Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Hierarchical Bayesian models for small area estimation of county-level private forest landowner population

Journal:	*Canadian Journal of Forest Research*
Manuscript ID	cjfr-2017-0154.R1
Manuscript Type:	Article
Date Submitted by the Author:	31-Jul-2017
Complete List of Authors:	Ver Planck, Neil; Michigan State University, Forestry Finley, Andrew; Michigan State University, Forestry Huff, Emily; Michigan State University, Forestry
Keyword:	cadastral, private forest ownerships, small area estimation, hierarchical Bayes, spatial area-level model
Is the invited manuscript for consideration in a Special Issue? :	N/A

Hierarchical Bayesian models for small area estimation of county-level private forest landowner population

Neil R. Ver Planck*[1], Andrew O. Finley[1], and Emily S. Huff[1]

[1]Department of Forestry, Michigan State University,
480 Wilson Road Room 126, East Lansing, MI 48824-1222, USA

September 18, 2017

*Corresponding author. Email: verplan6@msu.edu

Abstract

The National Woodland Owner Survey (NWOS), administered by the USDA Forest Service, provides estimates of private forest ownership characteristics, and owners' attitudes and behaviors at a national-, regional- and state-level. Due to sample sizes prescribed for inference at the state-level, there are insufficient data to support county-level estimates. However, county-level estimates of NWOS variables are desired because ownership programs and education initiatives often occur at the county-level and such information could help tailor these efforts to better match county-specific needs and demographics. Here, we present and assess methods to estimate the number of private forest ownerships at the county-level for two states, Montana and New Jersey. To assess model performance, true population parameters were derived from cadastral and remote sensing data. Two small area estimation (SAE) models, the Fay-Herriot (FH) and FH with conditional autoregressive random effects (FHCAR), improved estimated county-level population mean squared error (MSE) over that achieved by direct estimates. The proposed SAE models use covariates to improve accuracy and precision of county-level estimates. Results show total forest area and 2010 decennial Census population density covariates explained a significant portion of variability in county-level population size. These and other results suggest the proposed SAE methods yield a statistically robust approach to deliver reliable estimates of private ownership population size and could be extended to additional important NWOS variables at the county-level.

Keywords. cadastral, hierarchical Bayes, private forest ownerships, small area estimation, spatial area-level model

1 Introduction

There are approximately 192 million ha of private forest land in the United States (US) owned by corporations, families, individuals, non-governmental organizations, and tribal entities (Butler et al., 2016c). This forested land provides many social and ecosystem benefits, but is managed by millions of owners with potentially disparate goals and objectives (Butler et al., 2016a). Therefore, a deeper understanding of the demographics, attitudes, and management behaviors of private forest landowners is paramount to designing effective incentives, outreach programs, and support mechanisms that enable these owners to engage in sustainable forest management activities. The USDA Forest Service Forest Inventory and Analysis (FIA) National Woodland Owner Survey (NWOS) is the primary source of information about national-, regional-, and state-level private forest characteristics and owners' demographics, attitudes, and behaviors. NWOS results can be summarized by ownership type, e.g., corporate, family, other private, and tribal and is currently implemented every five years (Butler et al., 2016c).

Given limited resources to conduct the NWOS and low to moderate response rates—that are common in social surveys—small sample size restricts reliable inference to state-level estimates. Although state-level estimates are informative, county-level estimates might be more appropriate and effective when designing and delivering education and outreach to private forest owners. Furthermore, from a management and conservation standpoint, county-level estimates are more useful for tracking trends in private forest land parcellation, fragmentation, composition, and ownership demographics and characteristics (see, e.g., Kittredge et al., 2008; Pan et al., 2009; Poudyal and Hodges, 2009). Results from these, and similar, studies underscore the value of ownership information at fine spatial scales to encourage sustainable forest management and curb loss of ecosystem services. To this end, we present and assess methods to estimate the number of private forest ownerships at the county-level using the current NWOS sampling design.

The NWOS's target sample size is 250 respondents per state (Butler et al., 2016a) and a required 100 respondent minimum to generate state survey summaries (Butler et al., 2016b). Again, such small state-level sample sizes effectively preclude robust inference at county-levels, particularly when reporting results by ownership type. Small sample sizes can result in undesirably low precision of parameter estimates within a design-based framework (e.g., the NWOS design). Here, we refer to parameter estimates obtained using design-based estimators as *direct estimates*. These direct estimates are the current standard used in operational settings that employ design-based estimators; therefore, the direct estimates will be used for benchmarking against model-based approaches. Small area estimation (SAE) is a model-based approach that couples a direct estimate and possible covariates to improve the estimate precision and, in some cases, accuracy. Rao and Molina (2015) provide an excellent review of available SAE methods. Unlike a standard linear regression, the SAE framework is comprised of two component models: a sampling and a linking model (You and Zhou, 2011). The estimation of the SAE parameter of interest accounts for and balances between the sampling (i.e., direct estimator) and linking model errors. The linking model is a linear model with random effects that relate the small areas of interest with some error. Additional spatial structure may still remain in the linking model after accounting for possible covariates. Such residual structure can be further modeled using spatial random effects.

SAE is also of great interest to users of the core FIA biophysical variables with the small area being dependent on the application. The modeling framework applied here for private forest ownerships could easily be adapted to these biophysical variables. In the case of biophysical variables, several recent forestry studies, e.g., Goerndt et al. (2013) and Magnussen et al. (2014), use SAE to improve inference at county- and municipal-levels. A thorough literature review yielded no application of SAE to private forest landowners or related studies.

The primary contributions of this work are: *i*) producing county-level private forest ownership datasets for Montana and New Jersey; *ii*) defining and assessing SAE models to improve county-level inference of the number of private forest ownerships; *iii*) developing open source software to fit proposed SAE models.

The remainder of the manuscript is organized as follows. In Section 2 we detail the steps followed to create a spatially explicit private forest ownership dataset for Montana and New Jersey. Then we define the direct estimator for the number of county-level private forest ownerships along with two SAE models. We then describe our approach for comparing the proposed SAE models. Results are given in Section 3 that is followed by discussion and future directions in Section 4.

2 Materials and methods

2.1 True number of private forest ownerships

Recently, several states, including Montana and New Jersey, made cadastral data freely available for determining property ownership. Using these 2015 cadastral data and a geographic information system (GIS) based analysis defined in Ver Planck et al. (2016), we created a GIS layer that delineates our working definition of private forest. Here, *private forest* comprises NWOS ownership categories of corporate, family, and other private (excluding tribal lands) (Butler et al., 2016c). The number of ownerships with forest property area greater than two ha within each county is the parameter of interest in the subsequent analysis. The derived GIS layers provide the true parameter value within each county, denoted Y_{Ti}, with subscript noting the true T number of private forest ownerships in county i.

Developing the private forest ownership GIS layers began by downloading freely available county-level cadastral data from the respective state repositories, Base Map Service Center Montana State Library (2015) for Montana and New Jersey Office of Information Technology, Office of GIS (2015) for New Jersey. These cadastral data were combined with remotely sensed forest landcover data from the National Landcover Dataset (NLCD; Jin et al., 2013) to determine forest ownerships greater than two ha. Forested areas were defined to be at least 20% tree cover, at least 0.41 ha in area, and 30 m wide based upon the NLCD specifications available.

To begin processing county-level properties, we identified ownerships in the dataset with unknown names. Parcels for which ownership was unknown that fell within public lands boundaries were assumed public. Additionally, the Protected Areas Database of the US (US Geological Survey, Gap Analysis Program, 2016) was used to identify non-governmental conservation organizations, e.g., the Nature Conservancy, listed as unknown in the cadastral dataset. All other unknown properties were assigned a unique identifier and treated as unique private ownerships. Remaining properties with known private ownership names were combined by owner(s) name and street address so multiple properties within a given county were treated as a single ownership. Ownerships with forest industry or tribal affiliations were omitted from the final county-level private forest ownership GIS layers to be consistent with the private forest landowner definition used in Ver Planck et al. (2016) for Montana. This portion of the analysis was completed using a combination of QGIS (QGIS Development Team, 2014) and `R` statistical software (R Core Team, 2014).

2.2 Models

2.2.1 Direct estimator

The direct estimator used for private forest ownership population at the state-level (Butler et al., 2016a; Dickinson and Butler, 2013) was applied to individual counties (note, notation was modified slightly for consistency with SAE models presented in Section 2.2.2). The direct estimator and associated variance are based on probability proportional to size sampling (Hansen and Hurwitz, 1943):

$$Y_i = \frac{A_i}{n_i} \sum_{j=1}^{n_i} \frac{d_{ij}}{a_{ij}}; \tag{1}$$

$$\sigma_i^2 = \frac{1}{n_i(n_i - 1)} \sum_{j=1}^{n_i} \left(\frac{d_{ij}}{p_{ij}} - Y_i \right)^2, \tag{2}$$

where i indexes county, Y_i is population total, A_i is total private forest area, n_i is number of samples with replacement, d_{ij} is an indicator variable that is one if the jth sample fell in private forest and zero otherwise, a_{ij} is the forest area (ha) of the sampled ownership, and

$p_{ij} = a_{ij}/A_i$ is the ownership selection probability. The steps for drawing the ownership samples are described in Section 2.3. The total county-level private forest area, A_i, was fixed to equal the true private forest area derived from the GIS layer in Section 2.1.

2.2.2 Small area estimation models

The direct estimator, Section 2.2.1, was log transformed to meet the SAE model normality assumption. This transformation was also desirable because it ensures SAE model population estimates have the correct support, i.e., positive, following back transformation. Taking the log of the direct estimator necessitates transformation of the associated variance, accomplished here via the delta method (Casella and Berger, 2002). The Fay-Herriot (FH) SAE model (Fay and Herriot, 1979) for county i in $1, 2, \ldots, m$ counties is defined:

$$
\begin{aligned}
\widetilde{Y}_i &= \tilde{\theta}_i + \epsilon_i, \\
\tilde{\theta}_i &= \mathbf{x}_i'\boldsymbol{\beta} + v_i,
\end{aligned}
\tag{3}
$$

where $\widetilde{Y}_i$ is the log-transformed direct estimator, $\tilde{\theta}_i$ is the log population total, and ϵ_i is a normally distributed error with mean zero and variance $\tilde{\sigma}_i^2 = \sigma_i^2/Y_i^2$. The additive mean of $\tilde{\theta}_i$ comprises an intercept and county-level covariates held in the $p \times 1$ column vector $\mathbf{x}$ and associated $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}$. The county-level random effects term v_i is normally distributed with mean zero and variance σ_v^2.

It is reasonable to think that forest and ownership patterns, e.g., property size and owner's socio-economic or demographic characteristics, could exhibit spatial autocorrelation. In our setting, if direct estimate values are spatially correlated, i.e., adjacent counties have similar population values, then we should exploit this relationship to further improve inference by pooling information across proximate counties. Counties could be represented as either point locations (i.e., their centroid) or as areal units for defining this spatial structure. Representing counties by their centroid may misrepresent distances among neighboring counties due to irregularly shaped counties; therefore, the areal approach is implemented as it maintains the desired neighborhood structure of the county lattice. Hence, we augment model (3) by adding a spatially structured random effect that follows a conditional autoregressive (CAR) prior distribution, see, e.g., Banerjee et al. (2014) and You and Zhou (2011). This extended model called FHCAR is defined analogous to FH, with the exception that the unstructured random effects $\mathbf{v} = (v_1, v_2, \ldots, v_m)' \sim N(\mathbf{0}, \sigma_v^2 \boldsymbol{I})$ in model (3) are replaced with $\mathbf{v} \sim N(\mathbf{0}, \boldsymbol{\Sigma}(\sigma_v^2, \lambda))$. Here, the $m \times m$ covariance matrix $\boldsymbol{\Sigma}(\sigma_v^2, \lambda) = \sigma_v^2 \left[\lambda \boldsymbol{R} + (1-\lambda)\boldsymbol{I}\right]^{-1}$, where σ_v^2 is the spatial variance parameter, λ is the autocorrelation parameter, $\boldsymbol{R}$ is the neighborhood matrix with diagonal elements equal to the number of neighbors and off diagonal elements equal negative one or zero indicating if a neighbor is present or not, and $\boldsymbol{I}$ is the $m \times m$ identity matrix. Counties are only considered neighbors with adjoining borders.

The FH and FHCAR Bayesian model specifications are completed by assigning prior distributions to parameters (Gelman et al., 2014). We selected non-informative priors for all model parameters. Each regression coefficient in $\boldsymbol{\beta}$ was assigned a flat prior distribution, σ_v^2 was given an inverse-Gamma (IG) prior distribution, and, following You and Zhou (2011), λ's prior was uniform with support between zero and one. The IG's shape hyperparameter was set to two, which results in a prior mean equal to the scale hyperparameter and infinite variance. The IG's scale hyperparameter was set as $\sum_{i=1}^{m} \tilde{\sigma}_i^2/m$ to give equal prior weight to the sampling and CAR variances. A Markov chain Monte Carlo (MCMC) algorithm was used to sample from parameters' posterior distributions. Specifically, a Gibbs algorithm was developed to sample from $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_m)$, $\boldsymbol{\beta}$, and σ_v^2 with full conditional distributions given in You and Zhou (2011), and a Metropolis-Hastings algorithm was used to sample from λ's posterior distribution. The data and associated code are available in Harvard Dataverse doi:10.7910/DVN/A3ROXD.

Parameter posterior inference was based on 3000 post burn-in MCMC samples from each of $L = 3$ chains resulting in $K = 9000$ samples. Chain mixing and convergence were diagnosed using a multivariate potential scale reduction factor of less than 1.1 for all parameters considered (Gelman et al., 2014). For our parameter of interest $\hat{\theta}_i$ (posterior mean of the population total), the K posterior samples of $\tilde{\boldsymbol{\theta}}$ were exponentiated back to the original units.

2.3 Simulation study

For the NWOS and similar efforts, SAE models are viable from a statistical perspective if they improve county-level estimate precision without inducing substantial bias. Given the actual private forest ownerships for Montana and New Jersey, Section 2.1, we examine SAE model inference against truth using a simulation study. One iteration in the simulation study produces a set of county-level direct and SAE model estimates by: *i*) drawing a random probability proportional to size sample from the private forest ownership list sample frame; *ii*) computing direct estimates (Section 2.2.1); *iii*) estimating FH and FH-CAR models (Section 2.2.2); *iv*) evaluating differences between SAE model population estimates and truth. Summarizing results from *iv* for a large number of iterations allows us to assess precision and bias in SAE model population estimates.

A county specific sample size, n_i, is needed to conduct step *i*. Here too, we want the sample size to approximate that achieved by the NWOS design. To determine the sample size in each county, we randomly located 1000 points within each hexagon of a tessellated hexagonal grid laid over the states of Montana and New Jersey using DGGRID developed by Sahr (2011). Each of the 1000 points represented a unique sample iteration across an individual state. The area of each hexagon was 2407 ha. This grid roughly approximates the one used to spatially distribute FIA and NWOS samples (Bechtold and Patterson, 2005; Dickinson and Butler, 2013). From each set of hexagons sampled within a county, we calculated the mean number of points that fell within private forest ownerships across the 1000 sample iterations. This mean value was rounded up and used as n_i. Given fixed n_i—to reduce variation among repeated samples—we repeated the simulation study steps *i*-*iii* $N = 4000$ times for Montana and New Jersey. These large number of repeated simulations should empirically show unbiased estimates of truth for the direct estimator.

2.3.1 County-level covariates

Exploratory analysis using linear regression models showed population density (PD; people km^{-2}) from the 2010 decennial Census (US Census Bureau, 2016), and NLCD total forest area (TFA; ha) explain significant variability in log-transformed direct estimates. For Montana, the linear regressions of the simulation runs explained 45% of the variation on average with a range from 27% to 65%. For New Jersey, the explained variation was higher with a mean of 76% and a range from 50% to 95%. Therefore, PD and TFA covariates were used for the simulation study.

2.4 Simulation summary

The N iterations for the direct and SAE model estimates were evaluated for bias, relative bias, mean squared error (MSE), root mean squared error (RMSE), percent coverage for a 95% nominal rate, and 95% confidence interval width for direct estimates and 95% credible interval width for the SAE models. Each of these metrics were calculated in two ways: i) for an individual county (i.e., Eqs. 4 and 6), and ii) for an individual state (i.e., Eqs. 5 and 7).

Using the following two equations, bias was calculated as the average difference between the posterior mean of the population total, $\hat{\theta}_{ij}$ of county i and iteration j, and the truth Y_{Ti}:

$$Bias_i = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_{ij} - Y_{Ti}; \quad (4)$$

$$Bias = \frac{1}{m}\sum_{i=1}^{m} Bias_i. \quad (5)$$

Additionally relative bias, RB_i, is defined as the bias relative to the truth for each county ($RB_i = Bias_i \,/\, Y_{Ti}$) and summed across all counties for a state ($RB = \frac{1}{m}\sum_{i=1}^{m} RB_i$).

A trade-off between bias and precision is present when applying a SAE model. MSE was calculated as the average squared deviations of the posterior mean of the population total from the true population, and RMSE was the square root of these deviations:

$$MSE_i = \frac{1}{N}\sum_{j=1}^{N}(\hat{\theta}_{ij} - Y_{Ti})^2; \tag{6}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m} MSE_i. \tag{7}$$

Percent coverage was defined as the average number of times the 95% SAE credible interval for θ_{ij}, or the direct estimate 95% confidence interval, included truth for each county. The average 95% confidence interval width for the direct estimate was computed by a t-distribution with n_i minus one degrees of freedom; whereas, the average 95% credible interval width was determined from the .025 and .975 quantiles of the posterior distributions of θ_{ij}. Both percent coverage and average confidence interval width were also calculated at the state-level.

A final relative MSE comparison among the direct and SAE model estimates was made at the county-level based on MSE (Eq. 6) by

$$\frac{MSE_1 - MSE_2}{\frac{1}{2}MSE_1 + \frac{1}{2}MSE_2} \tag{8}$$

where MSE_1 indicates the direct or the first SAE model estimate to be compared to MSE_2 indicating the second SAE model estimate (Porter et al., 2015).

3 Results

Summing the sample sizes across all counties for a single iteration, Montana had a statewide sample size of 751. Individual counties ranged from a minimum of two samples to a maximum of 44 samples in Fergus county. The average sample size per county in Montana was 13. Figure 1a maps the sample sizes across all of the counties for Montana. The true population of Montana is 42 625 ownerships. Liberty county had the minimum population with 58 ownerships, and Flathead county had the maximum with 5209 ownerships (Fig. 1b). Montana is one of the least densely populated states in the US ranging from 0.097 to 21.6 people km^{-2} for an individual county with an average of 2.7 people km^{-2} (Fig. 1c). Montana's forest area is primarily concentrated in the western portion of the state with a range of 1077 ha (Sheridan county in the east) to 919 800 ha (Flathead county in the west) and a mean 151 200 ha per county (Fig. 1d).

New Jersey had a statewide sample size of 191 for each iteration. Individual counties ranged from a minimum of two samples to a maximum of 20 samples in Burlington county. The average sample size per county in New Jersey was nine. Figure 2a maps the sample sizes across all of the counties for New Jersey. The true population of New Jersey is 35 462 ownerships. Hudson county had the minimum population with 33 ownerships, and Atlantic county had the maximum with 3842 ownerships (Fig. 2b). New Jersey is one of the most densely populated states in the US ranging from 73.4 to 4753 people km^{-2} with an average of 806.6 people km^{-2} in a single county (Fig. 2c). New Jersey's forest area is primarily concentrated in the northwestern and southeastern portions of the state with a range of 562 ha (Hudson county) to 121 500 ha (Burlington county) and a mean of 41 210 ha per county (Fig. 2d).

approximate location for Figure 1

approximate location for Figure 2

In logarithmic form, the regression coefficients were significant for the intercept, PD and TF for the majority of the iterations. Averaged across all iterations for the Montana FH model, the mean point estimate for the intercept was 5.01, the PD regression coefficient was 0.0733, and the TF regression coefficient was 4.10x10^{-6}. For the FHCAR model, the point estimates were 5.04, 0.0676, and 4.01x10^{-6} for the intercept, PD and TF, respectively. None of the 95% credible intervals for the intercept or the TF regression coefficient included zero for the N iterations of either SAE model. For the PD regression coefficient, 3032 and 2914 iterations did not include zero in the credible interval of the FH and FHCAR models, respectively.

In the case of the New Jersey FH model, the point estimates were 7.26, -9.35x10^{-4}, and 1.04x10^{-5} for the intercept, PD and TF regression coefficients, respectively. For the FHCAR model, the point estimates were 7.25, -9.22x10^{-4}, and 1.08x10^{-5} for the intercept, PD and TF regression coefficients, respectively. None of the 95% credible intervals for the intercept included zero for any of the iterations of either model. None of the credible intervals for the PD regression coefficient included zero for any iterations of the FH model and 3996 iterations of the FHCAR model did not include zero. For the TF regression coefficient, 1735 and 2321 iterations did not include zero in the credible interval of the FH and FHCAR models, respectively.

SAE model parameter estimates for a randomly selected iteration are given in Table 1. All of the regression coefficients were significant with the exception of TF in New Jersey. For Montana, the mean of the sampling variances was much smaller than the random effect variance for both the FH and FHCAR models. Whereas, New Jersey had roughly equal mean sampling and random effect variances. The autocorrelation parameter in both states was fairly low with a much wider credible interval in New Jersey than Montana. Figure 3a confirms the posterior means of the sampling variances of all iterations in Montana were lower than the random effect variances of the FH and FHCAR models. However, for New Jersey, the FH and FHCAR random effect variance was smaller than the mean sampling variance (Fig. 3c). The posterior mean of all iterations for the FHCAR autocorrelation parameter was also fairly low in Montana and slightly greater in New Jersey (Fig. 3b and d).

approximate location for Table 1

approximate location for Figure 3

Table 2 shows the simulation summaries for the individual states. As an aside, and not surprisingly, this simulation study can empirically show the direct estimator is nearly unbiased (see Appendix, Tables A1 – A3). For Montana, the bias of the direct, FH and FHCAR model estimates was -3, -25, and -28 ownerships, respectively. A negative value indicates an underestimate of the true population and a positive value indicates an overestimate of the true population. In terms of relative bias, these statewide estimates represent less than negative one-tenth of a percent. The RMSE was largest for the direct estimates with a value of 467. Both the FH and FHCAR models had similar RMSE values of 407 and 405. The empirical coverage rates were low for the direct and both model estimates with values of 75.0%, 75.3%, and 75.3%. The 95% confidence or credible interval widths, ordered widest to narrowest, were 1663, 1109 and 1097 for the direct, FH and FHCAR model estimates. Overall the MSE is reduced by 24.0% from the direct to the FH model, 24.8% from the direct to the FHCAR model, and 1.0% from the FH to the FHCAR model.

For New Jersey, the bias of the direct estimates was one ownership, the FH model was 49 ownerships, and the FHCAR model was 40 ownerships. In terms of relative bias, these statewide estimates represent less than two-tenths of a percent. The RMSE was largest for the direct estimates with a value of 633. The FH model had the lowest RMSE of 545; whereas, the FHCAR model was between the two with a value of 564. The empirical coverage rate was highest for the direct estimates followed by the FH and FHCAR models with values of 93.7%, 90.7%, and 89.2%, respectively. The direct estimates had the widest 95% confidence interval width of 3979 followed by the FH model 95% credible interval width of 2207 and the FHCAR model of 2050. Overall the MSE is reduced by 25.9% from the direct to the FH model, 20.6% from the direct to the FHCAR model, and 6.6% from the FHCAR to the FH model.

approximate location for Table 2

For individual counties of Montana, the bias (Eq. 4) of the direct estimates ranged from -29 (for Lincoln county) to 12 (for Lewis and Clark county) ownerships (Appendix, Tables A1 and A2). Eight of the counties were unbiased for the direct estimates. In terms of relative bias, the direct estimates ranged from negative three to five percent. The FH model had a bias ranging from -289 (for Stillwater county) to 444 ownerships (for Flathead county). Treasure county was the lone unbiased county for the FH model. The relative bias ranged from -26% to 41% with a mean of -3%. The FHCAR model had a bias ranging from -288 (for Musselshell county) to 498 ownerships (for Flathead county). Daniels county was the only unbiased county for the FHCAR model. The relative bias ranged from -25% to 30% with a mean of -4%. These biases can be explained by the trade-off of increasing precision with an associated increasing bias. Based upon the relative MSE comparison of Eq. 8, 49 of 56 counties showed improvement from the direct to the FH model estimates, and 52 counties showed improvement from the direct to the FHCAR model estimates. Thirty counties showed improvement from the FH to the FHCAR model (Fig. 4). Both the FH and FHCAR models had 38 counties with greater percent coverage than the direct. The credible interval widths were narrower in all but Flathead county for both the SAE models compared to the direct estimate confidence interval widths.

For individual counties of New Jersey, the bias of the direct estimates ranged from -27 (for Cumberland county) to 18 (for Burlington county) ownerships (Appendix, Table A3). Four of the counties were unbiased for the direct estimates. In terms of relative bias, the direct estimates ranged from negative one to one percent. The FH model had a bias ranging from -477 (for Hunterdon county) to 673 ownerships (for Ocean county). No county was unbiased for either the FH or FHCAR model. The relative bias ranged from -15% to 372% (for Union county). Removing Union county, which has a small true ownership size of 46, the maximum becomes 55% (for Cape May county) and the mean is nine percent. The FHCAR model had a bias ranging from -572 (for Hunterdon county) to 764 ownerships (for Ocean county). The relative bias ranged from -17% to 420% (for Union county). Removing Union county again, the maximum becomes 60% (for Cape May county) and the mean nine percent. Sixteen of 21 counties showed improvement based upon relative MSE (Eq. 8) from the direct to the FH model and to the FHCAR model. Ten counties showed improvement from the FH to the FHCAR model (Fig. 5). The FH and FHCAR models had 16 and 14 counties, respectively, with greater percent coverage than the direct estimates. The credible interval widths were narrower in all counties for both the SAE models compared to the direct estimate confidence interval widths.

approximate location for Figure 4

approximate location for Figure 5

4 Discussion

Recent publications for the NWOS have focused on private ownerships 4.05 ha or greater in size (e.g., Butler et al., 2016c). Our original objective in this study was to include private forest ownerships of at least 0.41 ha, the minimum forest area considered by the FIA. These smallest ownerships were to be included as these ownerships comprise a large proportion of ownerships in states like New Jersey relative to Montana. However, preliminary analyses showed large positive biases in the direct estimator so we adjusted the minimum area threshold to two ha to attain empirically unbiased direct estimates. A two ha threshold still accounted for a large proportion of these smallest ownerships rather than increasing the threshold to 4.05 ha. The unbiased direct estimates were desirable for benchmarking against the SAE models, and the potential bias induced by these smallest ownerships may require further investigation.

We found population density and total forest area to be significant SAE model covariates in the majority of the iterations for both Montana and New Jersey. However, population density had a positive relationship with number of private ownerships for Montana and a negative association for New Jersey counties. A possible explanation for the differing relationships could be the very low population densities of Montana counties having not reached a critical threshold in the forested landscape; whereas, New Jersey is the most densely populated state in the US causing a corresponding loss of forested land. Kittredge et al. (2008) found population densities greater than 96.5–193 people km^{-2} (250–500 people mile^{-2}) for towns in Massachusetts increased the number of forested parcels less than eight ha and secondarily reduced the proportion of the landbase in forest due to land use change. Poudyal and Hodges (2009) also found a negative relationship between population density and woodland ownership population size for Texas. In studying all private forest landowners, Pan et al. (2009) found population density to decrease the mean forest landholding size in Alabama, which in turn could relate to a larger number of ownerships due to parcellation. The average county population density of Alabama was reported as 29 people km^{-2} (76 people mile^{-2}), slightly higher than the highest density found in Montana, so at these densities a positive relationship with number of ownerships is plausible. The positive relationship between number of ownerships and total forest area in both states is not surprising. As a county must have a forested landbase for private forest ownerships to be present. Poudyal and Hodges (2009) have also demonstrated the need for total private forest area at the county-level in modeling private forest owner population size.

Development of a true ownership population at the county-level is not novel; however, the methods here are a substantial contribution with complete coverage of two states. Other studies have used similar approaches to develop forestland ownerships for a single county with interest at the individual parcel-level (e.g., Cho and Newman, 2005; King and Butler, 2005). Cho and Newman (2005) were interested in the probability of land use conversion from forest to developed; King and Butler (2005) were interested in modeling landholding size for individual parcels. Each of these response variables are important to county-level forest management options available to an individual landowner. In each of these studies, the distance to the nearest road from the parcel was the most important independent variable. At the county level, road density may be another important covariate for future exploration; however, the inclusion of road density may be collinear with population density currently accounted for in the SAE models. Mehmood and Zhang (2001) found a decline in state-level average forest landholding size, which relates to an increase in the number of private forest ownerships for the period of 1978 to 1994. The five factors associated with this decrease in landholding size were death, urbanization, income, regulatory uncertainty, and financial assistance. A key objective of the current study was assessing the feasibility of applying SAE models to county-level ownerships and hence we did not exhaustively explore the full suite of potentially useful covariates, such as those identified in Mehmood and Zhang (2001). Next steps in developing this line of work will identify potential transformations to existing covariates, e.g., percentage of county forested, and an expanded set of useful SAE model covariates that explain variability in an expanded set of ownership characteristics and increase county-level estimate precision.

One interpretation of the conditional autoregressive (CAR) random effect in FHCAR model is to capture unobserved covariates that are spatially structured. For example, neighboring counties may have similar regulations or demographics that impact the number of ownerships. However, including a CAR random effect did not yield substantial gains in RMSE in Montana and actually increased RMSE for New Jersey. This, combined with the relatively small CAR correlation parameter estimates, suggests there is little local residual spatial structure and the FH model with unstructured random effects is adequate. The FH model was overall more successful than the direct estimates in terms of MSE and RMSE for each state and the majority of the individual counties of the two states in this study. Alternate covariates may be needed for applying SAE models to other responses of the NWOS, such as socio-economic variables of Poudyal and Hodges (2009).

For simplicity, our simulation study considered only iterations with at least two samples per county. This was done to ensure the direct estimator yielded estimates of θ and σ^2 for each county. In practice, however, when the state-wide sample size is small we would expect zero samples in many counties. In such settings, a SAE model can still be applied by imputing the "missing" θ and σ^2 estimates (ideally with uncertainty quantification). This can be viewed as a missing data problem that is easily handled within a Bayesian paradigm (Banerjee et al., 2014; Gelman et al., 2014). Here, given county-level observed covariates, prediction of missing θ's follow directly from the posited model, i.e., second line in model (3). The missing σ^2's can be predicted using a log-Normal regression comprising available covariates and spatially structured random effects, or other generalized variance functions (GVF) approaches common in SAE literature, see, e.g., Dick (1995). Our future work will explore the efficacy of such SAE missing data problems for county-level ownership characteristics.

With any survey implementation and sampling strategy, there is a trade-off between sample size (e.g., number of contacts, determined by expected response rate) and resources (e.g., time and money). Given limited resources, the number of contacts is typically constrained, and sample size is therefore heavily dependent on survey response rates. Although ownerships contacted in the NWOS respond at higher rates than many other mail-based surveys, the potential lack of sufficient responses at the county-level limits the usefulness of the data at scales where such data can enhance landowner outreach, incentive programs, and information campaigns. For statewide surveys with county-level coverage and possibly small county-level sample sizes, the FH model presented in this paper improves the ability of researchers and forest management practitioners to use the NWOS data, by adding simple and widely available covariates for population density and forest area. These covariates have been shown to relate to forest management (Zhang et al., 2005), as well, making their utility in SAE models applicable to more variables other than number of ownerships. Further research will assess SAE models use for county-level inference about other important response variables from the NWOS, e.g., ownerships with management plans. Such models will consider jointly modeling multiple response variables and those variables with non-Gaussian distributions, e.g., binary, count, and multinomial survey items.

Acknowledgements

We would like to acknowledge the valuable comments of the Associate Editor and two anonymous reviewers that improved the quality of this manuscript.

References

Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2014. *Hierarchical modeling and analysis for spatial data, second edition*. CRC Press, Boca Raton, Florida. 558 p.

Base Map Service Center Montana State Library. 2015. *Montana Cadastral Mapping Project*. Available online at https://svc.mt.gov/msl/mtcadastral; (accessed 12.05.15).

Bechtold, W.A., and P.L. Patterson. 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.

Butler, B.J., B.J. Dickinson, J.H. Hewes, S.M. Butler, K. Andrejczyk, and M. Markowski-Lindsay. 2016a. *USDA Forest Service National Woodland Owner Survey, 2011–2013: Design, implementation, and estimation methods*. Gen. Tech. Rep. NRS-157. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 43 p.

Butler, B.J., J.H. Hewes, B.J. Dickinson, K. Andrejczyk, M. Markowski-Lindsay, and S.M. Butler. 2016b. *USDA Forest Service National Woodland Owner Survey: national, regional, and state statistics for family forest and woodland ownerships with 10+ acres, 2011–2013*. Res. Bull. NRS-99. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 39 p.

Butler, B.J., J.H. Hewes, B.J. Dickinson, K. Andrejczyk, S.M. Butler, and M. Markowski-Lindsay. 2016c. Family Forest Ownerships of the United States, 2013: Findings from the USDA Forest Service's National Woodland Owner Survey. *J. Forest.* 114(6):638–647. doi:10.5849/jof.15-099.

Casella, G., and R.L. Berger. 2002. *Statistical Inference*. second edition. Duxbury, Pacific Grove, Califonia. 660 p.

Cho, S., and D.H. Newman. 2005. Spatial Analysis of Rural Land Development. *Forest Policy Econ.* 7(5): 732–744. doi:10.1016/j.forpol.2005.03.008.

Dick, P. 1995. Modelling net undercoverage in the 1991 Canadian census. *Survey Methodology* 21:45–54.

Dickinson, B.J., and B.J. Butler. 2013. Methods for Estimating Private Forest Ownership Statistics: Revised Methods for the USDA Forest Service's National Woodland Owner Survey. *J. Forest.* 111(5):319–325. doi:10.5849/jof.12-088.

Fay, R.E., and R.A. Herriot. 1979. Estimation of income for small places: An application of James-Stein procedures to census data. *J. Am. Stat. Assoc.* 74:268–277.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2014. *Bayesian data analysis*, third edition. Chapman & Hall/CRC Boca Raton, Florida. 661 p.

Goerndt, M.E., V.J. Monleon, and H. Temesgen. 2013. Small-area estimation of county-level forest attributes using ground data and remote sensed auxiliary information. *For. Sci.* 59(5):536–548. doi:10.5849/forsci.12-073.

Hansen, M.H., and W.N. Hurwitz. 1943. On the theory of sampling from finite populations. *Ann. Math. Stat.* 14(4):333–362.

Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. *Remote Sens. Environ.* 132:159–175. doi:10.1016/j.rse.2013.01.012.

King, S.L., and B.J. Butler. 2005. Generating a Forest Parcel Map for Madison County, NY. In, edited by Michael Bevers and Tara M. Barrett, Gen. Tech. Rep. PNW-GTR-656. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station 147–155.

Kittredge, D.B., A.W. D'Amato, P. Catanzaro, J. Fish, and B.J. Butler. 2008. Estimating ownerships and parcels of nonindustrial private forestland in Massachusetts. *North. J. Appl. For.* 25(2):93–98.

Magnussen, S., D. Mandallaz, J. Breidenbach, A. Lanz, and C. Ginzler. 2014. National forest inventories in the service of small area estimation of stem volume. *Can. J. For. Res.* 44(9):1079–1090. doi:10.1139/cjfr-2013-0448.

Mehmood, S.R., and D. Zhang. 2001. Forest Parcelization in the United States: A Study of Contributing Factors. *J. Forest.* 99(4): 30–34.

New Jersey Office of Information Technology, Office of GIS. 2015. *New Jersey Geographic Information Network*. Available online at https://njgin.state.nj.us/; (accessed 21.12.15).

Pan, Y., Y. Zhang, and I. Majumdar. 2009. Population, economic welfare and holding size distribution of private forestland in Alabama, USA. *Silva Fenn.* 43(1): 161–171.

Porter, A.T., C.K. Wikle, and S.H. Holan. 2015. Small area estimation via multivariate Fay–Herriot models with latent spatial dependence. *Aust. NZ J. Stat.* 57(1): 15–29. doi:10.1111/anzs.12101.

Poudyal, N.C., and D.G. Hodges. 2009. Property taxation and rural land values: Their effect on private forestland ownership structure in Texas. *Land Use Policy* 26(4): 1100–1105. doi:10.1016/j.landusepol.2009.01.005.

QGIS Development Team. 2014. *QGIS Geographic Information System*. Open Source Geospatial Foundation. URL http:/qgis.osgeo.org.

R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, J.N.K, and I. Molina. 2015. *Small area estimation*, second edition. John Wiley & Sons, Inc., Hoboken, New Jersey. 441 p.

Sahr, K. 2011. DGGRID version 6.1.

US Census Bureau. 2016. *2010 Census Data*. Available online at https://www.census.gov/2010census/data; (accessed 12.08.16).

US Geological Survey, Gap Analysis Program. 2016. *Protected Areas Database of the United States (PAD-US), version 1.4 Combined Feature Class*. Available online at https://gapanalysis.usgs.gov/padus; (accessed 29.12.15).

Ver Planck, N.R., A.O. Finley, E.S. Huff. 2017. Replication Data for: Hierarchical Bayesian models for small area estimation of county-level private forest landowner population. Harvard Dataverse, doi:10.7910/DVN/A3ROXD.

Ver Planck, N.R., A.L. Metcalf, A.O. Finley, and J.C. Finley. 2016. Evaluation of the USDA Forest Service National Woodland Owner Survey estimators for private forest area and landowners: a case study of Montana. *For. Sci.* 62(5):525–534. doi:10.5849/forsci.15-196.

You, Y., and Q.M. Zhou. 2011. Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Surv. Methodol.* 37(1):25–37.

Zhang Y., D. Zhang, and J. Schelhas. 2005. Small-scale non-industrial private forest ownership in the United States: rationale and implications for forest management. *Silva Fenn.* 39(3):443–454.

Table 1: The median and 95% credible intervals for Fay-Herriot (FH) and FH with conditional autoregressive (FHCAR) models of a single iteration for Montana (iteration 750) and New Jersey (iteration 350).

	Montana		New Jersey	
Parameter	FH	FHCAR	FH	FHCAR
β_0	4.92 (4.50, 5.34)	4.91 (4.38, 5.46)	7.42 (6.65, 8.13)	7.41 (6.59, 8.14)
β_1, PD	0.078 (0.001, 0.146)	0.072 (0.007, 0.135)	-1.11x10^{-3} (-1.48x10^{-3}, -7.42x10^{-4})	-1.09x10^{-3} (-1.45x10^{-3}, -7.00x10^{-4})
β_2, TF	4.40x10^{-6} (2.70x10^{-6}, 6.08x10^{-6})	4.52x10^{-6} (2.72x10^{-6}, 6.35x10^{-6})	1.12x10^{-5} (-3.56x10^{-7}, 2.27x10^{-5})	1.16x10^{-5} (1.14x10^{-6}, 2.23x10^{-5})
$\mathbb{E}(\tilde{\sigma}_i^2)$	0.16 (0.02, 0.67)	0.16 (0.02, 0.67)	0.23 (0.09, 0.75)	0.23 (0.09, 0.75)
σ_v^2	1.14 (0.76, 1.77)	1.71 (0.95, 3.73)	0.21 (0.07, 0.68)	0.22 (0.07, 0.93)
λ	—	0.14 (0.007, 0.62)	—	0.29 (0.01,0.92)

Table 2: Summary of bias, root mean squared error (RMSE), empirical coverage for a 95% nominal coverage rate and average 95% confidence interval width of the direct estimate and 95% credible interval width of the small area estimation model estimates for Montana and New Jersey across all counties and iterations for the direct, Fay-Herriot (FH) and FH with conditional autoregressive random effects (FHCAR) model estimates.

	Montana			New Jersey		
Metric	Direct	FH	FHCAR	Direct	FH	FHCAR
Bias	-3	-25	-28	1	49	40
RMSE	467	407	405	633	545	564
Percent Coverage	75.0	75.3	75.3	93.7	90.7	89.2
Confidence / Credible Interval Width	1663	1109	1097	3979	2207	2050

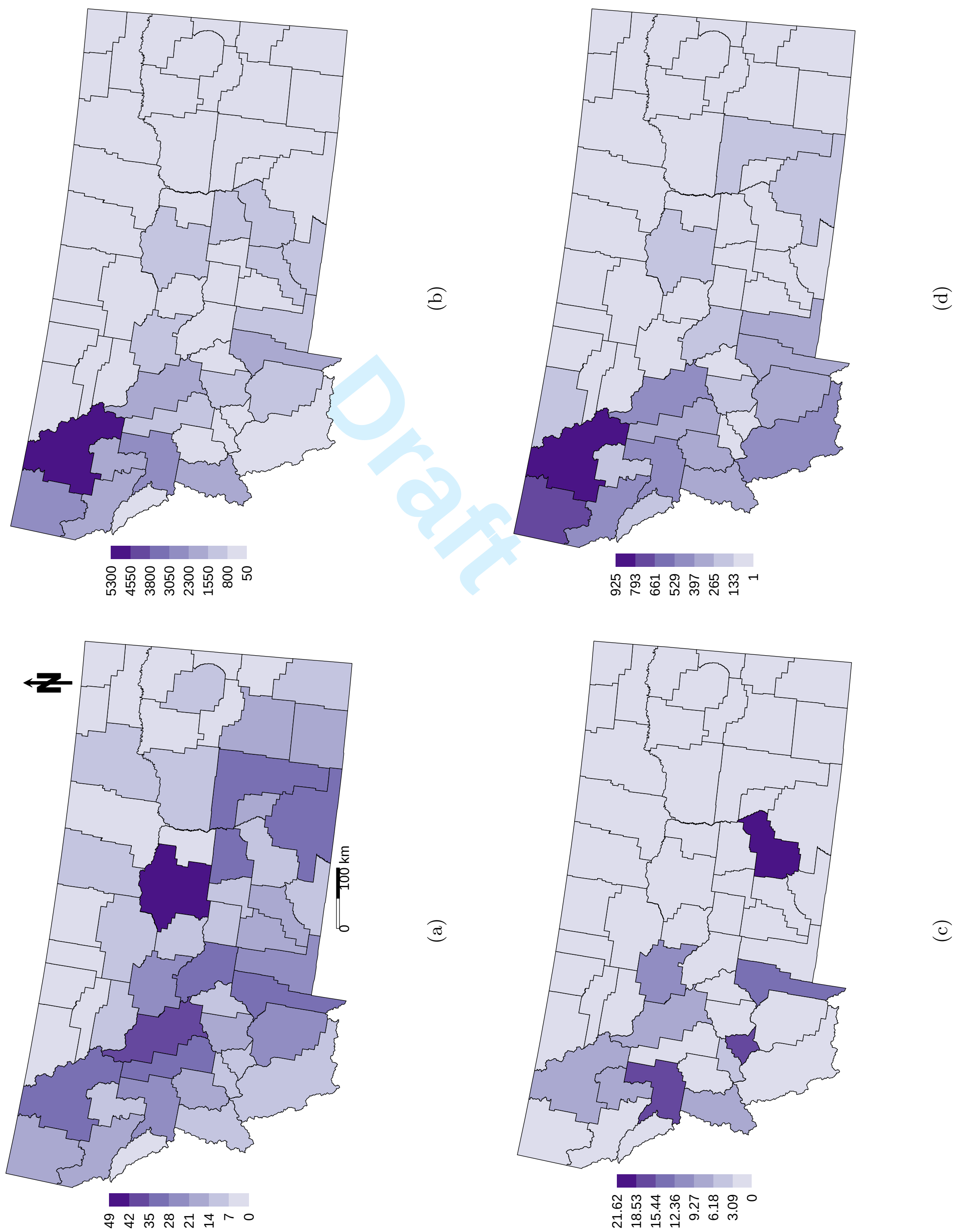
Figure 1: Montana counties by (a) sample size, (b) true population size with >2 ha forest, (c) 2010 Census population density (PD, people km^{-2}); (d) total forest area (TF, thousands ha).

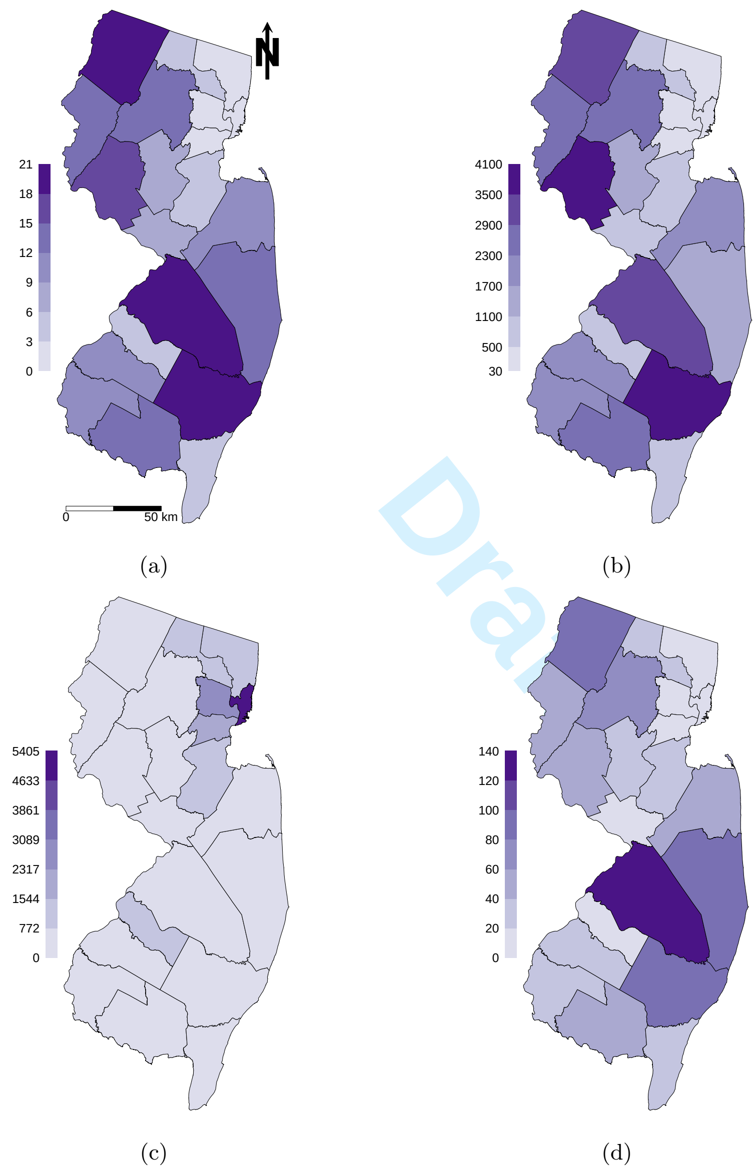
Figure 2: New Jersey counties by (a) sample size, (b) true population size with >2 ha forest, (c) 2010 Census population density (PD, people km^{-2}); (d) total forest area (TF, thousands ha).

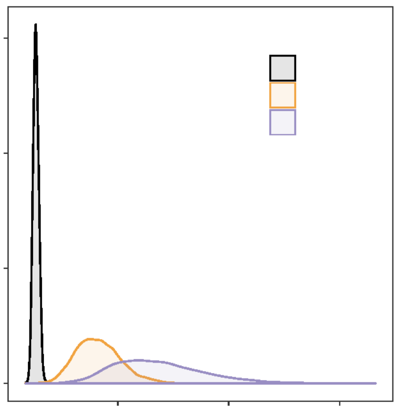
Figure 3: Distribution of posterior means from all iterations for select SAE model parameters: (a) variance parameters of Montana, (b) the FHCAR model autocorrelation parameter of Montana, (c) variance parameters of New Jersey; (d) the FHCAR model autocorrelation parameter of New Jersey.

Figure 4: Relative mean squared error comparisons for Montana (Eq. 8): (a) direct to FH model estimates, (b) direct to FHCAR model estimates; (c) FH to FHCAR model estimates.

Figure 5: Relative mean squared error comparisons for New Jersey (Eq. 8): (a) direct to FH model estimates, (b) direct to FHCAR model estimates; (c) FH to FHCAR model estimates.

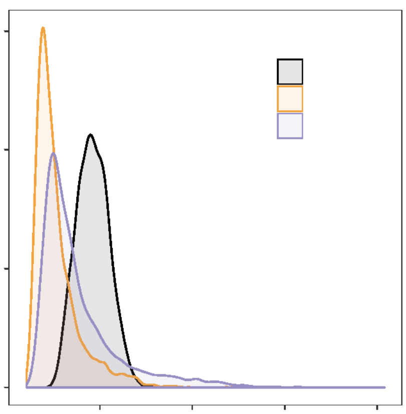

(b)
5300
4550
3800
3050
2300
1550
800
50
(d)
925
793
661
529
397
265
133
1
(a)
49
42
35
28
21
14
7
0
N
0
100 km
(c)
21.62
18.53
15.44
12.36
9.27
6.18
3.09
0


N
21
18
15
12
9
6
3
0
0
50 km
(a)
4100
3500
2900
2300
1700
1100
500
30
(b)
5405
4633
3861
3089
2317
1544
772
0
(c)
140
120
100
80
60
40
20
0
(d)

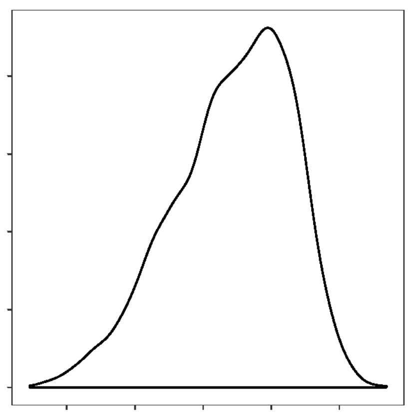


(a)

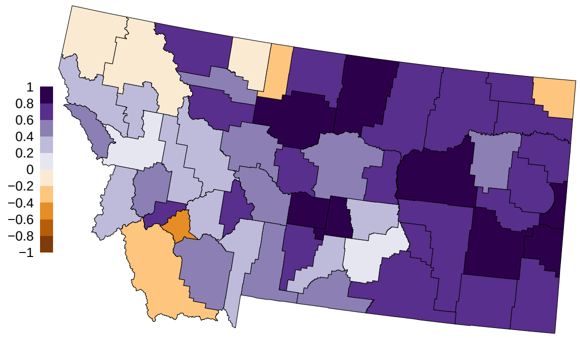
(b)



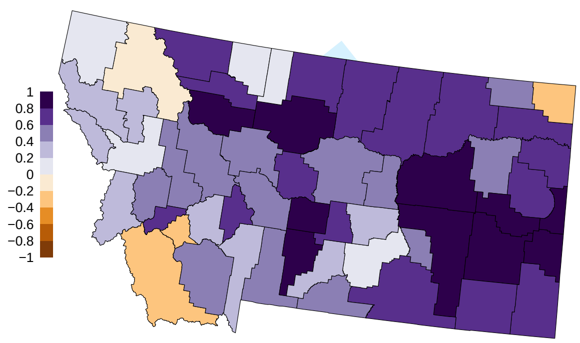
(c)



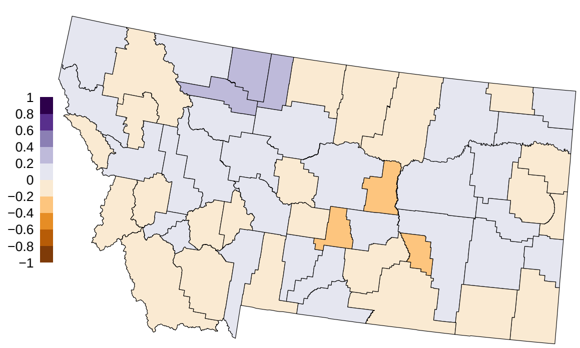
(d)


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

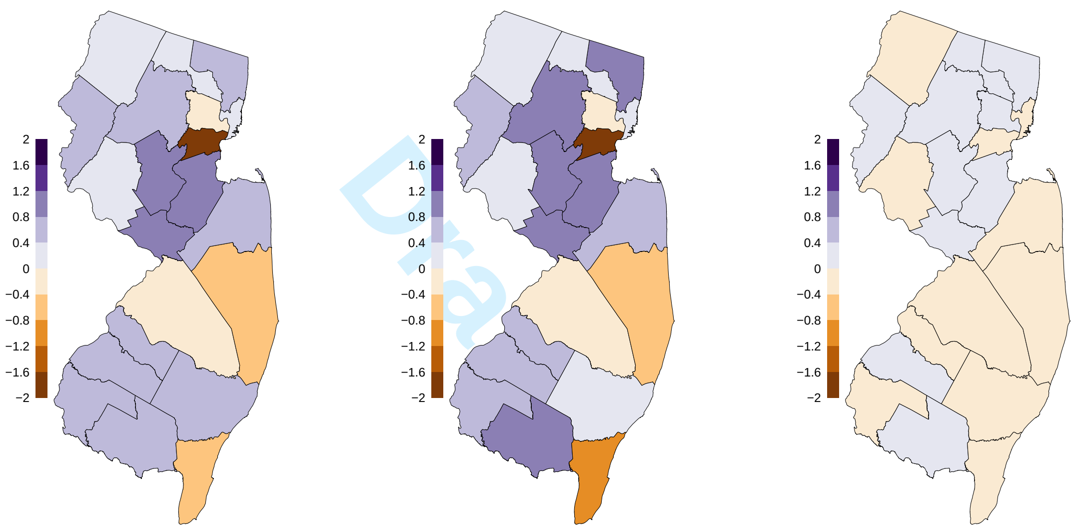
(a)


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

(b)


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

(c)



(a) (b) (c)

A Appendix

Table A1: Summary of the first 28 Montana counties by identification number and name for true population total, sample size, and bias across all repeated samples.

				Bias		
County ID	Name	Truth	n	Direct	FH	FHCAR
1	Silver Bow	439	7	-6	144	131
2	Cascade	1309	23	-12	-130	-160
3	Yellowstone	1062	12	4	54	97
4	Missoula	2782	27	-15	218	216
5	Lewis & Clark	1941	38	12	17	-35
6	Gallatin	1804	29	-9	17	17
7	Flathead	5209	30	-16	444	498
8	Fergus	1239	44	6	-173	-224
9	Powder River	344	18	-1	-22	-14
10	Carbon	868	9	3	-185	-142
11	Phillips	304	5	0	-36	-27
12	Hill	209	5	-5	-17	-22
13	Ravalli	1743	13	-8	-64	-103
14	Custer	461	18	-11	-101	-114
15	Lake	1589	10	9	-206	-190
16	Dawson	467	8	-10	-90	-97
17	Roosevelt	264	2	1	-27	-33
18	Beaverhead	411	9	-3	119	119
19	Chouteau	354	8	-4	-60	-85
20	Valley	505	7	-2	-90	-105
21	Toole	72	2	0	27	16
22	Big Horn	618	30	-9	-68	-56
23	Musselshell	1247	31	-13	-285	-288
24	Blaine	321	9	-6	-40	-32
25	Madison	838	21	0	-81	-87
26	Pondera	125	3	2	22	4
27	Richland	489	5	-2	-86	-91
28	Powell	825	30	1	1	-3

Table A2: Summary of the second 28 Montana counties by identification number and name for true population total, sample size, and bias across all repeated samples.

				Bias		
County ID	Name	Truth	n	Direct	FH	FHCAR
29	Rosebud	448	32	6	-31	-34
30	Deer Lodge	454	7	1	-17	-30
31	Teton	296	7	-9	-35	-42
32	Stillwater	1133	14	-11	-289	-278
33	Treasure	175	14	-1	0	7
34	Sheridan	79	2	0	26	23
35	Sanders	1816	16	-12	-37	-59
36	Judith Basin	412	8	2	-35	-28
37	Daniels	154	2	0	-1	0
38	Glacier	310	6	-7	-10	-30
39	Fallon	223	3	2	-18	-20
40	Sweet Grass	477	16	-1	-62	-58
41	McCone	156	2	1	-4	-1
42	Carter	296	9	-4	-34	-27
43	Broadwater	310	10	-8	-33	-21
44	Wheatland	180	8	0	-7	-2
45	Prairie	153	3	0	-2	-6
46	Granite	754	17	-6	-35	-27
47	Meagher	388	30	1	6	12
48	Liberty	58	2	0	24	17
49	Park	1032	26	-4	-92	-53
50	Garfield	256	8	-3	-31	-38
51	Jefferson	1385	16	2	-192	-254
52	Wibaux	148	4	3	-3	1
53	Golden Valley	153	7	8	6	27
54	Mineral	520	4	9	-2	18
55	Petroleum	163	6	-2	1	14
56	Lincoln	2857	19	-29	183	144

Table A3: Summary of New Jersey counties by identification number and name for true population total, sample size, and bias across all repeated samples.

				Bias		
County ID	Name	Truth	n	Direct	FH	FHCAR
1	Atlantic	3842	18	17	-214	-489
2	Bergen	350	2	5	113	102
3	Burlington	3210	20	18	491	649
4	Camden	627	3	0	157	178
5	Cape May	874	5	-9	483	524
6	Cumberland	2591	13	-27	-47	-93
7	Essex	132	2	0	49	49
8	Gloucester	2092	9	-2	-202	-266
9	Hudson	33	2	0	-5	-5
10	Hunterdon	3785	16	-2	-477	-572
11	Mercer	1001	6	2	63	103
12	Middlesex	917	5	8	-68	-75
13	Monmouth	2139	10	8	-249	-363
14	Morris	2762	14	9	-193	-313
15	Ocean	1507	12	1	673	764
16	Passaic	522	3	2	192	201
17	Salem	1831	9	4	54	90
18	Somerset	1629	8	-4	-60	-47
19	Sussex	3013	18	-4	224	284
20	Union	46	2	0	171	193
21	Warren	2739	14	-13	-123	-77

Table A4: Summary of the first 28 Montana counties by identification number for root mean squared error, empirical coverage for a 95% nominal coverage rate, and 95% confidence interval width for the direct estimator and 95% credible interval width for the two small area estimation models across all repeated samples.

	RMSE			Percent Coverage			Confidence / Credible interval width		
County ID	Direct	FH	FHCAR	Direct	FH	FHCAR	Direct	FH	FHCAR
1	341	418	405	70.2	78.9	79.0	1236	1195	1165
2	678	543	536	81.6	84.2	83.5	2513	1960	1894
3	571	536	556	80.3	84.0	84.5	2283	2001	2054
4	990	965	942	88.9	93.1	93.3	3964	3928	3862
5	867	742	703	85.6	90.3	90.4	3203	2881	2738
6	813	701	695	85.9	91.3	91.5	3117	2813	2791
7	1220	1304	1323	93.6	94.7	94.6	5252	5401	5426
8	531	414	410	79.7	79.2	77.2	1828	1399	1301
9	224	151	159	72.4	75.5	76.1	625	469	483
10	495	382	381	79.1	76.1	78.2	2018	1176	1260
11	264	174	175	70.0	72.5	73.5	1031	526	534
12	223	152	153	59.5	62.5	61.7	749	391	391
13	780	643	643	87.6	90.5	89.8	3327	2646	2585
14	384	245	235	56.5	59.0	58.2	1089	663	620
15	710	586	594	88.0	87.4	87.6	3269	2215	2272
16	319	213	214	72.7	72.4	71.6	1116	647	636
17	244	165	161	87.9	64.1	63.8	4181	441	421
18	314	365	367	66.7	76.1	75.6	1061	1007	1011
19	295	187	173	65.9	67.8	66.1	962	551	479
20	314	215	214	76.6	73.0	71.6	1159	662	631
21	73	78	70	84.8	68.9	69.6	1196	204	187
22	372	269	270	73.5	77.2	78.1	1223	918	929
23	570	488	486	82.2	77.2	76.8	2072	1427	1410
24	251	164	169	69.6	73.0	73.6	795	505	520
25	531	394	395	74.9	78.8	78.5	1823	1349	1343
26	163	126	107	68.2	70.7	69.5	818	330	282
27	308	217	219	80.8	77.0	76.3	1391	691	685
28	465	382	369	77.1	82.1	82.5	1564	1339	1301

Table A5: Summary of the second 28 Montana counties by identification number for root mean squared error, empirical coverage for a 95% nominal coverage rate, and 95% confidence interval width for the direct estimator and 95% credible interval width for the two small area estimation models across all repeated samples.

	RMSE			Percent Coverage			Confidence / Credible interval width		
County ID	Direct	FH	FHCAR	Direct	FH	FHCAR	Direct	FH	FHCAR
29	302	204	190	75.0	80.7	81.3	870	665	630
30	352	253	240	77.9	80.9	81.1	1377	889	837
31	292	193	187	61.7	67.8	67.4	952	552	530
32	599	499	491	81.8	76.6	77.6	2292	1401	1406
33	151	104	116	69.0	77.5	77.5	402	305	327
34	51	63	62	96.0	76.6	75.9	1139	199	202
35	671	586	572	87.8	89.8	89.8	2749	2343	2290
36	295	210	211	72.7	75.4	75.9	1057	679	683
37	166	118	122	84.8	64.0	63.7	2474	292	304
38	270	195	188	65.5	70.9	68.3	902	557	532
39	212	138	137	76.5	68.8	68.5	1147	384	381
40	352	232	230	71.0	74.8	75.4	1066	723	721
41	185	142	140	75.2	52.1	53.6	2854	315	307
42	212	140	149	68.3	69.9	70.2	650	417	439
43	301	198	209	59.6	65.7	66.8	864	551	581
44	194	122	125	62.2	67.9	68.7	509	323	327
45	172	117	109	70.2	66.5	67.0	842	297	279
46	462	356	362	75.5	81.2	81.3	1603	1252	1270
47	306	245	237	60.8	67.2	68.6	892	727	696
48	74	82	73	71.2	57.0	58.1	1111	184	167
49	570	441	449	77.8	81.0	82.0	2037	1574	1628
50	247	159	147	57.2	64.2	64.5	737	436	400
51	632	519	520	86.3	85.8	83.2	2509	1810	1678
52	196	122	127	59.4	60.7	61.0	674	306	316
53	191	121	137	63.4	73.3	76.5	532	332	363
54	382	300	328	72.5	70.3	69.8	1860	944	1026
55	153	105	117	71.9	78.5	79.5	487	307	329
56	893	896	886	88.9	91.8	91.6	3701	3589	3562

Table A6: Summary of New Jersey counties by identification number for root mean squared error, empirical coverage for a 95% nominal coverage rate, and confidence interval width for the direct estimator and credible interval width for the two small area estimation models across all repeated samples.

	RMSE			Percent Coverage			Confidence / Credible interval width		
County ID	Direct	FH	FHCAR	Direct	FH	FHCAR	Direct	FH	FHCAR
1	942	732	790	97.3	98.5	96.4	5254	3947	3359
2	247	162	157	95.7	98.4	98.6	12481	939	843
3	975	1041	1131	93.1	95.6	95.2	4499	4439	4471
4	354	276	288	91.4	93.8	93.8	3925	1302	1265
5	417	637	682	90.2	89.2	86.5	2668	2058	2095
6	842	578	528	93.1	96.7	97.0	4114	2978	2712
7	75	78	78	98.9	96.2	93.5	2933	347	293
8	644	473	472	94.9	96.0	95.2	3839	2362	2125
9	18	17	17	96.2	67.8	66.7	448	54	52
10	906	832	864	96.1	94.3	92.0	4898	3437	3190
11	457	292	284	91.4	96.0	96.7	2803	1587	1449
12	451	270	248	89.8	92.3	92.9	2777	1251	1125
13	621	454	499	97.2	97.9	96.2	4080	2425	2124
14	779	522	510	97.8	99.2	98.8	5137	3368	2761
15	701	874	937	88.4	93.9	90.0	3413	3215	3074
16	359	323	320	85.2	89.0	87.6	3480	1179	1061
17	617	484	494	90.8	93.3	93.7	3000	2177	2192
18	550	324	295	96.4	98.8	99.1	3972	2279	2015
19	912	767	776	92.0	96.2	96.2	4310	3660	3610
20	30	186	207	100.0	26.9	13.6	1521	448	401
21	847	616	588	91.6	93.8	94.6	4009	2898	2833