



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

Journal:	<i>Canadian Journal of Forest Research</i>
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

SCHOLARONE™
Manuscripts

¹ Hierarchical Bayesian models for small area estimation
² of county-level private forest landowner population

³ Neil R. Ver Planck^{*1}, Andrew O. Finley¹, and Emily S. Huff¹

⁴ ¹Department of Forestry, Michigan State University,

⁵ 480 Wilson Road Room 126, East Lansing, MI 48824-1222, USA

⁶ September 18, 2017

Draft

*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

30 1 Introduction

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

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

56 The NWOS's target sample size is 250 respondents per state (Butler et al., 2016a) and
57 a required 100 respondent minimum to generate state survey summaries (Butler et al.,
58 2016b). Again, such small state-level sample sizes effectively preclude robust inference

59 at county-levels, particularly when reporting results by ownership type. Small sample
60 sizes can result in undesirably low precision of parameter estimates within a design-based
61 framework (e.g., the NWOS design). Here, we refer to parameter estimates obtained
62 using design-based estimators as *direct estimates*. These direct estimates are the current
63 standard used in operational settings that employ design-based estimators; therefore, the
64 direct estimates will be used for benchmarking against model-based approaches. Small
65 area estimation (SAE) is a model-based approach that couples a direct estimate and pos-
66 sible covariates to improve the estimate precision and, in some cases, accuracy. Rao and
67 Molina (2015) provide an excellent review of available SAE methods. Unlike a standard
68 linear regression, the SAE framework is comprised of two component models: a sampling
69 and a linking model (You and Zhou, 2011). The estimation of the SAE parameter of in-
70 terest accounts for and balances between the sampling (i.e., direct estimator) and linking
71 model errors. The linking model is a linear model with random effects that relate the
72 small areas of interest with some error. Additional spatial structure may still remain in
73 the linking model after accounting for possible covariates. Such residual structure can be
74 further modeled using spatial random effects.

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

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

86 The remainder of the manuscript is organized as follows. In Section 2 we detail the
87 steps followed to create a spatially explicit private forest ownership dataset for Montana

88 and New Jersey. Then we define the direct estimator for the number of county-level
89 private forest ownerships along with two SAE models. We then describe our approach
90 for comparing the proposed SAE models. Results are given in Section 3 that is followed
91 by discussion and future directions in Section 4.

92 **2 Materials and methods**

93 **2.1 True number of private forest ownerships**

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

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

112 To begin processing county-level properties, we identified ownerships in the dataset
113 with unknown names. Parcels for which ownership was unknown that fell within pub-
114 lic 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):

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

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

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 j th sample fell in private forest and zero otherwise, a_{ij} is the forest area (ha) of the sampled ownership, and

¹³⁶ $p_{ij} = a_{ij}/A_i$ is the ownership selection probability. The steps for drawing the ownership
¹³⁷ samples are described in Section 2.3. The total county-level private forest area, A_i , was
¹³⁸ fixed to equal the true private forest area derived from the GIS layer in Section 2.1.

¹³⁹ **2.2.2 Small area estimation models**

¹⁴⁰ The direct estimator, Section 2.2.1, was log transformed to meet the SAE model normal-
¹⁴¹ ity assumption. This transformation was also desirable because it ensures SAE model
¹⁴² population estimates have the correct support, i.e., positive, following back transforma-
¹⁴³ tion. Taking the log of the direct estimator necessitates transformation of the associated
¹⁴⁴ variance, accomplished here via the delta method (Casella and Berger, 2002). The Fay-
¹⁴⁵ Herriot (FH) SAE model (Fay and Herriot, 1979) for county i in $1, 2, \dots, m$ counties is
¹⁴⁶ defined:

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

¹⁴⁷ where \tilde{Y}_i is the log-transformed direct estimator, $\tilde{\theta}_i$ is the log population total, and ϵ_i is a
¹⁴⁸ normally distributed error with mean zero and variance $\tilde{\sigma}_i^2 = \sigma_i^2/Y_i^2$. The additive mean
¹⁴⁹ of $\tilde{\theta}_i$ comprises an intercept and county-level covariates held in the $p \times 1$ column vector \mathbf{x}
¹⁵⁰ and associated $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}$. The county-level random effects
¹⁵¹ term v_i is normally distributed with mean zero and variance σ_v^2 .

¹⁵² It is reasonable to think that forest and ownership patterns, e.g., property size and
¹⁵³ owner's socio-economic or demographic characteristics, could exhibit spatial autocorre-
¹⁵⁴ lation. In our setting, if direct estimate values are spatially correlated, i.e., adjacent
¹⁵⁵ counties have similar population values, then we should exploit this relationship to fur-
¹⁵⁶ ther improve inference by pooling information across proximate counties. Counties could
¹⁵⁷ be represented as either point locations (i.e., their centroid) or as areal units for defin-
¹⁵⁸ ing this spatial structure. Representing counties by their centroid may misrepresent

159 distances among neighboring counties due to irregularly shaped counties; therefore, the
160 areal approach is implemented as it maintains the desired neighborhood structure of
161 the county lattice. Hence, we augment model (3) by adding a spatially structured ran-
162 dom effect that follows a conditional autoregressive (CAR) prior distribution, see, e.g.,
163 Banerjee et al. (2014) and You and Zhou (2011). This extended model called FHCAR
164 is defined analogous to FH, with the exception that the unstructured random effects
165 $\mathbf{v} = (v_1, v_2, \dots, v_m)' \sim N(\mathbf{0}, \sigma_v^2 \mathbf{I})$ in model (3) are replaced with $\mathbf{v} \sim N(\mathbf{0}, \Sigma(\sigma_v^2, \lambda))$.
166 Here, the $m \times m$ covariance matrix $\Sigma(\sigma_v^2, \lambda) = \sigma_v^2 [\lambda \mathbf{R} + (1 - \lambda) \mathbf{I}]^{-1}$, where σ_v^2 is the
167 spatial variance parameter, λ is the autocorrelation parameter, \mathbf{R} is the neighborhood
168 matrix with diagonal elements equal to the number of neighbors and off diagonal elements
169 equal negative one or zero indicating if a neighbor is present or not, and \mathbf{I} is the $m \times m$
170 identity matrix. Counties are only considered neighbors with adjoining borders.

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

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

¹⁸⁸ mean of the population total), the K posterior samples of $\hat{\theta}$ were exponentiated back to
¹⁸⁹ the original units.

¹⁹⁰ **2.3 Simulation study**

¹⁹¹ For the NWOS and similar efforts, SAE models are viable from a statistical perspective
¹⁹² if they improve county-level estimate precision without inducing substantial bias. Given
¹⁹³ the actual private forest ownerships for Montana and New Jersey, Section 2.1, we examine
¹⁹⁴ SAE model inference against truth using a simulation study. One iteration in the simula-
¹⁹⁵ tion study produces a set of county-level direct and SAE model estimates by: *i*) drawing
¹⁹⁶ a random probability proportional to size sample from the private forest ownership list
¹⁹⁷ sample frame; *ii*) computing direct estimates (Section 2.2.1); *iii*) estimating FH and FH-
¹⁹⁸ CAR models (Section 2.2.2); *iv*) evaluating differences between SAE model population
¹⁹⁹ estimates and truth. Summarizing results from *iv* for a large number of iterations allows
²⁰⁰ us to assess precision and bias in SAE model population estimates.

²⁰¹ A county specific sample size, n_i , is needed to conduct step *i*. Here too, we want
²⁰² the sample size to approximate that achieved by the NWOS design. To determine the
²⁰³ sample size in each county, we randomly located 1000 points within each hexagon of a
²⁰⁴ tessellated hexagonal grid laid over the states of Montana and New Jersey using DGGRID
²⁰⁵ developed by Sahr (2011). Each of the 1000 points represented a unique sample iteration
²⁰⁶ across an individual state. The area of each hexagon was 2407 ha. This grid roughly
²⁰⁷ approximates the one used to spatially distribute FIA and NWOS samples (Bechtold
²⁰⁸ and Patterson, 2005; Dickinson and Butler, 2013). From each set of hexagons sampled
²⁰⁹ within a county, we calculated the mean number of points that fell within private forest
²¹⁰ ownerships across the 1000 sample iterations. This mean value was rounded up and used
²¹¹ as n_i . Given fixed n_i —to reduce variation among repeated samples—we repeated the
²¹² simulation study steps *i-iii* $N = 4000$ times for Montana and New Jersey. These large
²¹³ number of repeated simulations should empirically show unbiased estimates of truth for
²¹⁴ the direct estimator.

²¹⁵ **2.3.1 County-level covariates**

²¹⁶ 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
²¹⁷ forest area (TFA; ha) explain significant variability in log-transformed direct estimates.
²¹⁸ For Montana, the linear regressions of the simulation runs explained 45% of the variation
²¹⁹ on average with a range from 27% to 65%. For New Jersey, the explained variation was
²²⁰ higher with a mean of 76% and a range from 50% to 95%. Therefore, PD and TFA
²²¹ covariates were used for the simulation study.

²²³ **2.4 Simulation summary**

²²⁴ The N iterations for the direct and SAE model estimates were evaluated for bias, relative
²²⁵ bias, mean squared error (MSE), root mean squared error (RMSE), percent coverage for
²²⁶ a 95% nominal rate, and 95% confidence interval width for direct estimates and 95%
²²⁷ credible interval width for the SAE models. Each of these metrics were calculated in two
²²⁸ ways: *i*) for an individual county (i.e., Eqs. 4 and 6), and *ii*) for an individual state (i.e.,
²²⁹ Eqs. 5 and 7).

²³⁰ Using the following two equations, bias was calculated as the average difference be-
²³¹ tween the posterior mean of the population total, $\hat{\theta}_{ij}$ of county i and iteration j , and the
²³² truth Y_{Ti} :

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

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

²³³ Additionally relative bias, RB_i , is defined as the bias relative to the truth for each county
²³⁴ ($RB_i = Bias_i / Y_{Ti}$) and summed across all counties for a state ($RB = \frac{1}{m} \sum_{i=1}^m RB_i$).

²³⁵ A trade-off between bias and precision is present when applying a SAE model. MSE
²³⁶ was calculated as the average squared deviations of the posterior mean of the population

²³⁷ total from the true population, and RMSE was the square root of these deviations:

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

$$(7) \quad MSE = \frac{1}{m} \sum_{i=1}^m MSE_i.$$

²³⁸ Percent coverage was defined as the average number of times the 95% SAE credible
²³⁹ interval for θ_{ij} , or the direct estimate 95% confidence interval, included truth for each
²⁴⁰ county. The average 95% confidence interval width for the direct estimate was computed
²⁴¹ by a t-distribution with n_i minus one degrees of freedom; whereas, the average 95%
²⁴² credible interval width was determined from the .025 and .975 quantiles of the posterior
²⁴³ distributions of θ_{ij} . Both percent coverage and average confidence interval width were
²⁴⁴ also calculated at the state-level.

²⁴⁵ A final relative MSE comparison among the direct and SAE model estimates was
²⁴⁶ made at the county-level based on MSE (Eq. 6) by

$$(8) \quad \frac{MSE_1 - MSE_2}{\frac{1}{2}MSE_1 + \frac{1}{2}MSE_2}$$

²⁴⁷ where MSE_1 indicates the direct or the first SAE model estimate to be compared to
²⁴⁸ MSE_2 indicating the second SAE model estimate (Porter et al., 2015).

²⁴⁹ 3 Results

²⁵⁰ Summing the sample sizes across all counties for a single iteration, Montana had a
²⁵¹ statewide sample size of 751. Individual counties ranged from a minimum of two samples
²⁵² to a maximum of 44 samples in Fergus county. The average sample size per county in
²⁵³ Montana was 13. Figure 1a maps the sample sizes across all of the counties for Montana.
²⁵⁴ The true population of Montana is 42 625 ownerships. Liberty county had the minimum
²⁵⁵ population with 58 ownerships, and Flathead county had the maximum with 5209 own-

256 erships (Fig. 1b). Montana is one of the least densely populated states in the US ranging
257 from 0.097 to 21.6 people km^{-2} for an individual county with an average of 2.7 people
258 km^{-2} (Fig. 1c). Montana's forest area is primarily concentrated in the western portion of
259 the state with a range of 1077 ha (Sheridan county in the east) to 919 800 ha (Flathead
260 county in the west) and a mean 151 200 ha per county (Fig. 1d).

261 New Jersey had a statewide sample size of 191 for each iteration. Individual counties
262 ranged from a minimum of two samples to a maximum of 20 samples in Burlington county.
263 The average sample size per county in New Jersey was nine. Figure 2a maps the sample
264 sizes across all of the counties for New Jersey. The true population of New Jersey is 35
265 462 ownerships. Hudson county had the minimum population with 33 ownerships, and
266 Atlantic county had the maximum with 3842 ownerships (Fig. 2b). New Jersey is one of
267 the most densely populated states in the US ranging from 73.4 to 4753 people km^{-2} with
268 an average of 806.6 people km^{-2} in a single county (Fig. 2c). New Jersey's forest area is
269 primarily concentrated in the northwestern and southeastern portions of the state with
270 a range of 562 ha (Hudson county) to 121 500 ha (Burlington county) and a mean of 41
271 210 ha per county (Fig. 2d).

272 approximate location for Figure 1

273 approximate location for Figure 2

274 In logarithmic form, the regression coefficients were significant for the intercept, PD
275 and TF for the majority of the iterations. Averaged across all iterations for the Mon-
276 tana FH model, the mean point estimate for the intercept was 5.01, the PD regression
277 coefficient was 0.0733, and the TF regression coefficient was 4.10×10^{-6} . For the FHCAR
278 model, the point estimates were 5.04, 0.0676, and 4.01×10^{-6} for the intercept, PD and
279 TF, respectively. None of the 95% credible intervals for the intercept or the TF regression
280 coefficient included zero for the N iterations of either SAE model. For the PD regression
281 coefficient, 3032 and 2914 iterations did not include zero in the credible interval of the
282 FH and FHCAR models, respectively.

283 In the case of the New Jersey FH model, the point estimates were 7.26, -9.35×10^{-4} ,
284 and 1.04×10^{-5} for the intercept, PD and TF regression coefficients, respectively. For

285 the FHCAR model, the point estimates were 7.25, -9.22×10^{-4} , and 1.08×10^{-5} for the
286 intercept, PD and TF regression coefficients, respectively. None of the 95% credible
287 intervals for the intercept included zero for any of the iterations of either model. None
288 of the credible intervals for the PD regression coefficient included zero for any iterations
289 of the FH model and 3996 iterations of the FHCAR model did not include zero. For the
290 TF regression coefficient, 1735 and 2321 iterations did not include zero in the credible
291 interval of the FH and FHCAR models, respectively.

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

304 approximate location for Table 1

305 approximate location for Figure 3

306 Table 2 shows the simulation summaries for the individual states. As an aside, and
307 not surprisingly, this simulation study can empirically show the direct estimator is nearly
308 unbiased (see Appendix, Tables A1 – A3). For Montana, the bias of the direct, FH
309 and FHCAR model estimates was -3, -25, and -28 ownerships, respectively. A negative
310 value indicates an underestimate of the true population and a positive value indicates an
311 overestimate of the true population. In terms of relative bias, these statewide estimates
312 represent less than negative one-tenth of a percent. The RMSE was largest for the direct

313 estimates with a value of 467. Both the FH and FHCAR models had similar RMSE
314 values of 407 and 405. The empirical coverage rates were low for the direct and both
315 model estimates with values of 75.0%, 75.3%, and 75.3%. The 95% confidence or credible
316 interval widths, ordered widest to narrowest, were 1663, 1109 and 1097 for the direct, FH
317 and FHCAR model estimates. Overall the MSE is reduced by 24.0% from the direct to
318 the FH model, 24.8% from the direct to the FHCAR model, and 1.0% from the FH to
319 the FHCAR model.

320 For New Jersey, the bias of the direct estimates was one ownership, the FH model was
321 49 ownerships, and the FHCAR model was 40 ownerships. In terms of relative bias, these
322 statewide estimates represent less than two-tenths of a percent. The RMSE was largest
323 for the direct estimates with a value of 633. The FH model had the lowest RMSE of 545;
324 whereas, the FHCAR model was between the two with a value of 564. The empirical
325 coverage rate was highest for the direct estimates followed by the FH and FHCAR models
326 with values of 93.7%, 90.7%, and 89.2%, respectively. The direct estimates had the widest
327 95% confidence interval width of 3979 followed by the FH model 95% credible interval
328 width of 2207 and the FHCAR model of 2050. Overall the MSE is reduced by 25.9%
329 from the direct to the FH model, 20.6% from the direct to the FHCAR model, and 6.6%
330 from the FHCAR to the FH model.

331 approximate location for Table 2

332 For individual counties of Montana, the bias (Eq. 4) of the direct estimates ranged
333 from -29 (for Lincoln county) to 12 (for Lewis and Clark county) ownerships (Appendix,
334 Tables A1 and A2). Eight of the counties were unbiased for the direct estimates. In
335 terms of relative bias, the direct estimates ranged from negative three to five percent.
336 The FH model had a bias ranging from -289 (for Stillwater county) to 444 ownerships (for
337 Flathead county). Treasure county was the lone unbiased county for the FH model. The
338 relative bias ranged from -26% to 41% with a mean of -3%. The FHCAR model had a
339 bias ranging from -288 (for Musselshell county) to 498 ownerships (for Flathead county).
340 Daniels county was the only unbiased county for the FHCAR model. The relative bias
341 ranged from -25% to 30% with a mean of -4%. These biases can be explained by the

342 trade-off of increasing precision with an associated increasing bias. Based upon the
343 relative MSE comparison of Eq. 8, 49 of 56 counties showed improvement from the direct
344 to the FH model estimates, and 52 counties showed improvement from the direct to
345 the FHCAR model estimates. Thirty counties showed improvement from the FH to the
346 FHCAR model (Fig. 4). Both the FH and FHCAR models had 38 counties with greater
347 percent coverage than the direct. The credible interval widths were narrower in all but
348 Flathead county for both the SAE models compared to the direct estimate confidence
349 interval widths.

350 For individual counties of New Jersey, the bias of the direct estimates ranged from
351 -27 (for Cumberland county) to 18 (for Burlington county) ownerships (Appendix, Table
352 A3). Four of the counties were unbiased for the direct estimates. In terms of relative
353 bias, the direct estimates ranged from negative one to one percent. The FH model had
354 a bias ranging from -477 (for Hunterdon county) to 673 ownerships (for Ocean county).
355 No county was unbiased for either the FH or FHCAR model. The relative bias ranged
356 from -15% to 372% (for Union county). Removing Union county, which has a small true
357 ownership size of 46, the maximum becomes 55% (for Cape May county) and the mean is
358 nine percent. The FHCAR model had a bias ranging from -572 (for Hunterdon county)
359 to 764 ownerships (for Ocean county). The relative bias ranged from -17% to 420% (for
360 Union county). Removing Union county again, the maximum becomes 60% (for Cape
361 May county) and the mean nine percent. Sixteen of 21 counties showed improvement
362 based upon relative MSE (Eq. 8) from the direct to the FH model and to the FHCAR
363 model. Ten counties showed improvement from the FH to the FHCAR model (Fig. 5).
364 The FH and FHCAR models had 16 and 14 counties, respectively, with greater percent
365 coverage than the direct estimates. The credible interval widths were narrower in all
366 counties for both the SAE models compared to the direct estimate confidence interval
367 widths.

368 approximate location for Figure 4

369 approximate location for Figure 5

370 4 Discussion

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

382 We found population density and total forest area to be significant SAE model co-
383 variates in the majority of the iterations for both Montana and New Jersey. However,
384 population density had a positive relationship with number of private ownerships for
385 Montana and a negative association for New Jersey counties. A possible explanation for
386 the differing relationships could be the very low population densities of Montana counties
387 having not reached a critical threshold in the forested landscape; whereas, New Jersey is
388 the most densely populated state in the US causing a corresponding loss of forested land.
389 Kittredge et al. (2008) found population densities greater than $96.5\text{--}193 \text{ people km}^{-2}$
390 ($250\text{--}500 \text{ people mile}^{-2}$) for towns in Massachusetts increased the number of forested
391 parcels less than eight ha and secondarily reduced the proportion of the landbase in
392 forest due to land use change. Poudyal and Hodges (2009) also found a negative rela-
393 tionship between population density and woodland ownership population size for Texas.
394 In studying all private forest landowners, Pan et al. (2009) found population density to
395 decrease the mean forest landholding size in Alabama, which in turn could relate to a
396 larger number of ownerships due to parcellation. The average county population density
397 of Alabama was reported as $29 \text{ people km}^{-2}$ ($76 \text{ people mile}^{-2}$), slightly higher than the
398 highest density found in Montana, so at these densities a positive relationship with num-

ber 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

457 mail-based surveys, the potential lack of sufficient responses at the county-level limits
458 the usefulness of the data at scales where such data can enhance landowner outreach,
459 incentive programs, and information campaigns. For statewide surveys with county-level
460 coverage and possibly small county-level sample sizes, the FH model presented in this
461 paper improves the ability of researchers and forest management practitioners to use the
462 NWOS data, by adding simple and widely available covariates for population density and
463 forest area. These covariates have been shown to relate to forest management (Zhang
464 et al., 2005), as well, making their utility in SAE models applicable to more variables
465 other than number of ownerships. Further research will assess SAE models use for county-
466 level inference about other important response variables from the NWOS, e.g., ownerships
467 with management plans. Such models will consider jointly modeling multiple response
468 variables and those variables with non-Gaussian distributions, e.g., binary, count, and
469 multinomial survey items.

470 Acknowledgements

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

473 References

- 474 Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2014. *Hierarchical modeling and analysis for*
475 *spatial data, second edition.* CRC Press, Boca Raton, Florida. 558 p.
- 476 Base Map Service Center Montana State Library. 2015. *Montana Cadastral Mapping*
477 *Project.* Available online at <https://svc.mt.gov/msl/mtcadastral>; (accessed 12.05.15).
- 478 Bechtold, W.A., and P.L. Patterson. 2005. *The enhanced Forest Inventory and Analysis*
479 *program-national sampling design and estimation procedures.* Gen. Tech. Rep. SRS-
480 80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research
481 Station. 85 p.

- 482 Butler, B.J., B.J. Dickinson, J.H. Hewes, S.M. Butler, K. Andrejczyk, and M. Markowski-
483 Lindsay. 2016a. *USDA Forest Service National Woodland Owner Survey, 2011–2013: Design, implementation, and estimation methods*. Gen. Tech. Rep. NRS-157. Newtown
484 Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Sta-
485 tion. 43 p.
- 486
- 487 Butler, B.J., J.H. Hewes, B.J. Dickinson, K. Andrejczyk, M. Markowski-Lindsay, and
488 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,
489 Forest Service, Northern Research Station. 39 p.
- 490
- 491
- 492 Butler, B.J., J.H. Hewes, B.J. Dickinson, K. Andrejczyk, S.M. Butler, and M. Markowski-
493 Lindsay. 2016c. Family Forest Ownerships of the United States, 2013: Findings from
494 the USDA Forest Service's National Woodland Owner Survey. *J. Forest.* 114(6):638–
495 647. doi:10.5849/jof.15-099.
- 496
- 497 Casella, G., and R.L. Berger. 2002. *Statistical Inference*. second edition. Duxbury, Pacific
498 Grove, California. 660 p.
- 499
- 500 Cho, S., and D.H. Newman. 2005. Spatial Analysis of Rural Land Development. *Forest
501 Policy Econ.* 7(5): 732–744. doi:10.1016/j.forpol.2005.03.008.
- 502
- 503 Dick, P. 1995. Modelling net undercoverage in the 1991 Canadian census. *Survey Method-
504 ology* 21:45–54.
- 505
- 506 Dickinson, B.J., and B.J. Butler. 2013. Methods for Estimating Private Forest Ownership
507 Statistics: Revised Methods for the USDA Forest Service's National Woodland Owner
508 Survey. *J. Forest.* 111(5):319–325. doi:10.5849/jof.12-088.
- 509
- 510 Fay, R.E., and R.A. Herriot. 1979. Estimation of income for small places: An application
511 of James-Stein procedures to census data. *J. Am. Stat. Assoc.* 74:268–277.
- 512
- 513 Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2014. *Bayesian data analysis*, third
514 edition. Chapman & Hall/CRC Boca Raton, Florida. 661 p.

- 509 Goerndt, M.E., V.J. Monleon, and H. Temesgen. 2013. Small-area estimation of county-
510 level forest attributes using ground data and remote sensed auxiliary information. *For.*
511 *Sci.* 59(5):536–548. doi:10.5849/forsci.12-073.
- 512 Hansen, M.H., and W.N. Hurwitz. 1943. On the theory of sampling from finite popula-
513 tions. *Ann. Math. Stat.* 14(4):333–362.
- 514 Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. A comprehensive
515 change detection method for updating the National Land Cover Database to circa 2011.
516 *Remote Sens. Environ.* 132:159–175. doi:10.1016/j.rse.2013.01.012.
- 517 King, S.L., and B.J. Butler. 2005. Generating a Forest Parcel Map for Madison County,
518 NY. In, edited by Michael Bevers and Tara M. Barrett, Gen. Tech. Rep. PNW-GTR-
519 656. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest
520 Research Station 147–155.
- 521 Kittredge, D.B., A.W. D'Amato, P. Catanzaro, J. Fish, and B.J. Butler. 2008. Estimating
522 ownerships and parcels of nonindustrial private forestland in Massachusetts. *North. J.*
523 *Appl. For.* 25(2):93–98.
- 524 Magnussen, S., D. Mandallaz, J. Breidenbach, A. Lanz, and C. Ginzler. 2014. National
525 forest inventories in the service of small area estimation of stem volume. *Can. J. For.*
526 *Res.* 44(9):1079–1090. doi:10.1139/cjfr-2013-0448.
- 527 Mehmood, S.R., and D. Zhang. 2001. Forest Parcelization in the United States: A Study
528 of Contributing Factors. *J. Forest.* 99(4): 30–34.
- 529 New Jersey Office of Information Technology, Office of GIS. 2015. *New Jersey Geo-*
530 *graphic Information Network*. Available online at <https://njgin.state.nj.us/>; (accessed
531 21.12.15).
- 532 Pan, Y., Y. Zhang, and I. Majumdar. 2009. Population, economic welfare and holding
533 size distribution of private forestland in Alabama, USA. *Silva Fenn.* 43(1): 161–171.

- 534 Porter, A.T., C.K. Wikle, and S.H. Holan. 2015. Small area estimation via multivariate
535 Fay–Herriot models with latent spatial dependence. *Aust. NZ J. Stat.* 57(1): 15–29.
536 doi:10.1111/anzs.12101.
- 537 Poudyal, N.C., and D.G. Hodges. 2009. Property taxation and rural land values: Their
538 effect on private forestland ownership structure in Texas. *Land Use Policy* 26(4): 1100–
539 1105. doi:10.1016/j.landusepol.2009.01.005.
- 540 QGIS Development Team. 2014. *QGIS Geographic Information System*. Open Source
541 Geospatial Foundation. URL <http://qgis.osgeo.org>.
- 542 R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foun-
543 dation for Statistical Computing, Vienna, Austria.
- 544 Rao, J.N.K, and I. Molina. 2015. *Small area estimation*, second edition. John Wiley &
545 Sons, Inc., Hoboken, New Jersey. 441 p.
- 546 Sahr, K. 2011. DGGRID version 6.1.
- 547 US Census Bureau. 2016. *2010 Census Data*. Available online at
548 <https://www.census.gov/2010census/data>; (accessed 12.08.16).
- 549 US Geological Survey, Gap Analysis Program. 2016. *Protected Areas Database of the
550 United States (PAD-US), version 1.4 Combined Feature Class*. Available online at
551 <https://gapanalysis.usgs.gov/padus>; (accessed 29.12.15).
- 552 Ver Planck, N.R., A.O. Finley, E.S. Huff. 2017. Replication Data for: Hierarchical
553 Bayesian models for small area estimation of county-level private forest landowner
554 population. Harvard Dataverse, doi:10.7910/DVN/A3ROXD.
- 555 Ver Planck, N.R., A.L. Metcalf, A.O. Finley, and J.C. Finley. 2016. Evaluation of
556 the USDA Forest Service National Woodland Owner Survey estimators for private
557 forest area and landowners: a case study of Montana. *For. Sci.* 62(5):525–534.
558 doi:10.5849/forsci.15-196.

- 559 You, Y., and Q.M. Zhou. 2011. Hierarchical Bayes small area estimation under a spatial
560 model with application to health survey data. *Surv. Methodol.* 37(1):25–37.
- 561 Zhang Y., D. Zhang, and J. Schelhas. 2005. Small-scale non-industrial private forest
562 ownership in the United States: rationale and implications for forest management.
563 *Silva Fenn.* 39(3):443–454.

Draft

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).

Parameter	Montana		New Jersey	
	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.

Metric	Montana			New Jersey		
	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

Draft

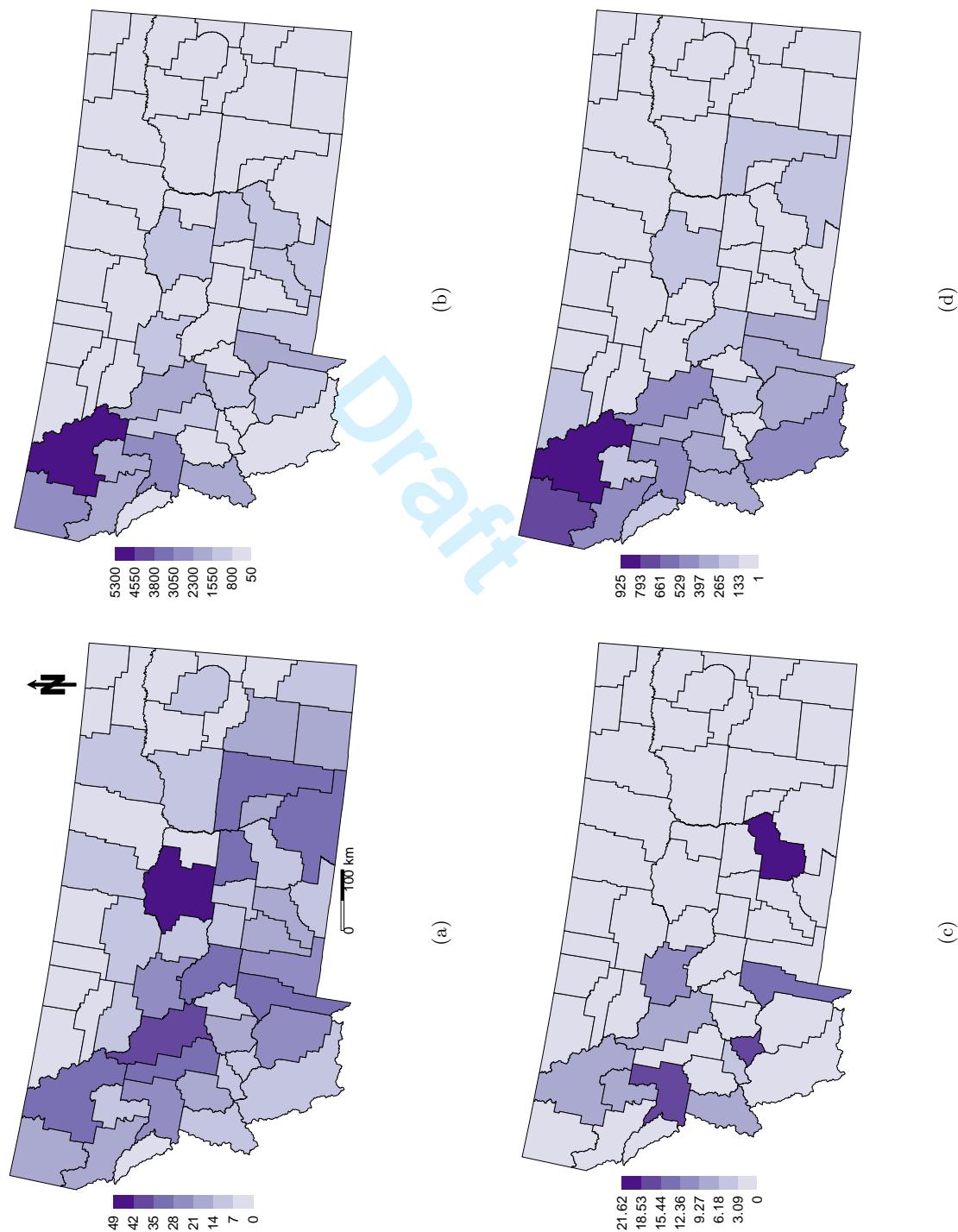
Figure 1: Montana counties by (a) sample size, (b) true population size with >2 ha forest, (c) 2010 Census population density (PD, people km^{-2}); (d) total forest area (TF, thousands ha).

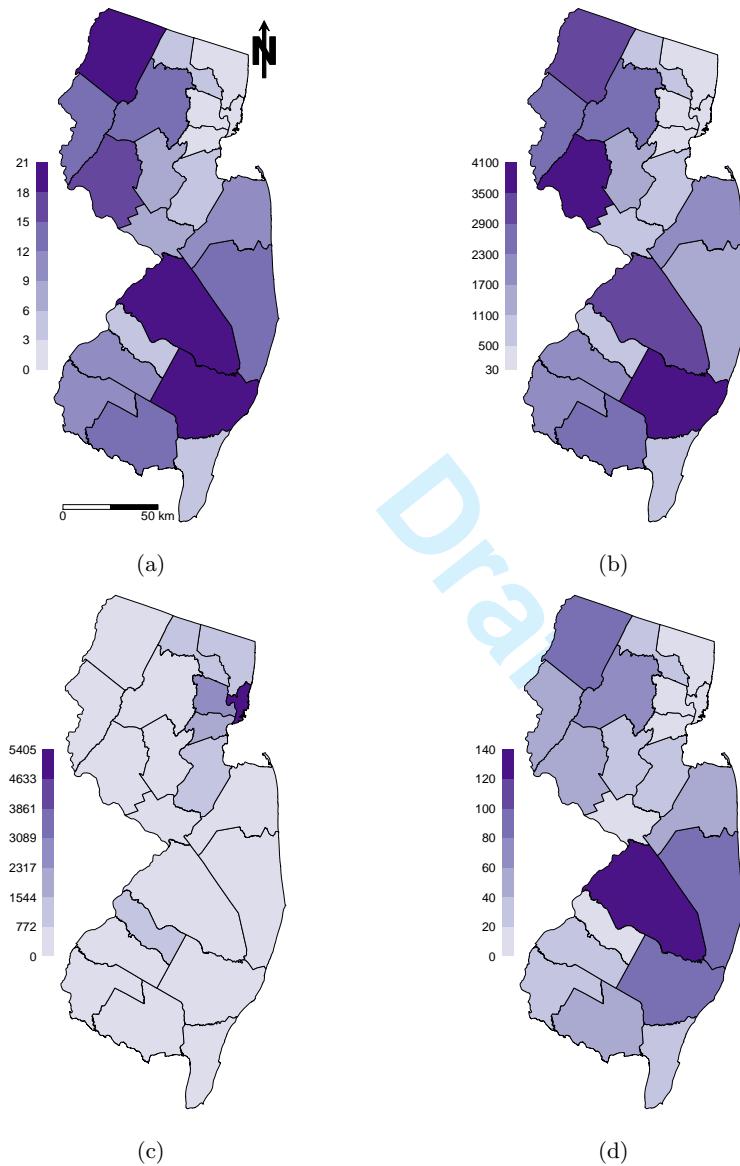
Figure 2: New Jersey counties by (a) sample size, (b) true population size with >2 ha forest, (c) 2010 Census population density (PD, people km^{-2}); (d) total forest area (TF, thousands ha).

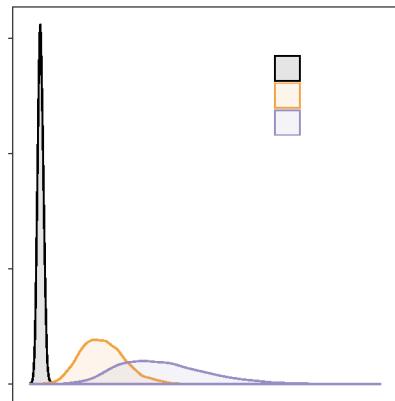
Figure 3: Distribution of posterior means from all iterations for select SAE model parameters: (a) variance parameters of Montana, (b) the FHCAR model autocorrelation parameter of Montana, (c) variance parameters of New Jersey; (d) the FHCAR model autocorrelation parameter of New Jersey.

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.



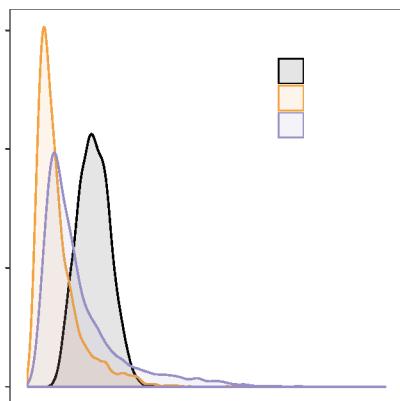




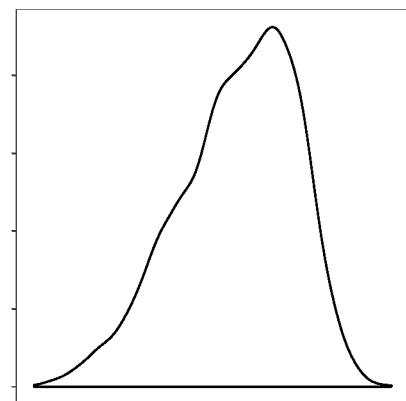
(a)



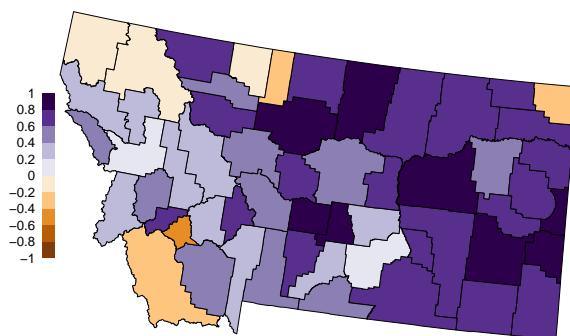
(b)



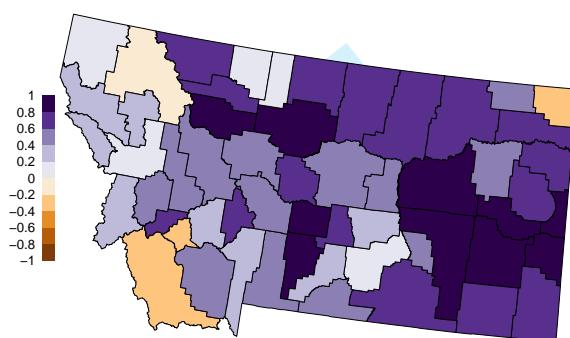
(c)



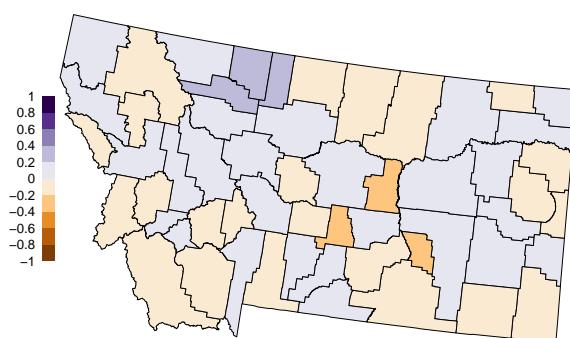
(d)



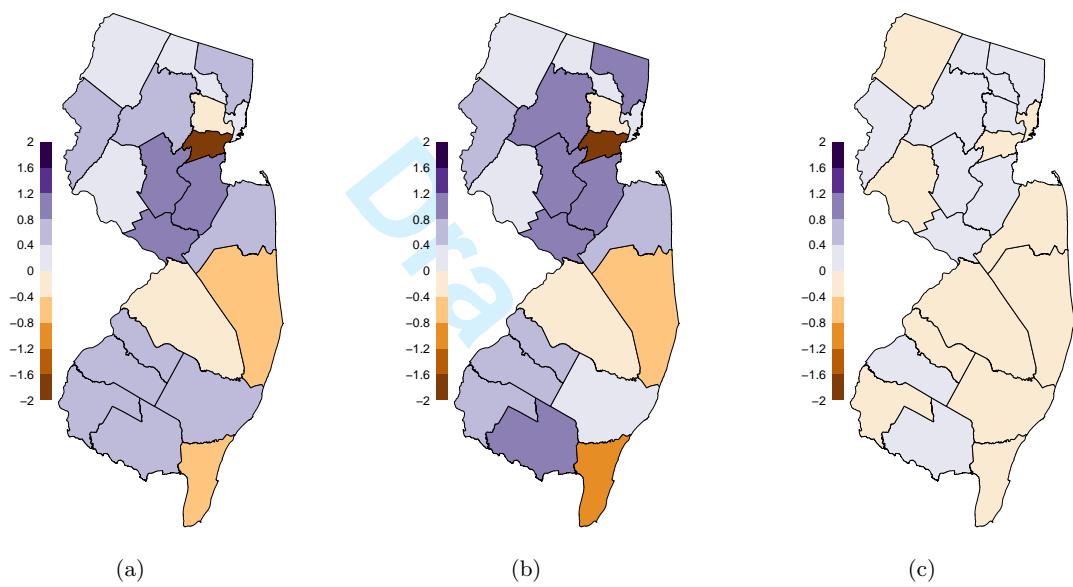
(a)



(b)



(c)



⁵⁶⁴ **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.

County ID	Name	Truth	n	Bias		
				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.

County ID	Name	Truth	n	Bias		
				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.

County ID	Name	Truth	n	Bias		
				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.

County ID	RMSE			Percent Coverage			Confidence / Credible interval width		
	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.

County ID	RMSE			Percent Coverage			Confidence / Credible interval width		
	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.

County ID	RMSE			Percent Coverage			Confidence / Credible interval width		
	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