity; MN plots were sampled at twice as many, resulting in one plot per 1200 ha. These data are publicly available and can be extracted from a large, relational FIA Database (FIADB, 2018).

3. Methods

3.1. Expansion estimator

With a simple random sampling or (quasi) systematic sampling design, a population parameter such as the total or mean can be estimated using the expansion estimator (Royall and Herson, 1973). This estimator is simple to use, but may not be efficient at reducing variances, particularly for small sample sizes, n, or relative to the spatial heterogeneity of the attribute of interest in a population (Brewer and Gregoire, 2009; Hou et al., 2015). The expansion estimator is probability-based and design-unbiased, a favorable feature ensuring the unbiasedness of inference, which is particularly important for official reporting instruments. In this study, the expansion estimator was applied to FIA panel data for obtaining the initial annual state- and county-level estimates that will be updated to the same year with the mixed estimator in Section 3.3.

For estimating the total forestland area, A, the attribute of interest is the proportion of the sample plot i categorized or mapped as forestland, $P_{id} = \frac{\sum_j^4 \sum_k^{K_{ij}} a_{mijk}\delta_{ijkd}}{a_m \overline{P}_m}$, where a_{mijk} is the mapped area of the j^{th} subplot covering condition k in the i^{th} sample plot, i.e. accessible forestland;δ_{ijkd} is a zero-one domain indicator, which is one if a_{mijk} belongs to the domain of interest d; K_{ij} is the number of conditions; $a_m$is the plot area; and $\overline{P}_m = \sum_i^n \sum_j^4 \sum_k^{K_{ij}} \frac{a_{mijk}\delta_{ijk}}{a_m n}$, is an adjustment factor representing the mean proportion of mapped plot areas within the geographical boundary of the population, with δ_{ijk} indicating a zero-one domain indicator, which is one if a_{mijk} is inside the boundary.

The mean forestland proportion of the population can then be estimated using $\overline{P}_d = \frac{1}{n}\sum_i^n P_{id}$ with $\widehat{Var}\left(\overline{P}_d\right) = \frac{\sum_i^n P_{id}^2 - n\overline{P}_d^2}{n(n-1)}$. Finally, the expansion estimator for the total forestland area can be expressed as $\widehat{A}_{EXP} = A_P \overline{P}_d$ with $\widehat{Var}\left(\widehat{A}_{EXP}\right) = A_P^2 \widehat{Var}\left(\overline{P}_d\right)$, where A_P is the total area of the population (Cochran, 1977; Thompson, 2012).

3.2. Post-stratified estimator

For post-stratified estimation, a population is partitioned into strata, and a previously established sample is assigned with those strata. Because the assignments are made independently between strata, variances estimated with a post-stratified estimator for individual strata can be summed together to obtain the variance for the population (Cochran, 1977; Holt and Smith, 1979). Variance reduction and flexibility for assigning different stratifications to a fixed sample are the advantages of using post-stratified estimation, but the disadvantages are that the homogeneity of post-stratification may affect the inferential precision and a small stratum could contain few sample plots (Ghosh and Vogt, 1993; Little, 1993). The post-stratified estimator can reduce to the expansion estimator in a way that any hierarchical structures in the sample data are discarded as if there was just a single stratum (Bechtold and Patterson, 2005, p.49). Like the expansion estimator, the post-stratified estimator is also design-unbiased, applied to FIA panel data for obtaining the initial annual state- and county-level estimates that will be updated to the same year with the mixed estimator in the next section.

The attribute of interest is the proportion of the sample plot categorized as forestland, $P_{hid} = \frac{\sum_j^4 \sum_k^{K_{hij}} a_{mhijk}\delta_{hijkd}}{a_m \overline{P}_{mh}}$, where P_{hid} is the proportion of plot i in the domain d for plots assigned to stratum h; a_{mhijk} is the mapped area of subplot j covering condition k, i.e. accessible forestland;δ_{hijkd} is a zero-one domain indicator, which is one only if a_{mhijk} belongs to d; K_{hij} is the number of conditions; $a_m$is the plot area; and $\overline{P}_{mh} = \sum_i^{n_h} \sum_j^4 \sum_k^{K_{hij}} \frac{a_{mhijk}\delta_{hijk}}{a_m n_h}$, is an adjustment factor representing the mean proportion of mapped plot areas within the geographical boundary of the population, with n_h indicating the sample size assigned to stratum h, δ_{hijk} another zero-one domain indicator, which is one if a_{mhijk} falls inside the boundary.

The stratum-specific mean can then be estimated using $\overline{P}_{hd} = \frac{1}{n_h}\sum_i^{n_h} P_{hid}$ with $\widehat{Var}\left(\overline{P}_{hd}\right) = \frac{\sum_i^{n_h} P_{hid}^2 - n_h \overline{P}_{hd}^2}{n_h(n_h-1)}$. Finally, the post-stratified estimator for the total forestland area takes the form $\widehat{A}_{STR} = A_P \sum_h^H W_h \overline{P}_{hd}$ with $\widehat{Var}\left(\widehat{A}_{STR}\right) = \frac{A_P^2}{n}\left[\sum_h^H W_h n_h \widehat{Var}\left(\overline{P}_{hd}\right) + \sum_h^H (1-W_h)\frac{n_h}{n}\widehat{Var}\left(\overline{P}_{hd}\right)\right]$, where W_h is the weight for stratum h (Bechtold and Patterson, 2005, section 4.3). Conveniently with FIA data, a population was readily post-stratified using satellite data, and the information about each variable required for applying the post-stratified estimator was readily available (FIADB, 2018).

3.3. Data assimilation with ME

Data assimilation with the mixed estimator (ME) was used for updating the annual forestland area estimates initially obtained from the expansion estimator and the post-stratified estimator at the state- and county-levels. ME is model-based, developed from Durbin (1953), Theil and Goldberger (1961) and Theil (1971), and related to mixed modeling, because ME can be regarded as a special case of a mixed model without any fixed effects (Goldberger, 1962; Henderson, 1963, 1973; Robinson, 1991). Under a common construct, the context of BLUP is a linear mixed model, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, where $\boldsymbol{y}$ is a vector of observations; $\boldsymbol{\beta}$ is a vector of unknown parameters having fixed values, i.e. the fixed effects; $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known matrices; and $\boldsymbol{u}$ and $\boldsymbol{e}$ are vectors of random effects such that $E(\boldsymbol{u}) = 0$, $E(\boldsymbol{e}) = 0$ and $Var\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G} & 0\\ 0 & \boldsymbol{R}\end{bmatrix}$, where $\boldsymbol{G}$ and $\boldsymbol{R}$ are known positive definite matrices. The BLUP for $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are solutions to the simultaneous equations $\begin{cases}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X\beta} + \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Zu} = \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y}\\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X\beta} + (\boldsymbol{G}^{-1} + \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z})\boldsymbol{u} = \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y}\end{cases}$ (Henderson, 1950); when there are no fixed effects, the random effects are simply $\boldsymbol{u} = \left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right)^{-1}\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y}$ with $Var(\boldsymbol{u}) = \left(\boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1}\right)^{-1}$(Robinson, 1991, section 5.2), elucidating the origin of ME and relationships with other theories in statistics.

Consider a model that denotes the total forestland area estimated using the expansion estimator or post-stratified estimator at year t, $A_t = a_t + \varepsilon_t$, where ε_t is a random error with a mean of zero and variance γ_t^2 (Cassel et al., 1977). This formulation indicates a_t as a random variable instead of a fixed parameter. To constrain the change of a_t over time, a transition equation was applied, $a_t - a_{t-1} = v_t, t = 2, \cdots, T$, where v_t is another random variable, independent of ε_t, with a mean of zero and variance $q\gamma_t^2$; and q is an unknown scaling factor to be estimated. This transition constraint is essentially a numerical approximation to the first derivative of the underlying temporal trend in the forestland areas over years (Van Deusen, 1996, 1999). Compactly, these relationships can be expressed in a matrix form as $\begin{bmatrix}\boldsymbol{A}\\ 0\end{bmatrix} = \begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{C}\end{bmatrix}\boldsymbol{a} + \begin{bmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{v}\end{bmatrix}$ (Theil, 1963, 1971; Theil and Goldberger, 1961), where $\boldsymbol{A} = [A_1, A_2, \cdots, A_T]'$, $\boldsymbol{a} = [a_1, a_2, \cdots, a_T]'$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{v}$ are vectors of random errors; 0 is a vector of zeros; $\boldsymbol{I}$ is an identity matrix; and $\boldsymbol{C}$ is a constraint matrix implementing the transition equation, with a dimension of $(T-1)\times T$, and on each row a sequence of $[-1 \quad 1]$ for elements starting at column t and zeros elsewhere. The covariance matrix of $\boldsymbol{\varepsilon}$ is denoted as $\Sigma = E(\boldsymbol{\varepsilon\varepsilon}')$, and the transition covariance matrix as $\Omega = E(\boldsymbol{vv}')$ consisting of a $(T-1)\times(T-1)$ submatrix of Σ multiplied by q, i.e. $\Omega = \mathrm{diag}\left(q\gamma_2^2, \cdots, q\gamma_T^2\right)$. When Σ and q are estimated, the form of ME under this construct can be expressed as $\widehat{\boldsymbol{a}} = \left(\frac{1}{q}\boldsymbol{C}'\widehat{\Omega}^{-1}\boldsymbol{C} + \widehat{\Sigma}^{-1}\right)^{-1}\widehat{\Sigma}^{-1}\boldsymbol{A}$ with

$\widehat{Var}(\widehat{\boldsymbol{a}}) = \left(\frac{1}{q}\boldsymbol{C}^{'}\widehat{\Omega}^{-1}\boldsymbol{C} + \widehat{\Sigma}^{-1}\right)^{-1}$ (Kangas, 1991; Van Deusen, 1999). Note the similarity between $\widehat{\boldsymbol{a}}$ and $\boldsymbol{u}$, $\widehat{Var}(\widehat{\boldsymbol{a}})$ and $Var(\boldsymbol{u})$. For consistency, we denote $\widehat{A}_{ME} = \widehat{\boldsymbol{a}}$ and $\widehat{Var}\left(\widehat{A}_{ME}\right) = \widehat{Var}(\widehat{\boldsymbol{a}})$. The $\widehat{A}_{ME}$ and $\widehat{Var}\left(\widehat{A}_{ME}\right)$ give respectively the updated annual estimates and the updated variance estimates for the total forestland area.

Technically, all terms required by applying ME were estimated based on either the expansion estimator or the post-stratified estimator introduced in Sections 3.1 and 3.2. The vector $\boldsymbol{A}$ consisted of the area estimates $\widehat{A}_{EXP}$ or $\widehat{A}_{STR}$ at respective years. The covariance matrix $\widehat{\Sigma}$ was assumed to be diagonal, consisting of elements $\left[\widehat{\gamma}_1^2, \widehat{\gamma}_2^2, \cdots, \widehat{\gamma}_T^2\right]$, where $\widehat{\gamma}_t^2$ is the estimated variance $\widehat{Var}\left(\widehat{A}_{EXP}\right)$ or $\widehat{Var}\left(\widehat{A}_{STR}\right)$ at year t. Although the diagonal assumption does not account for covariances among observations for remeasured plots, the updating procedure illustrated here allows for incorporating covariances. The unknown scaling factor q was estimated with the maximum likelihood by assuming $\boldsymbol{\varepsilon}$ and $\boldsymbol{v}$ come from a multivariate normal distribution. The form of the joint log-likelihood function of $\boldsymbol{\varepsilon}$ and $\boldsymbol{v}$ is $L \propto \frac{1}{2}\left[\log(|\Sigma|) + (\boldsymbol{A}-\boldsymbol{a})^{'}\Sigma^{-1}(\boldsymbol{A}-\boldsymbol{a}) + \frac{1}{q}\boldsymbol{a}^{'}\boldsymbol{C}^{'}\Omega^{-1}\boldsymbol{C}\boldsymbol{a} + \log(|q\Omega|)\right]$, and $\widehat{q}$, $\widehat{\boldsymbol{a}}$ and $\widehat{Var}(\widehat{\boldsymbol{a}})$ were simultaneously obtained by minimizing L through conducting a grid search for q over the range of 0.01 to 1 at 0.01 increment (Van Deusen, 2002). In this study, ME was applied to updating the initial annual estimates respectively obtained with the expansion estimator and the post-stratified estimator at the state- and county-levels. Under this construct, ME is unbiased because these two estimators are design-unbiased. Excellent discussions about the unbiasedness of ME and the correction when a bias does occur are given by Teräsvirta (1981), Toutenburg (1982) and Kangas (1991).

3.4. Comparison of inferential precision

The sampling error, $SE_{\%} = 100 \times \frac{\sqrt{\widehat{Var}(\widehat{A})}}{\widehat{A}}$, reflects the inferential precision on a percentage basis and enables comparisons among estimators. Given the fact that the updating procedure is unbiased, a smaller $SE_{\%}$ represents greater inferential precision, or equivalently, a less inferential uncertainty. The $SE_{\%}$ is a measure officially used by FIA as the basis for 67% and 95% confidence intervals which can be approximated under a normality assumption for the sampling distribution of estimates (Bechtold and Patterson, 2005; Hou et al., 2015). $SE_{\%}$ was evaluated for comparing the inferential precision obtained from the post-stratified estimator, expansion estimator, and the mixed estimator at the state- and county-levels.

Further, the sampling error standard, $SE_{\%STD} = 100 \times 0.03 \times \sqrt{\frac{404686}{\widehat{A}}}$, was evaluated for assessing if an $SE_{\%}$ met the official precision standard mandated by the Forest Service directive (USDA, 2008). This directive requires that, for area estimation, the $SE_{\%}$ shall not exceed 3% per 404686 ha (approximately one million acres) of forestland. Although achieving $SE_{\%STD}$ is not a specific objective of this study, $SE_{\%STD}$ reflects from an official reporting perspective the target precision for a forest inventory estimate. A satisfactory official inventory estimate in the USA, either on an annual or periodic basis, is expected to meet $SE_{\%} \leq SE_{\%STD}$.

4. Results and discussion

4.1. Annual inventories at the state-level

Although the national forest inventory in the USA has been in a transition from the periodic to an annual system, the estimates used for official reporting purposes do not have an annual temporal resolution and rely on pooling all sample plots measured during an inventory cycle of five or 10 years (Bechtold and Patterson, 2005). An important motif of using pooled panels roots in meeting the mandated $SE_{\%STD}$ per the Forest Service directive, and therefore effective estimators that support annual inferences to satisfy this criterion without altering current FIA protocols would contribute to the success of this transition (McRoberts, 2001; Roesch, 2009; Van Deusen, 2002).

The estimates obtained with the post-stratified and expansion estimators for MN, GA and CA using pooled and annual FIA data collected from an inventory cycle of five or 10 years are summarized in Table 1. The estimates using the pooled data are categorized by "all-yr" (all-years). For the post-stratified estimator, the "all-yr" estimates of respective states were identical to those appearing in official reports (Brandeis, 2015; Christensen et al., 2016; Miles and Kepler, 2017). The $SE_{\%STD}$ criterion was met for MN and GA, but not for CA even with 10 years of pooled data. This would, on one hand, suggest extraordinary heterogeneity in the CA data, and, on the other hand, motivate steps to improve the inferential precision. Although the estimates $\widehat{A}_{EXP}$ were close to $\widehat{A}_{STR}$, the $SE_{\%}$ of the expansion estimators was 42% to 108% larger than those of the post-stratified estimator, violating $SE_{\%} \leq SE_{\%STD}$ for the three states.

The annual estimates using the FIA panel data are categorized by respective years. For the post-stratified estimator, the annual $\widehat{A}_{STR}$ estimates were somewhat erratic but still similar to each other, indicating a minor variation in areal changes over the years. This was also reflected by the annual $\widehat{A}_{EXP}$ estimates. For $\widehat{A}_{STR}$, $SE_{\%} \leq SE_{\%STD}$ was achieved for pooled estimates in MN and GA, but not for CA, nor for $\widehat{A}_{EXP}$ estimates in any of the three states. For both estimators, $SE_{\%} \leq SE_{\%STD}$ was not satisfied for any annual estimates, for any of the three states, suggesting that the use of design-based inference alone would not be an efficient or viable option. Nevertheless, it is interesting to conclude that the $SE_{\%}$ of an individual year is about $\sqrt{T}$ times the $SE_{\%}$ of "all-yr" for the post-stratified estimator, and about T times for the expansion estimator, reiterating the benefit of post-stratification from improving the precision.

4.2. Updated annual inventories at the state-level

The annual estimates updated by ME based on the initial post-stratified and expansion estimations for MN, GA and CA are summarized in Table 2. Two updating phenomena were noted. First, the updating procedure was quite effective for increasing inferential precision. For the post-stratified estimator, the updated annual estimates mostly met $SE_{\%} \leq SE_{\%STD}$ for MN and GA, with the $SE_{\%}$ reduced by 55% on average from the initial estimates (Table 1). Although for CA this criterion was not satisfied, the $SE_{\%}$ was significantly smaller, by 66% on average, illustrating the power of data assimilation. Similarly, for the expansion estimator, the $SE_{\%}$ was reduced by 54% to 66% from the initial estimates for the three states, but none of them met $SE_{\%} \leq SE_{\%STD}$.

Second, the smaller the initial $SE_{\%}$ obtained with an estimator, the smaller the $SE_{\%}$ obtained with the updating procedure, which is consistent with our findings for a different type of data assimilation being temporally invariant (Hou et al., 2019). After updating, the $SE_{\%}$ for an individual year is slightly larger than the $SE_{\%}$ of "all-yr" for the post-stratified estimator – more so for CA than for MN and GA, and about $T/2$ times larger than for the expansion estimator. An estimator used for obtaining the initial estimates also reflects the magnitude of inferential precision after implementing the updating procedure. Therefore, the estimator selected for making the initial estimation should be as precise as possible, which can be a rule of thumb for selecting candidate estimators in preparation for the data assimilation.

4.3. Temporal trends in forestland areas

The temporal trends in forestland areas over an inventory cycle of

Table 1
Initial estimates for Minnesota (MN), Georgia (GA) and California (CA) based on annual and pooled (all-yr) FIA data collected for a complete inventory cycle of five or ten years.

State	Year	n	Expansion estimation				Post-stratified estimation			
			$\hat{A}_{EXP}$ha	$\widehat{Var}\left(\hat{A}_{EXP}\right)$	$SE_{\%}$	$SE_{\%STD}$	$\hat{A}_{STR}$ha	$\widehat{Var}\left(\hat{A}_{STR}\right)$	$SE_{\%}$	$SE_{\%STD}$
MN	all-yr	17764	6740546	5256399726	1.08	0.74	6997405	1331732401	0.52	0.72
	2006	3546	6469742	132047887058	5.62	0.75	6790954	6314979330	1.17	0.73
	2007	3550	7062823	131793482602	5.14	0.72	7155721	6842220410	1.16	0.71
	2008	3551	6926595	131621671743	5.24	0.73	7036936	6354969624	1.13	0.72
	2009	3563	6529995	130750051814	5.54	0.75	6972608	7479573380	1.24	0.72
	2010	3554	6714018	131351820595	5.40	0.74	7001025	7122041508	1.21	0.72
GA	all-yr	6323	10087766	7213516937	0.84	0.60	10037894	3514922164	0.59	0.60
	2005	1275	10302876	177699233521	4.09	0.59	10036741	16178157609	1.27	0.60
	2006	1290	9939856	173496369761	4.19	0.61	10082548	18521115004	1.35	0.60
	2007	1281	10330597	176086808266	4.06	0.59	10065393	17779467554	1.32	0.60
	2008	1224	9807896	192957710052	4.48	0.61	10007579	19376082853	1.39	0.60
	2009	1253	10046290	183817015835	4.27	0.60	10014772	17810295395	1.33	0.60
CA	all-yr	16740	12957850	21757827197	1.14	0.53	13297969	6753369244	0.62	0.52
	2001	1713	13123324	2078351256595	10.99	0.53	13104235	66599498652	1.97	0.53
	2002	1677	13305233	2169128372414	11.07	0.52	13678681	68095577961	1.91	0.52
	2003	1672	12756528	2181629623118	11.58	0.53	12959839	68750679939	2.02	0.53
	2004	1677	13036517	2168439149784	11.30	0.53	13166207	73921199944	2.07	0.53
	2005	1677	13582265	2170739397142	10.85	0.52	13409791	67081921449	1.93	0.52
	2006	1700	13127039	2110276072273	11.07	0.53	13159545	65841613369	1.95	0.53
	2007	1685	12850706	2147917604148	11.40	0.53	12769068	73927418753	2.13	0.53
	2008	1658	12748697	2218658711914	11.68	0.53	13054659	71476538933	2.05	0.53
	2009	1655	12418151	2228216360656	12.02	0.54	12581240	66309590568	2.05	0.54
	2010	1626	12580213	2307494153021	12.07	0.54	12962762	62695908898	1.93	0.53

Table 2
Updated annual estimates for Minnesota (MN), Georgia (GA) and California (CA) based on the initial estimates of the post-stratified and expansion estimations.

State	Year	n	Updated expansion estimation				Updated post-stratified estimation			
			$\hat{A}_{ME}$ha	$\widehat{Var}\left(\hat{A}_{ME}\right)$	$SE_{\%}$	$SE_{\%STD}$	$\hat{A}_{ME}$ha	$\widehat{Var}\left(\hat{A}_{ME}\right)$	$SE_{\%}$	$SE_{\%STD}$
MN	2006	3546	6739576	27852683905	2.48	0.75	6987140	1434036715	0.54	0.73
	2007	3550	6742269	27080350123	2.44	0.72	6989265	1396116817	0.53	0.71
	2008	3551	6741758	26824100123	2.43	0.73	6989694	1385593092	0.53	0.72
	2009	3563	6739413	27071832342	2.44	0.75	6989642	1404050590	0.54	0.72
	2010	3554	6739162	27838923370	2.48	0.74	6989754	1446900509	0.54	0.72
GA	2005	1275	10093234	38132719235	1.93	0.59	10041990	3778492848	0.61	0.60
	2006	1290	10091187	37129067692	1.91	0.61	10042050	3678170630	0.60	0.60
	2007	1281	10090646	36830241194	1.90	0.59	10041719	3648825200	0.60	0.60
	2008	1224	10087423	37263804513	1.91	0.61	10041100	3691655975	0.61	0.60
	2009	1253	10087016	38349530804	1.94	0.60	10040840	3795255326	0.61	0.60
CA	2001	1713	12980773	274067200254	4.03	0.53	13107011	8665995301	0.71	0.53
	2002	1677	12979285	257900137826	3.91	0.52	13107040	8156196300	0.69	0.52
	2003	1672	12974510	246291718266	3.83	0.53	13101297	7789470727	0.67	0.53
	2004	1677	12971931	238885475339	3.77	0.53	13096643	7535556344	0.66	0.53
	2005	1677	12968703	235342500333	3.74	0.52	13091789	7413096292	0.66	0.52
	2006	1700	12959599	235531484746	3.74	0.53	13083903	7409531518	0.66	0.53
	2007	1685	12948630	239569322939	3.78	0.53	13074200	7537786845	0.66	0.53
	2008	1658	12938310	247769932204	3.85	0.53	13067768	7775586897	0.67	0.53
	2009	1655	12929850	260208600884	3.95	0.54	13061923	8116537830	0.69	0.54
	2010	1626	12926388	277927940231	4.08	0.54	13060941	8577361543	0.71	0.53

five years for MN and GA and 10 years for CA are graphed in Fig. 1 using the initial and updated annual estimates, with the shaded areas representing the 95% confidence intervals. The left panels of subgraphs A, C and E contrast the initial with the updated trends, and the right panels of subgraphs B, D and F contrast the updated trends with the flat lines using "all-yr" estimates.

The initial trends were somewhat erratic between adjacent years, although generally flat across the complete inventory cycles. Fluctuations in trend lines and confidence intervals are inherent in both the panel data and the estimator. Compared with the expansion estimator, the trend lines and confidence intervals were smoother and narrower for the post-stratified estimator. However, whether the change in forestland areas was truly fluctuated is questionable per the initial estimates due to the large confidence intervals, apart from the fact that forestland was less likely to experience drastic gain and loss under natural or undisturbed conditions.

Compared with the initial trends, the ME updated estimates were much smoother trend lines and narrower confidence intervals that better depict credible changes in forestland areas for the three states. The updated trend lines and confidence intervals were within the confidence intervals of the initial estimation, and to a large extent intersected or overlapped the "all-yr" results, suggesting improved credibility and consistency of the updated trends. This advancement would better serve the gain-loss approach for estimating emissions and removals in greenhouse gas inventories for which forestland area is a key variable (GFOI, 2016; IPCC, 2006).

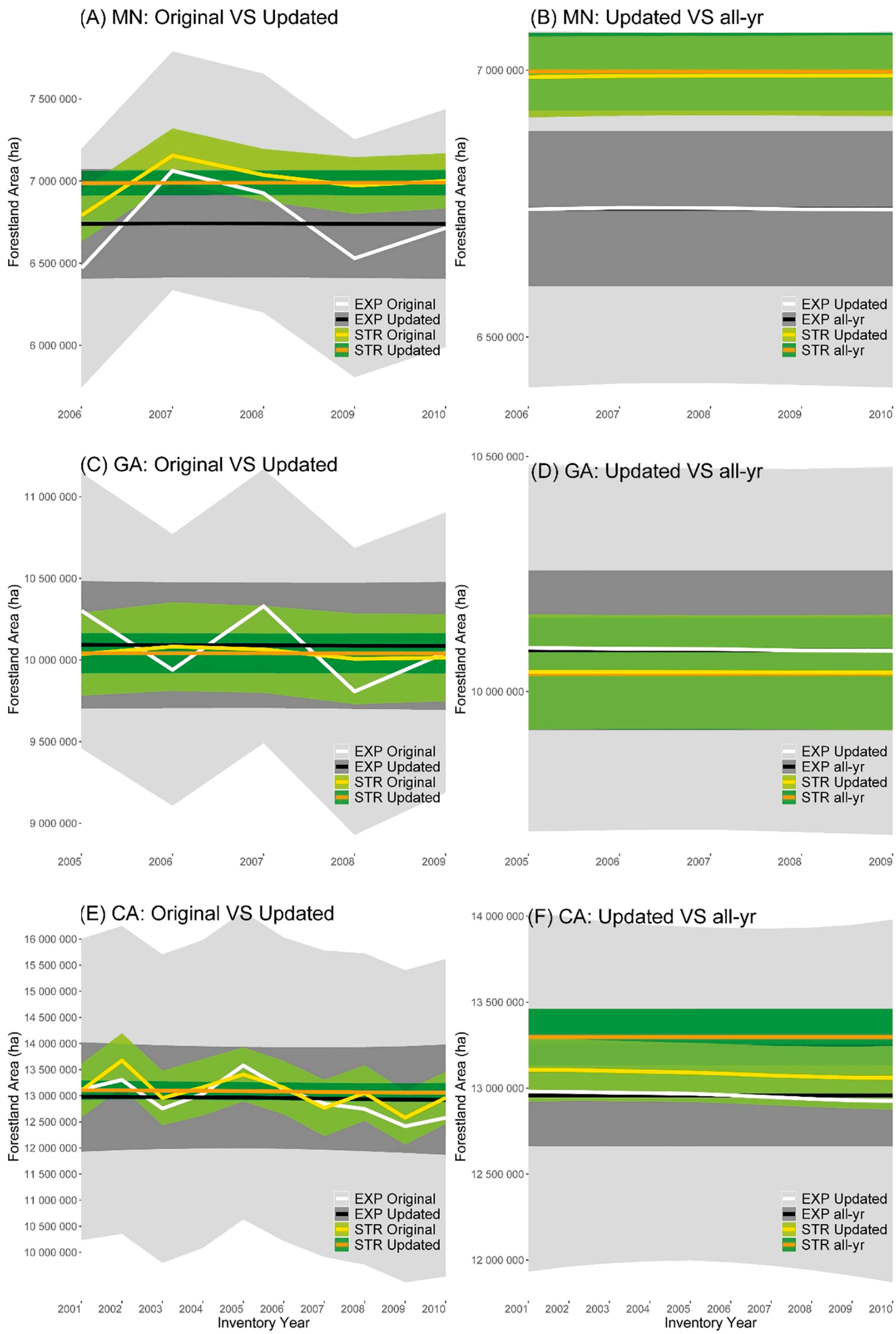

Fig. 1. Temporal trends in forestland areas over an inventory cycle of five- or ten-years for Minnesota (MN), Georgia (GA) and California (CA). Shaded areas represent the 95% confidence intervals. Note the scale of y-axis between the left and right panels is different.

Although the present study did not focus on analyzing ecological and human drivers behind the temporal changes, the updated estimates and trends are in a better position to serve this purpose by constraining the propagation of statistical uncertainty to subsequent analyses (Hou et al., 2017, 2018; McRoberts, 2006, 2011). This is particularly useful for FIA data users such as forest ecologists and decision-makers for elaborating reliable conclusions and consistent policies (Frické, 2009; Rowley, 2007).

Except for CA where the updated trend lines over a duration of 10 years indicated a clear decrease in forestland areas, the forestland areas are relatively stable for MN and GA. Wildfires seem to be responsible for the decrease in CA (Calfire, 2019), but wildfires typically have no effect on FIA estimates of forestland area because FIA's definition includes land use. Unless wildfires – or any other canopy disturbance are directly associated with a permanent land use change, those areas will continue to be labeled as forestland. It is more feasible that land use changes like housing development or conversion to agriculture are affecting permanent loss of forests.

Technically, the burn-in process (Fox et al., 2018) for the ME updating procedure was instant. This partially supports the statement in Van Deusen (1999) that for assimilating a short time series such as an inventory cycle of five or10 years, ME is more efficient than the Kalman filter, because the Kalman filter requires using unknown values to initiate the burn-in process before stabilizing. However, for a long time series, typically longer than 20 years, the Kalman filter may yield better performance (Van Deusen, 1999), despite that ME is essentially equivalent to the Kalman filter under specific assumptions (Dixon and Howitt, 1979).

4.4. Annual estimates and updated annual estimates at the county-level

The annual county-level estimates before and after implementing the updating procedure are reported in Fig. 2. For simplicity, results are graphed just for MN because similar findings apply to GA and CA. Five

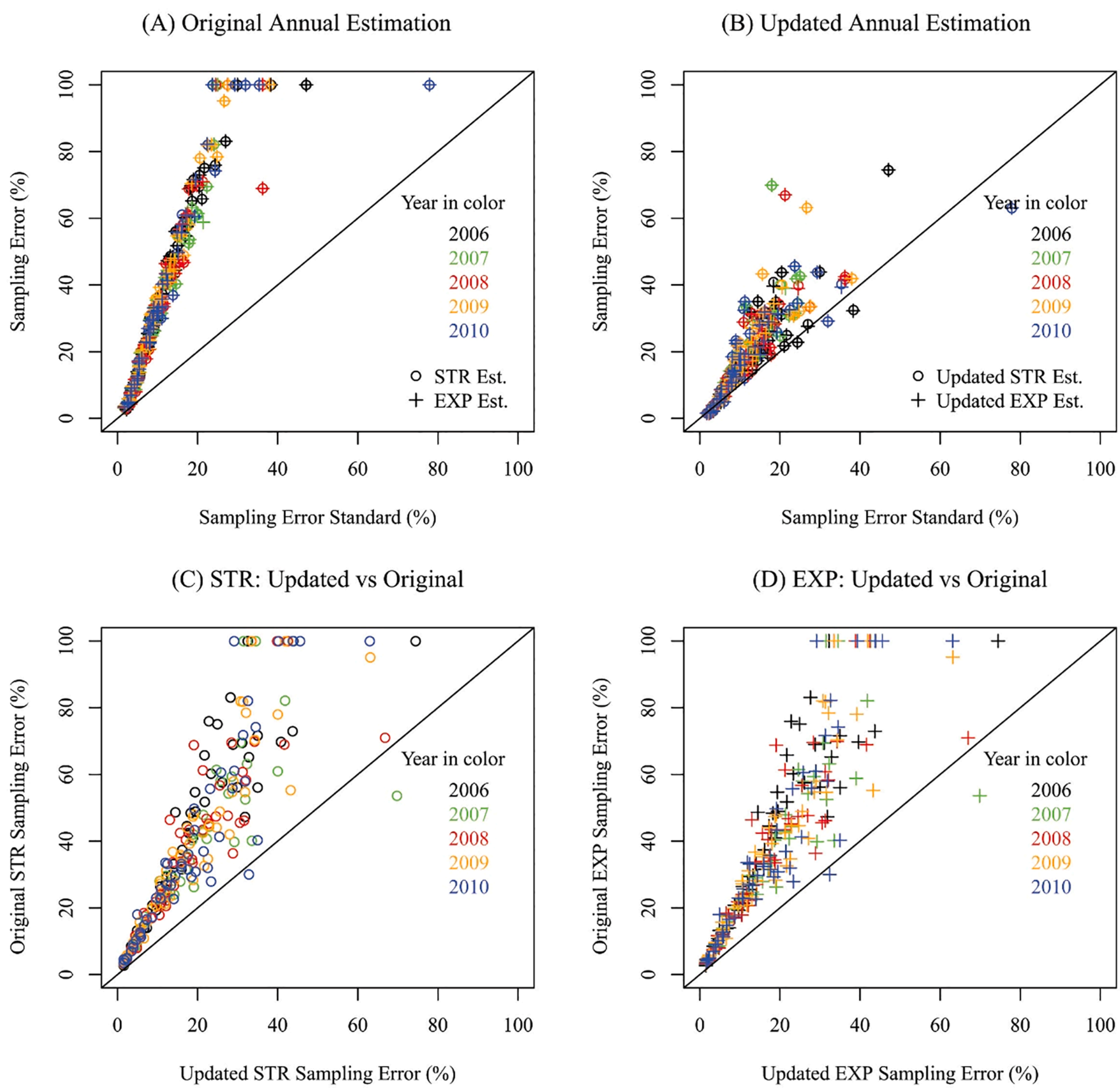

Fig. 2. County level estimation for Minnesota. Each circle or cross represents a county.

findings are relevant at the county-level. First, there are 87 counties in MN of which 31 were excluded from the analysis because no data existed for one or more years. A list of these counties missing data can be extracted from the FIA database using EVALIDator (2019). Furthermore, 62 of 87 MN counties contained fewer than eight forested plots during one or more years; eight plots is a suggested minimum sample size for producing FIA estimates. When all panel years were pooled, no counties contained zero forested plots, and only 28 of 87 counties did not meet or exceeded an eight-plot minimum. This emphasizes the point that even in the best of conditions where MN is the only FIA state with both a 5-year cycle and double intensity, per-county sample sizes are still limiting. These counties were relatively small and in sparsely forested portions of the State, causing counties to have insufficient sample sizes (Bechtold and Patterson, 2005). FIA panel data are generally adequate for making county-level inference for pooled panels, but often not for annual estimates.

Second, the difference in $SE_{\%}$ between the post-stratified and expansion estimators dissipated at the county-level (Fig. 2A). As the sample size decreases, the inferential precision by the post-stratified estimator reduces to that for the expansion estimator, because the hierarchical structures in the sample data became less distinguishable from a single stratum. Third, compared with the state-level estimation, a greater allowance for $SE_{\%}$ in smaller areas at the county-level reflects the decrease in sample size as the estimation area decreases. FIA inventories were primarily designed for large areas, and as the panel data are subdivided into smaller areas, the $SE_{\%}$ increases, as does $SE_{\%STD}$. Fourth, regardless, the updating procedure effectively reduced the $SE_{\%}$ from the initial estimates, and the smaller the initial $SE_{\%}$ obtained with an estimator, the smaller the $SE_{\%}$ obtained with the updating procedure, consistent with the state-level inferences (Fig. 2 C&D). Fifth, the criterion $SE_{\%} \leq SE_{\%STD}$ was not met for all counties using the post-stratified and expansion estimators, or for most counties using the mixed estimator (Fig. 2B).

A number of challenges arise in the county-level estimations. These include the missing data issue and the reduced precision for small areas. With the exception of enlarging the sample size, which unfortunately is rarely considered in a national forest inventory context, it is worthwhile to consider statistical techniques employing missing data recovery, small-area estimation, model-assisted inference, and/or a joint use of these to obtain the initial estimates (Battese et al., 1988; Fuller and Harter, 1987; Henderson et al., 1959; Prasad and Rao, 1990). Data assimilation techniques can subsequently be used to update these estimates.

5. Conclusions

A distinguishing feature of this updating procedure resides in its compatibility with the FIA program by not imposing any modifications to the data collection protocols, and in its compatibility with alternative design- or model-based inferences for making the initial estimation. A highlight of the mixed estimator is its unbiasedness for updating design-based inference (Kangas, 1991); however, when updating model-based inference, cautions must be exercised for uncertainties related to adopting auxiliary data from remote sensing as an example (Hou et al., 2017; Xu et al., 2018), because bias associated with these uncertainties will propagate to the mixed estimator, requiring a correction as a result (Teräsvirta, 1981).

Five conclusions are relevant. First, the sampling error, $SE_{\%}$, of an annual inventory was significantly smaller for the post-stratified estimator than the expansion estimator at the state-level, but similar at the county-level. Second, only statewide estimates for MN and GA, based on post-stratified estimation and pooled across all years, met $SE_{\%} \leq SE_{\%STD}$, a criterion indicating that, for area estimation, the $SE_{\%}$ shall not exceed 3% per 404686 ha (ca. one million acres). However, the statewide expansion estimates did not meet this standard, nor did either design-based estimator for any statewide annual estimate, suggesting that design-based inference alone would not be an efficient option for making annual inventories using FIA panel data, especially at the county-level. Third, at both population levels (state and county), the updating procedure using the mixed estimator effectively reduced the initial annual $SE_{\%}$ obtained with the stratified and expansion estimators over 55% on average, as well as achieving updated annual $SE_{\%}$ comparable with the $SE_{\%}$ based on using a five- or ten-years pooled FIA dataset. Fourth, a rule of thumb is that candidate estimators selected for making the initial estimation should be as precise as possible, because the estimator used for obtaining the initial estimates also decides the magnitude of inferential precision after implementing the updating procedure. Last but not the least, compared with the initial temporal trends, the ME updated trends were of much smoother trend lines and narrower confidence intervals that better depict temporal trends in areal changes for a population in question.

CRediT authorship contribution statement

Zhengyang Hou: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Writing - original draft, Writing - review & editing. **Grant M. Domke:** Conceptualization, Data curation, Funding acquisition, Project administration. **Matthew B. Russell:** Conceptualization, Data curation, Funding acquisition, Project administration. **John W. Coulston:** Supervision, Validation. **Mark D. Nelson:** Validation. **Qing Xu:** Formal analysis, Funding acquisition, Methodology, Project administration, Writing - original draft, Writing - review & editing. **Ronald E. McRoberts:** Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Dr. Zhengyang Hou and Dr. Qing Xu were supported by the National Natural Science Foundation of China (Grant No. 32001252) and the International Center for Bamboo and Rattan (Grant No. 1632020029). This work was also supported by the U.S. Department of Agriculture Forest Service - Northern Research Station (No. 14-JV-11242305-012) and the Minnesota Agricultural Experiment Station (No. MIN-42-063). Dr. Margarida Tomé and two anonymous reviewers are sincerely acknowledged for their excellent services and helpful comments.

References

Battese, G.E., Harter, R.M., Fuller, W.A., 1988. An error-components model for prediction of county crop areas using survey and satellite data. J. Amer. Statist. Assoc. 83, 28–36.

Bechtold, W.A., Patterson, P.L., 2005. The enhanced forest inventory and analysis program – national sampling design and estimation procedures. General Technical Report SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC. Available online at: https://www.fs.usda.gov/treesearch/pubs/20371.

Blayo, E., Bocquet, M., Cosme, E., Cugliandolo, L., 2014. Advanced data assimilation for geosciences. Oxford University Press.

Brandeis, T.J., 2015. Georgia's forests, 2009. Resource Bulletin SRS-207. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC.

Brewer, K.R.W., Gregoire, T.G., 2009. Introduction to survey sampling. Handbook of Statistics – Vol 29A. Sample Surveys: Design, Methods and Applications. Atlanta: Elsevier B.V.

Calfire, 2019. Stats and events. Available online at https://www.fire.ca.gov/stats-events/.

Cassel, C.M., Särndal, C.E., Wretman, J.H., 1977. Foundations on inference in survey sampling. John Wiley & Sons, New York.

Christensen, G.A., Waddell, K.L., Stanton, S.M., Kuegler, O., 2016. California's forest resources: Forest Inventory and Analysis, 2001–2010. Gen. Tech. Rep. PNW-GTR-913. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Cochran, W.G., 1977. Sampling Techniques. Wiley, New York.

Czaplewski, R.L., 1990. Kalman filter to update forest cover estimates. In: Labau, V.J., Cunia, T. (Eds.), State-of-the-art methodology of forest inventory. USDA For. Serv., Pacific Northwest Research Station.
Dixon, B., Howitt, R., 1979. Continuous forest inventory using a linear filter. Forest Sci. 25, 675–698.
Durbin, J., 1953. A note on regression when there is extraneous information about one of the coefficients. J. Amer. Statist. Assoc. 48, 799–808.
Edgar, C.B., Westfall, J.A., Klockow, P.A., Vogel, J.G., Moore, G.W., 2019. Interpreting effects of multiple, large-scale disturbances using national forest inventory data: A case study of standing dead trees in east Texas, USA. Forest Ecol. Manag. 437, 27–40.
EVALIDator, 2019. Forest inventory and analysis national program - data and tools. Available online at https://www.fia.fs.fed.us/tools-data/.
FIA, 2019. Forest Inventory and Analysis National Program. Available online at https://www.fia.fs.fed.us/tools-data/default.asp.
FIADB, 2018. The Forest Inventory and Analysis Database: Database Description and User Guide for Phase 2 (Version 8.0). Available online at https://www.fia.fs.fed.us/library/database-documentation/index.php.
FIAFG, 2019. Forest Inventory and Analysis, National Core Field Guide, Version 9.0, Volume I: Field Data Collection Procedures for Phase 2 Plots. Available online at https://www.fia.fs.fed.us/library/field-guides-methods-proc/index.php.
Fletcher, S., 2017. Data assimilation for the geosciences: from theory to applications. Elsevier.
Fox, A.M., Hoar, T.J., Anderson, J.L., Arellano, A.F., Smith, W.K., Litvak, M.E., MacBean, N., Schimel, D.S., Moore, D.J.P., 2018. Evaluation of a data assimilation system for land surface models using CLM4.5. J. Adv. Model. Earth Syst. 10, 2471–2494.
Frické, M., 2009. The knowledge pyramid: a critique of the DIKW hierarchy. J. Inform. Sci. 35, 131–142.
Fuller, W.A., Harter, R.M., 1987. The multivariate components of variance model for small area estimation. Small Area Statistics. Wiley, New York.
GFOI, 2016. Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance from the Global Forest Observations Initiative, Edition 2.0. Food and Agriculture Organization, Rome. Available online at: https://www.reddcompass.org/download-the-mgd.
Ghosh, D., Vogt, A., 1993. Some theorems relating post-stratification and sample configuration. Am. Stat. Assoc. Proc. Sect. Surv. Res. Methods 341–345.
Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized linear regression model. J. Amer. Statist. Assoc. 57, 369–375.
GHRRP, 2019. Greenhouse Gas Reporting Program. Available online at https://www.epa.gov/ghgreporting.
Henderson, C.R., 1950. Estimation of genetic parameters. Ann. Math. Statist. 21, 309–310.
Henderson, C.R., Kemfthorne, O., Searle, S.R., Von Krosigk, C.N., 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 13, 192–218.
Henderson, C.R., 1963. Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding. Nat. Acad. Sci., Nat. Res. Council, Publication 982, Washington, DC.
Henderson, C.R., 1973. Sire evaluation and genetic trends. In: Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush 10-41. Amer. Soc. Animal Sci. -Amer. Dairy Sci. Assn. -Poultry Sci. Assn., Champaign, Ill.
Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423–447.
Holt, D., Smith, T.M.F., 1979. Post stratification. J. Roy. Statist. Soc. A 142, 33–46.
Hou, Z., McRoberts, R.E., Ståhl, G., Packalen, P., Greenberg, J., Xu, Q., 2018. How much can natural resource inventory benefit from finer resolution auxiliary data? Remote Sens. Environ. 209, 31–40.
Hou, Z., Mehtätalo, L., McRoberts, R.E., Ståhl, G., Tokola, T., Rana, P., Siipilehto, J., Xu, Q., 2019. Remote sensing-assisted data assimilation and simultaneous inference for forest inventory. Remote Sens. Environ. 234, 1114431.
Hou, Z., Xu, Q., Hartikainen, S., Antilla, P., Packalen, T., Maltamo, M., Tokola, T., 2015. Impact of plot size and spatial pattern of forest attributes on sampling efficacy. Forest Sci. 61, 847–860.
Hou, Z., Xu, Q., McRoberts, R.E., Greenberg, J.A., Liu, J., Heiskanen, J., Pitkänen, S., Packalen, P., 2017. Effects of temporally external auxiliary data on model-based inference. Remote Sens. Environ. 198, 150–159.
IPCC, 2006. Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. (Eds.), 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use. Institute for Global Environmental Strategies, Hayama, Japan. Available online at: http://www.ipccnggip.iges.or.jp/public/2006gl/index.html.
Katzfuss, M., Stroud, J.R., Wikle, C.K., 2016. Understanding the Ensemble Kalman Filter. Am. Stat. 70, 350–357.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng.-T ASME 82 (Series D), 35–45.
Kangas, A., 1991. Updated measurement data as prior information in forest inventory. Silva Fenn. 25, 181–191.
Lahoz, W., Khattatov, B., Menard, R., 2010. Data assimilation: making sense of observations. Springer.
Lahoz, W.A., Schneider, P., 2014. Data assimilation: making sense of earth observation. Front. Environ. Sci. 2 (article 16), 1–28.
Little, R.J.A., 1993. Post-stratification. A modeler's perspective. J. Amer. Statist. Assoc. 88, 1001–1012.
McRoberts, R.E., 1999. Joint annual forest inventory and monitoring system: the North Central perspective. J. For. 97, 27–31.
McRoberts, R.E., 2001. Imputation and model-based updating techniques for annual forest inventories. For. Sci. 47, 322–330.
McRoberts, R.E., 2006. A model-based approach to estimating forest area. Remote Sens. Environ. 103, 56–66.
McRoberts, R.E., Hansen, M.H., Smith, W.B., 2010. United States of America. In: Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (Eds.), National Forest Inventories, Pathways for Common Reporting. Springer.
McRoberts, R.E., 2011. Satellite image-based maps: scientific inference or pretty pictures? Remote Sens. Environ. 115, 715–724.
Miles, P.D., Kepler, D., 2017. Forests of Minnesota, 2015. Resource Update FS-105. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station.
NRSF, 2019. National Report on Sustainable Forests. Available online at https://www.fs.fed.us/research/sustain/national-report/.
Patterson, P.L., Reams, G.A., 2005. Combining panels for forest inventory and analysis estimation. In: Bechtold, W.A., Patterson, P.L. (Eds.),The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures. General Technical Report SRS-80, Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. Available online at: https://www.fs.usda.gov/treesearch/pubs/20371.
Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of mean square error of small-area estimators. J. Amer. Statist. Assoc. 85, 163–171.
Robinson, G.K., 1991. That BLUP is a Good Thing: the estimation of random effects. Stat. Sci. 6, 15–32.
Roesch, F., 2009. Spatial-temporal models for improved county-level annual estimates. In: McWilliams, W., G. Moisen, R. Czaplewski, comps. Forest Inventory and Analysis (FIA) symposium 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.
Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. J. Inform. Sci. 33, 163–180.
Royall, R.M., Herson, J., 1973. Robust estimation in finite populations I. J. Amer. Statist. Assoc. 68, 880–889.
Teräsvirta, T., 1981. Some results on improving the least squares estimation of linear models by mixed estimation. Scand. J. Stat. 8, 33–38.
Theil, H., 1963. On the use of incomplete prior information in regression analysis. J. Amer. Statist. Assoc. 58, 401–414.
Theil, H., 1971. Principles of Econometrics. John Wiley & Sons, New York.
Theil, H., Goldberger, A.S., 1961. On pure and mixed statistical estimation in economics. Int. Econ. Rev. 2, 65–78.
Thompson, S.K., 2012. Sampling, third ed. Wiley, New Jersey, United States.
Toutenburg, H., 1982. Prior information in linear models. John Wiley & Sons.
USDA, 2008. Accuracy standards. In: Forest Survey Handbook. FSH 4809.11, Chapter 10. Washington, DC: U.S. Department of Agriculture, Forest Service.
Van Deusen, P.C., 1991. Trend monitoring with varying coefficient models. Forest Sci. 37, 1365–1375.
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, 1824–1828.
Van Deusen, P.C., 2002. Comparison of some annual forest inventory estimators. Can. J. For. Res. 32, 1992–1995.
Xu, Q., Man, A., Fredrickson, M., Hou, Z., Pitkänen, J., Wing, B., Ramirez, C., Li, B., Greenberg, J.A., 2018. Quantification of uncertainty in aboveground biomass estimates derived from small-footprint airborne LiDAR. Remote Sens. Environ. 216, 514–528.