Leveraging national forest inventory data to estimate forest carbon density status and trends for small areas

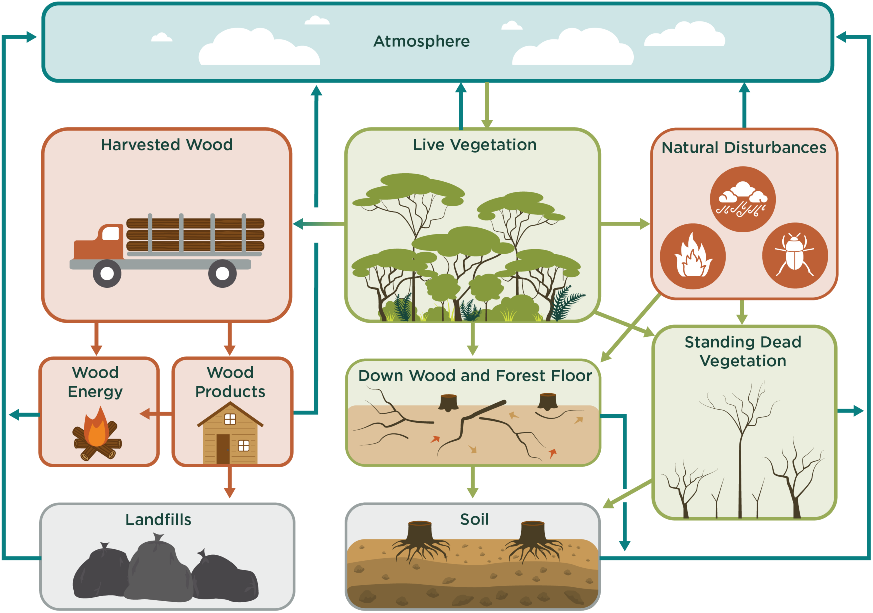
Elliot S. Shannon[1,2], Andrew O. Finley[1,2], Paul B. May[3], Grant M. Domke[4], Hans-Erik Andersen[5], George C. Gaines III[6], Arne Nothdurft[7] and Sudipto Banerjee[8]

[1]Dept of Forestry, Michigan State University, East Lansing, MI, USA.
[2]Dept of Statistics and Probability, Michigan State University, East Lansing, MI, USA.
[3]Dept of Mathematics, South Dakota School of Mines & Technology, Rapid City, SD, USA.
[4]USDA Forest Service, Northern Research Station, St. Paul, MN, USA.
[5]USDA Forest Service, Pacific Northwest Research Station, Seattle, WA, USA.
[6]USDA Forest Service, Rocky Mountain Research Station, Missoula, MT, USA.
[7]Dept of Forest and Soil Sciences, University of Natural Resources and Life Sciences, Vienna, Austria.
[8]Dept of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA.

June 17, 2025

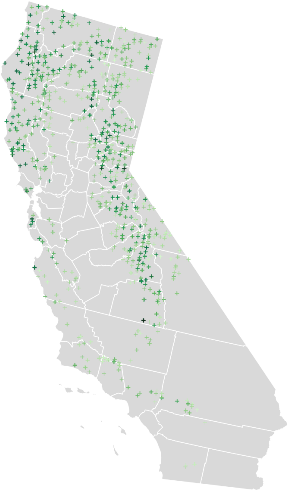
Carbon Cycle

Forest Carbon and Climate Program
Department of Forestry
MICHIGAN STATE UNIVERSITY



National Forest Inventory Data

- National Forest Inventory (NFI) programs provide critical information on forest health, sustainable management, and ecosystem change.
- Users require higher spatial and temporal resolution forest status and change parameter estimates.
- Design-based estimates are limited to large spatial and temporal scales.

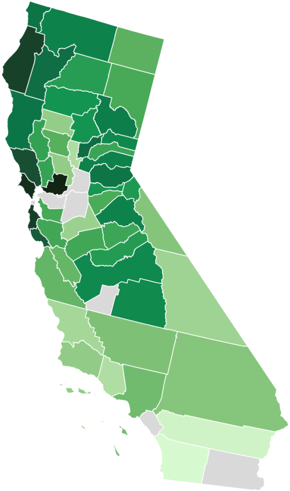


Small Area Estimation

- Small area estimation (SAE) methods have gained attention for estimating forest parameters in data-sparse settings.
  - employ statistical models to relate forest response variables to auxiliary data.
  - improve accuracy and precision over design-based approaches.

Fay-Herriot Model

- The Fay-Herriot (FH) model is widely used in SAE applications for NFI data.
  - fit to small area direct estimates.
  - does not require exact plot locations.
- Direct estimates are often missing when sample sizes are too small or measurements are homogeneous.



Bayesian Spatio-Temporal SAE Model

We propose a Bayesian spatio-temporal SAE model of live forest carbon density (LFCD) that

- directly uses NFI plot-level measurements,
- incorporates auxiliary covariates,
- accommodates spatially and temporally varying regression coefficients,
- appropriately quantifies uncertainty.

Data

- We have 593,368 United States Forest Service Forest Inventory and Analysis (FIA) plot measurements collected across 3,108 counties in the CONUS from 2008 to 2021.
  - Exact plot locations are unknown, but plot measurements may be assigned to counties.
- We leverage remotely sensed percent tree canopy cover (TCC) as a covariate.
  - Averaged among all pixels within a given county and year.

Notation

Let

- $j = 1, \ldots, J$ index counties,
- $t = 1, \ldots, T$ index discrete years,
- $i = 1, \ldots, n_{j,t}$ index FIA plots measured in county j in year t,
  - Note, we may have $n_{j,t} = 0$ for some j and t.
- $y_{i,j,t}$ be the LFCD (Mg/ha) at FIA plot i in county j in year t,
- $\mu_{j,t}$ be the latent (unobserved) mean LFCD for county j in year t,
- $\mathbf{x}_{j,t} = (1, x_{1,j,t}, \ldots x_{P,j,t})^{\mathrm{T}}$ be the length $P+1$ vector of covarites associated with county j in year t,
- $\tilde{\mathbf{x}}_{j,t} = (\tilde{x}_{1,j,t}, \ldots \tilde{x}_{Q,j,t})^{\mathrm{T}}$ be the length $Q \subseteq P$ vector of covariates with space-varying impact on $\mu_{j,t}$.

Model

For county j in year t, the proposed model is then

$$y_{i,j,t} = \underbrace{\mathbf{x}_{j,t}^{\mathrm{T}}\boldsymbol{\beta}_t + \tilde{\mathbf{x}}_{j,t}^{\mathrm{T}}\boldsymbol{\eta}_j + u_{j,t}}_{\mu_{j,t}} + \varepsilon_{i,j,t}, \quad i = 1, \ldots, n_{j,t}, \tag{1}$$

where

- $\varepsilon_{i,j,t} \overset{\text{ind}}{\sim} N(0, \sigma_t^2)$,
- $\boldsymbol{\beta}_t$ is a length $P+1$ vector of temporally-varying regression coefficients,
- $\boldsymbol{\eta}_j$ is a length Q vector of space-varying regression coefficients,
- $u_{j,t}$ is a dynamically evolving spatio-temporal intercept term.

Temporally-varying Regression Coefficients

$\boldsymbol{\beta}_t$ is modeled dynamically as

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\xi}_t, \text{with} \tag{2}$$

$$\boldsymbol{\xi}_t \overset{\text{ind}}{\sim} MVN\left(\mathbf{0}, \boldsymbol{\Sigma}_\xi\right), \quad t = 1, \ldots, T, \tag{3}$$

which allows the effect of covariates in $\mathbf{x}_{j,t}$ to have time-varying impact on the response according to the covariance structure in $\boldsymbol{\Sigma}_\xi$.

Space-varying Regression Coefficients

Writing $\boldsymbol{\eta}_j = (\eta_{1,j}, \ldots, \eta_{Q,j})^{\mathrm{T}}$ and collecting $\boldsymbol{\eta}_q^* = (\eta_{q,1}, \ldots, \eta_{q,J})^{\mathrm{T}}$, we model $\boldsymbol{\eta}_q^*$ as a conditional autoregressive (CAR) random effect,

$$\boldsymbol{\eta}_q^* \sim MVN\left(\mathbf{0}, \tau_{\eta,q}^2 \mathbf{Q}(\rho_{\eta,q})\right), \quad q = 1, \ldots, Q, \tag{4}$$

where

- $\tau_{\eta,q}^2$ is a scalar variance parameter,
- $\rho_{\eta,q}$ is a spatial correlation parameter $(0 < \rho_{\eta,q} < 1)$,
- $\mathbf{Q}(\rho_{\eta,q})$ is a $J \times J$ correlation matrix reflecting the county neighborhood structure. (See Banerjee et al. (2004) for details).

Dynamic Spatio-temporal Intercept

$u_{j,t}$ is modeled as a dynamically evolving CAR spatial random effect,

$$u_{j,t} = u_{j,t-1} + \omega_{j,t}, \tag{5}$$

where $u_{j,0} \equiv 0$ for all j.

Then, collecting all $\omega_{j,t}$ for time t as $\boldsymbol{\omega}_t = (\omega_{1,t}, \ldots, \omega_{J,t})^{\mathrm{T}}$, we specify a CAR spatial structure for $\boldsymbol{\omega}_t$ as

$$\boldsymbol{\omega}_t \sim MVN\left(\mathbf{0}, \tau^2_{\omega,t}\mathbf{Q}(\rho_\omega)\right). \tag{6}$$

Direct Estimates

- Traditionally, NFI-derived quantities of interest have been estimated using design-based direct estimates.

Specifically, the direct estimate mean for $\mu_{j,t}$ is calculated as

$$\hat{\mu}_{j,t} = \frac{1}{n_{j,t}} \sum_{i=1}^{n_j t} y_{i,j,t}, \tag{7}$$

with associated estimate variance

$$\hat{\sigma}^2_{j,t} = \frac{1}{n_{j,t}(n_{j,t} - 1)} \sum_{i=1}^{n_j t} (y_{i,j,t} - \hat{\mu}_{j,t})^2. \tag{8}$$

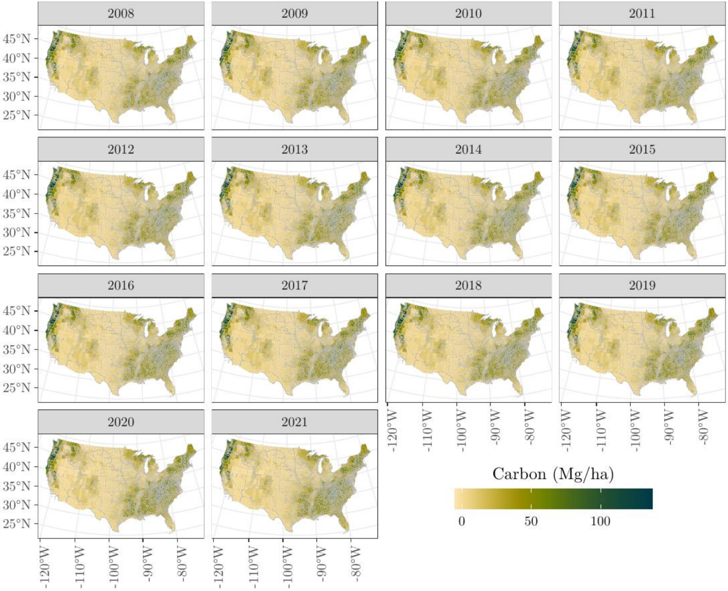


Figure 1: Posterior mean values of live forest carbon density ($\mu_{j,t}$).

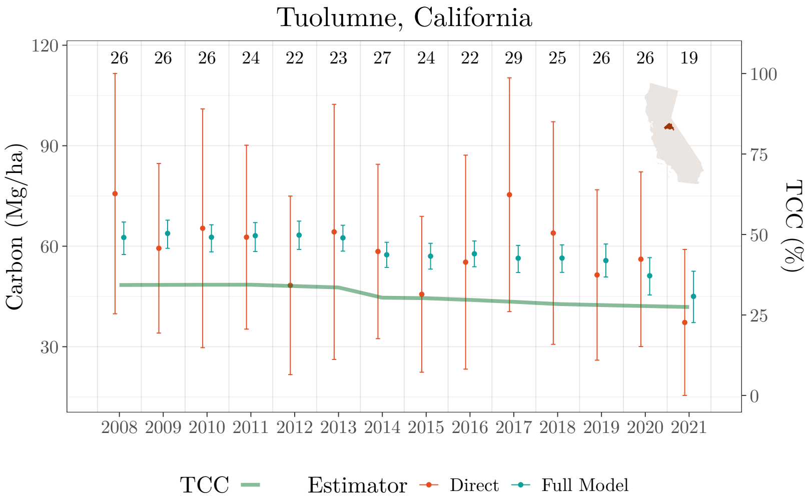


Figure 2: Posterior mean and 95% credible intervals of LFCD ($\mu_{j,t}$) for Tuolumne County, California, compared to direct estimate means ($\hat{\mu}_{j,t}$) and 95% confidence intervals over time. Top row displays sample sizes ($n_{j,t}$).

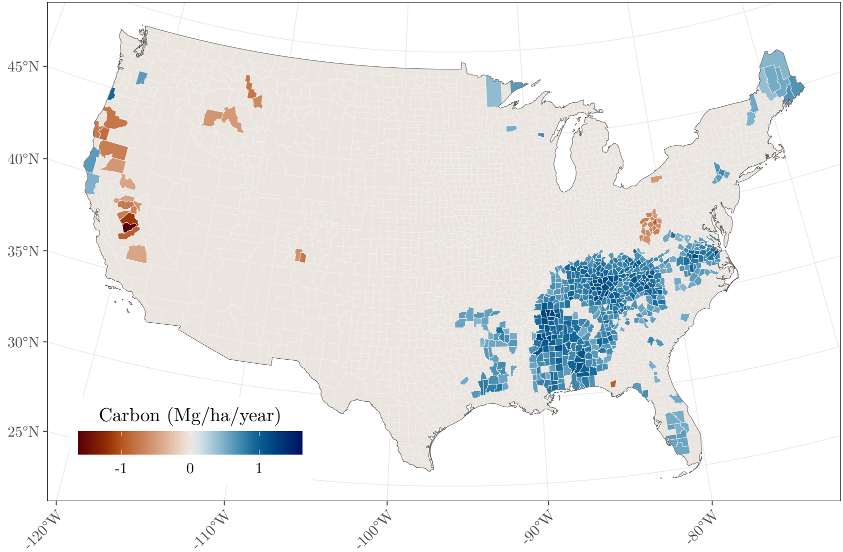


Figure 3: Significant live forest carbon density trends (Mg/ha/year).

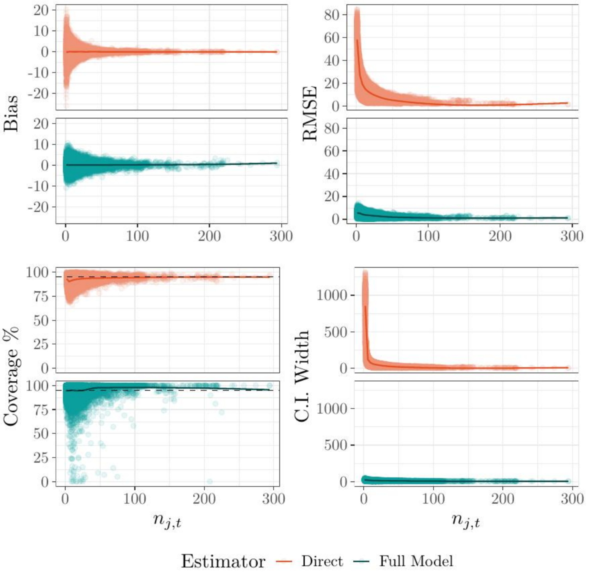
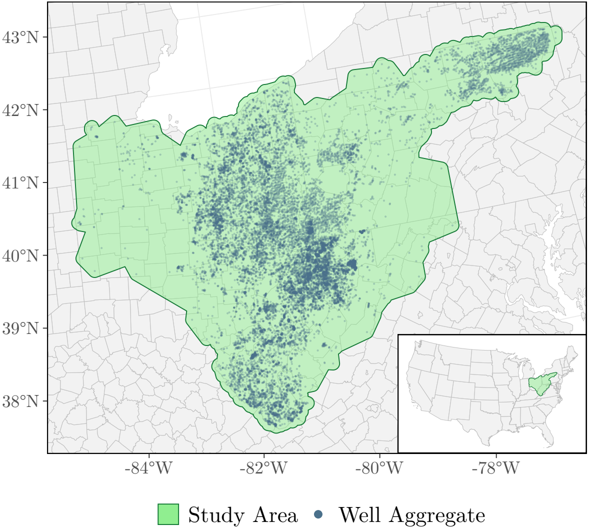


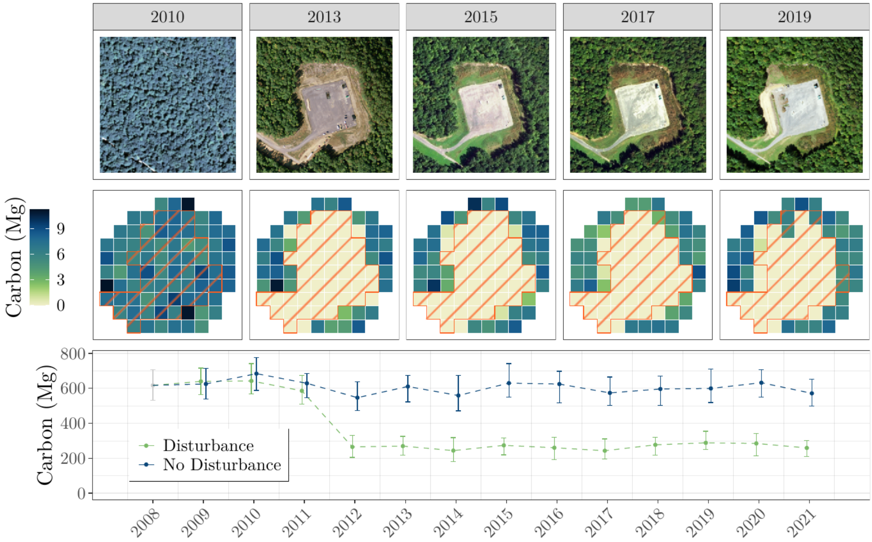
Figure 4: Average measures of bias, root mean square error (RMSE), coverage percentage, and coverage interval widths for the model and direct estimator arranged according to sample size $n_{j,t}$.

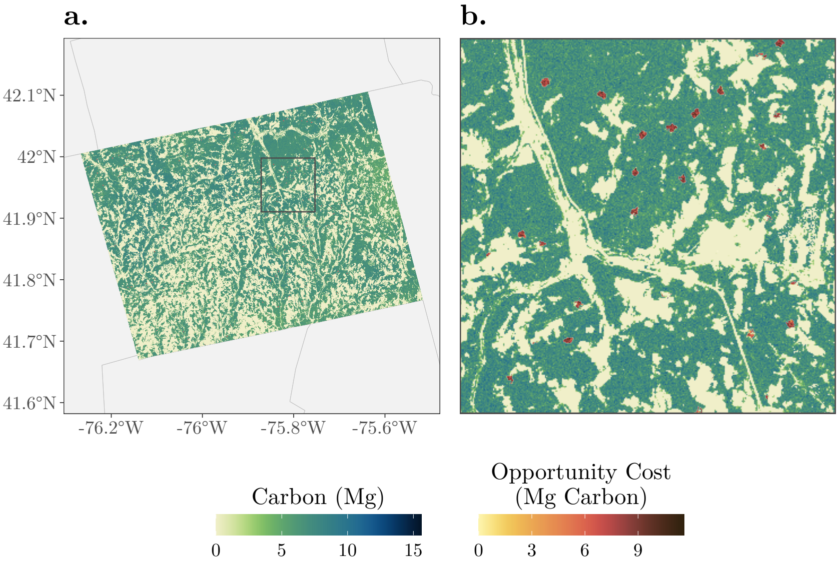
Each point represents the mean metric value for estimating $\mu_{j,t}$ averaged over $R = 100$ simulated population replicates.

Shannon et al. (2025)

Looking Forward




2010
2013
2015
2017
2019
Carbon (Mg)
9
6
3
0
800
600
400
200
0
Carbon (Mg)
Disturbance
No Disturbance
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021


a.
b.
42.1°N
42°N
41.9°N
41.8°N
41.7°N
41.6°N
-76.2°W
-76°W
-75.8°W
-75.6°W
Carbon (Mg)
0
5
10
15
Opportunity Cost
(Mg Carbon)
0
3
6
9

References

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*, volume 101.

Shannon, E. S., Finley, A. O., May, P. B., Domke, G. M., Andersen, H.-E., III, G. C. G., Nothdurft, A., and Banerjee, S. (2025). Leveraging national forest inventory data to estimate forest carbon density status and trends for small areas.