

# Bayesian spatio-temporal models for forest inventory small area estimation

**Elliot S. Shannon<sup>1,2</sup>**

<sup>1</sup>Dept of Forestry, Michigan State University, East Lansing, MI, USA.

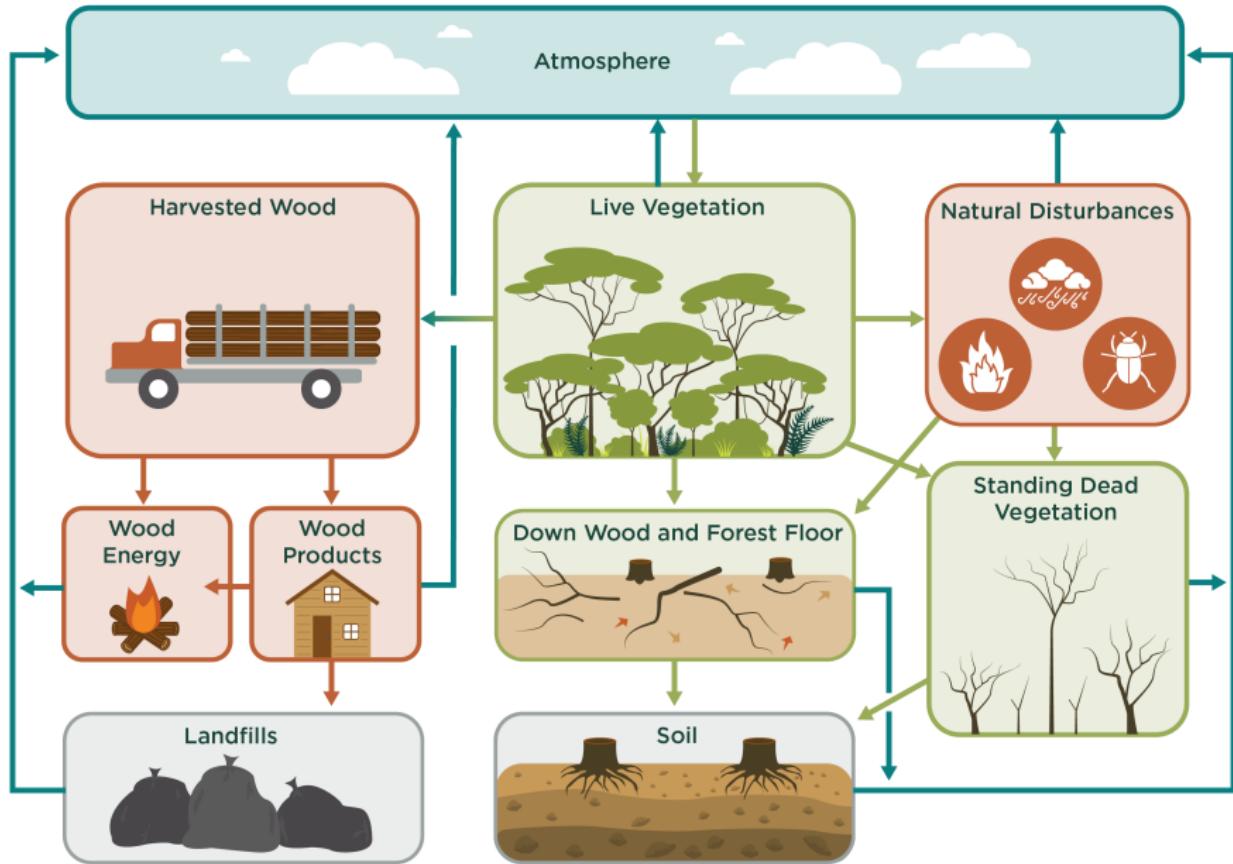
<sup>2</sup>Dept of Statistics and Probability, Michigan State University, East Lansing, MI, USA.

October 28, 2025

# Outline

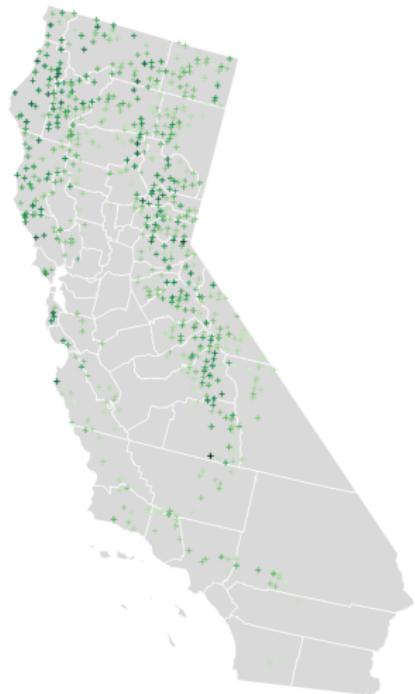
- Chapter 1: Introduction
- Chapter 2: Spatio-temporal Small Area Estimation Model
- Chapter 3: Improved Model
- Chapter 4: Natural Gas Application
- Chapter 5: Wildfire Application
- Chapter 6: Conclusion

# Carbon Cycle



# Introduction: National Forest Inventory and FIA

- Forest carbon estimates are based on National Forest Inventory (NFI) data and estimators.
- The US NFI is conducted by the USDA Forest Service Forest Inventory and Analysis (FIA) program.
- FIA maintains 300,000 inventory plots across the contiguous US (CONUS), which are measured on rotating basis every 5 to 10 years.



# Introduction: FIA Design

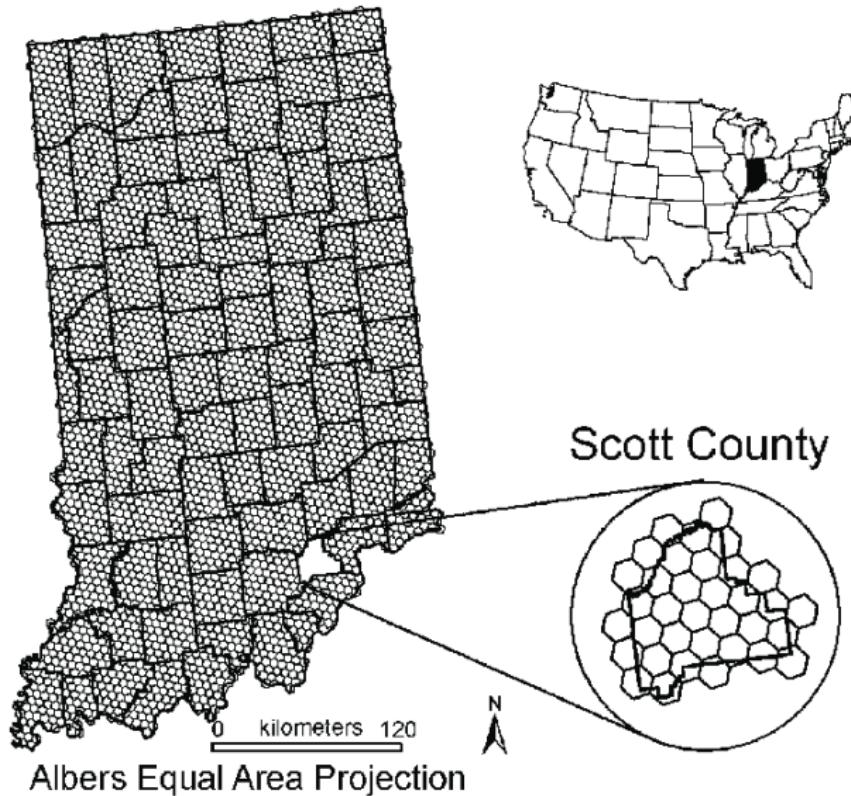


Figure 1: Mcroberts et al. (2005)

# Introduction: FIA Design

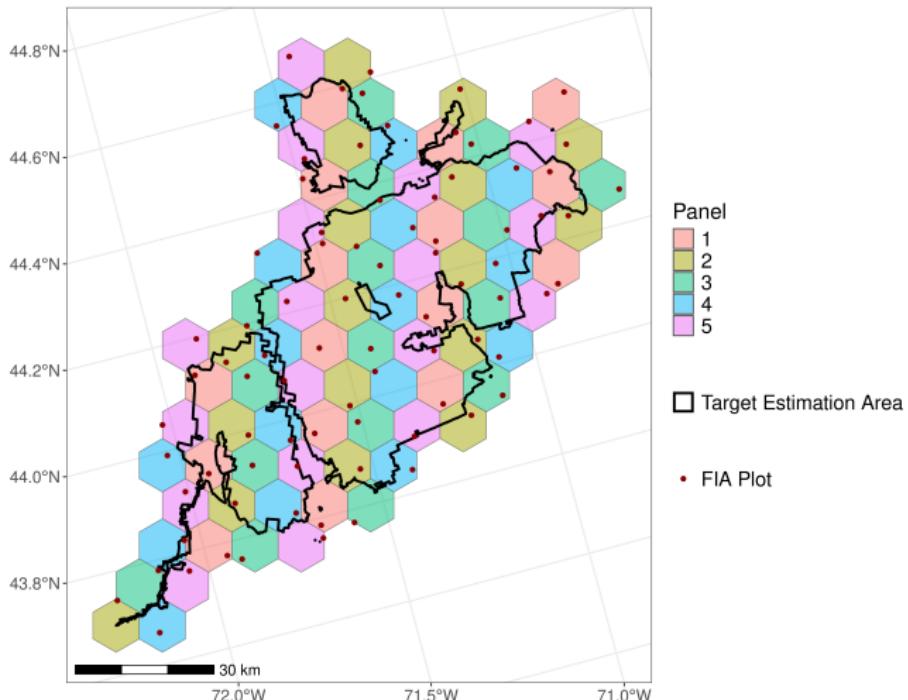


Figure 2: FIA panel design (May and Finley, 2025).

# Introduction: Design-based Estimation

- Traditionally, NFI programs provide design-based estimates for forest parameters based on forest inventory plot measurements.
- Design-based estimates assume a fixed finite population, with population parameters accessible without error if all population units are observed.
- To achieve desired levels of estimate precision, these methods require repeated costly measurements from a dense network of inventory plots.
- Given the high costs associated with data acquisition, FIA plot measurements are sparse, limiting reliable design-based estimates to large spatial and temporal scales.

# Introduction: Demand for Small Area Estimates

Increasingly, users groups require higher spatial and temporal resolution forest status and change parameter estimates to evaluate existing land use policies and management practices, and inform future activities.

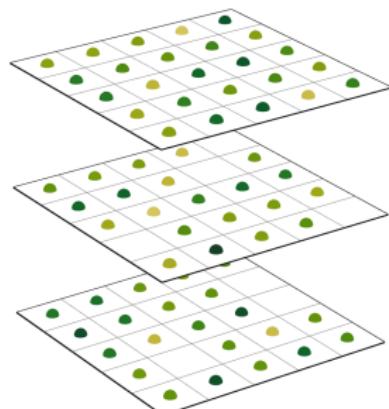
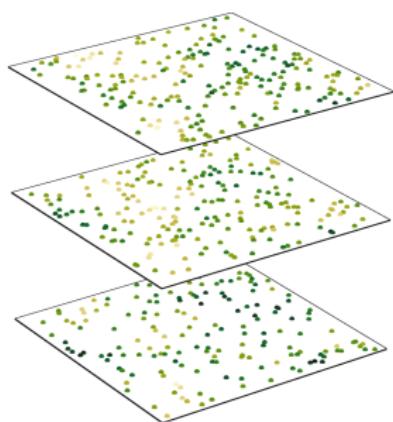


# Introduction: Small Area Estimation

- Model-based small area estimation (SAE) methods have gained attention for estimating forest parameters in data-sparse settings (Schroeder et al., 2014; Lister et al., 2020; Hou et al., 2021; Coulston et al., 2021; Finley et al., 2024; Shannon et al., 2024)
- SAE methods employ statistical models to relate forest response variables to auxiliary data.
- Accuracy and precision are improved over design-based approaches when strong relationships exist between response variables and auxiliary information.

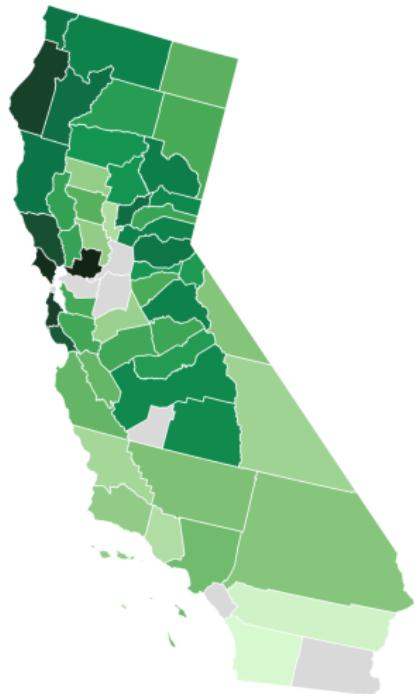
# Introduction: Unit-level vs. Area-level SAE methods

- **Unit-level** models are constructed at the population unit level (inventory plots), which is the minimal unit that can be sampled from a population.
- **Area-level** models characterize the relationship between area-specific (design-based) direct estimates and auxiliary data.



# Introduction: 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.



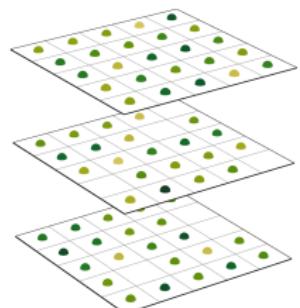
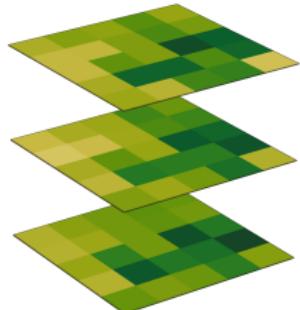
# Introduction: Direct Estimates

- For area  $j$  at time  $t$ , our interest lies in the latent (unobserved) mean  $\mu_{j,t}$ .
- Let  $y_{i,j,t}$  be individual plot measurements,  $i = 1, \dots, n_{j,t}$ .
- The direct estimate mean for area  $j$  at time  $t$  is

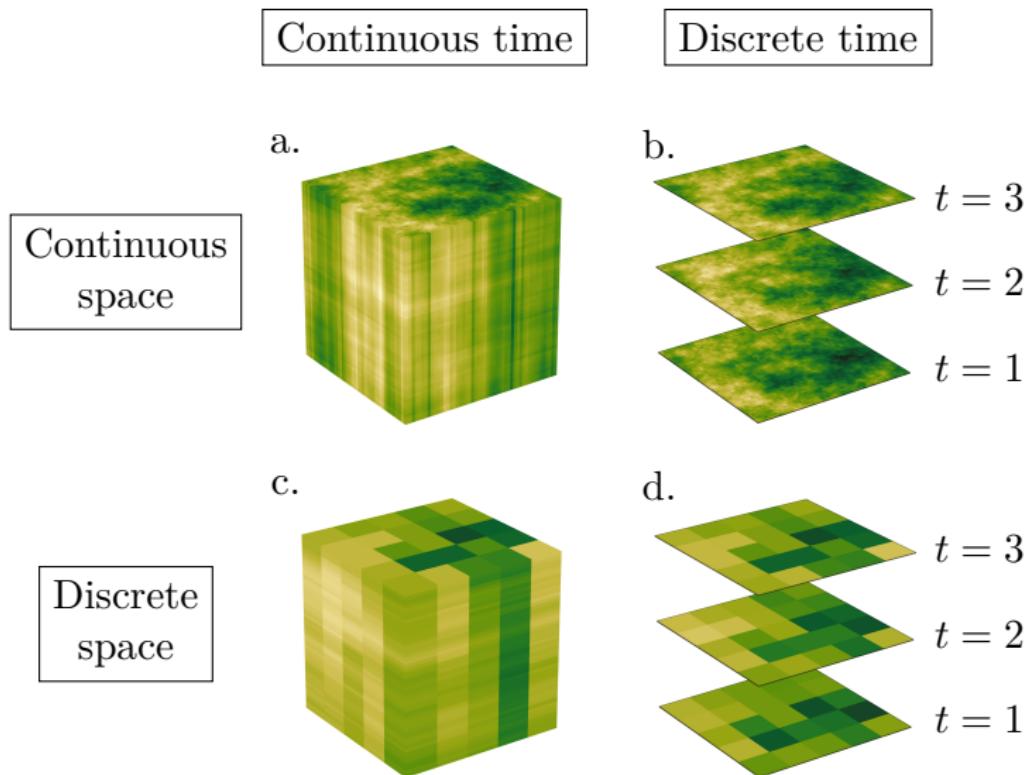
$$\hat{\mu}_{j,t} = \frac{1}{n_{j,t}} \sum_{i=1}^{n_{j,t}} y_{i,j,t}. \quad (1)$$

- The associated estimate variance is

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



# Introduction: Spatio-temporal Models



## Chapter 2: Inferential Goals

Provide CONUS county estimates of live aboveground forest carbon density for 2008-2021 using FIA county-level direct (i.e., design-based) estimates.

Desired qualities of the estimator:

- ① leverage spatially and temporally proximate information from FIA county estimates and ancillary data to improve estimate accuracy and precision,
- ② provide robust and flexible uncertainty quantification,
- ③ facilitate statistically defensible exploration of carbon status, trend, and change,
- ④ accommodate missing direct estimates,
- ⑤ scale to allow for a large number of areal units and time steps.

## Chapter 2: Data

Data for CONUS 2008-2021:

- ① 100k+ plot-level aboveground live tree carbon measurements expressed on as Mg/ha,
- ② a complete set of annual county-level predictor variables.

Let  $y_{i,j,t}$  represent the observed carbon density for FIA plot  $i$  in county  $j$  and year  $t$ , where  $i = 1, \dots, n_{j,t}$ ,  $j = 1, \dots, J$ , and  $t = 1, \dots, T$ , with  $n_{j,t}$  being the number of FIA plots measured in county  $j$  and year  $t$ .

We consider  $J = 3,108$  counties and  $T = 14$  years.

A set of  $p$  annual county-level predictors  $x_{k,j,t}$  for  $k = 1, \dots, p$ .

# Chapter 2: Data

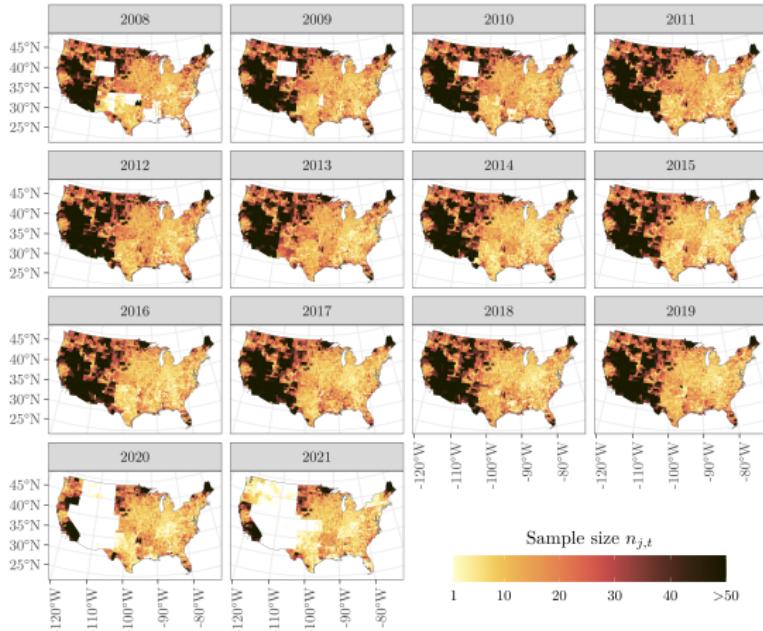


Figure 3: Number of observed FIA plots within each county and year ( $n_{j,t}$ ). Transparent counties have zero observed FIA plots.

# Chapter 2: Data

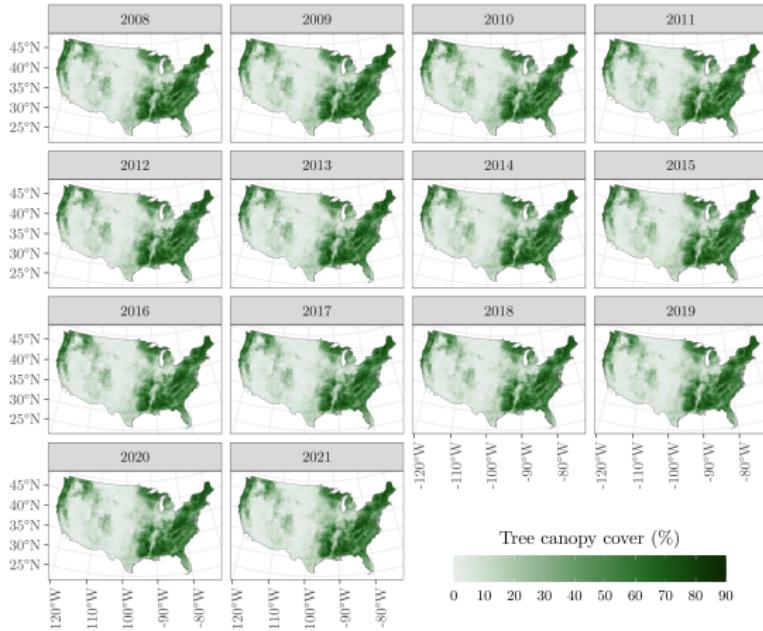


Figure 4: Annual National Land Cover Database percent tree canopy cover (TCC) averaged within each county.

## Chapter 2: Model

The model we propose is an extension to the traditional Fay-Herriot model. For county  $j$  at time  $t$  the model is

$$\hat{\mu}_{j,t} = \mu_{j,t} + \delta_{j,t}, \quad (3)$$

$$\mu_{j,t} = \beta_0 + \eta_{0,j,t} + \sum_{k=1}^p x_{k,j,t} \beta_k + \sum_{k=1}^q \tilde{x}_{k,j,t} \eta_{k,j,t} + \epsilon_{j,t}, \quad (4)$$

where  $\delta_{j,t}$  and  $\epsilon_{j,t}$  are mutually exclusive error terms with  $\delta_{j,t} \stackrel{\text{ind}}{\sim} N(0, \sigma_{\delta,j,t}^2)$  and  $\epsilon_{j,t} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\epsilon}^2)$ .

Here,  $\delta_{j,t}$ 's variance  $\sigma_{\delta,j,t}^2 \sim IG\left(\frac{n_{j,t}}{2}, \frac{(n_{j,t}-1)\hat{\sigma}_{j,t}^2}{2}\right)$ . Modeling  $\sigma_{\delta,j,t}^2$  in this way provides a more coherent hierarchical model and allows us to obtain potential information from the observed sample size  $n_{j,t}$ .

## Chapter 2: Spatial random effects

Consider specifications for the  $(q + 1)$  spatial and temporal random effects, i.e., intercept and  $q$  for those predictor variables with space-and/or time-varying relationships with the outcome.

We define a  $J \times 1$  vector of spatial random effects as

$$\boldsymbol{\eta}^s \sim MVN \left( \mathbf{0}, \sigma_{\eta^s}^2 \mathbf{R}(\rho_{\eta^s}) \right), \quad (5)$$

where  $\sigma_{\eta^s}^2$  is the scalar variance,  $\rho_{\eta^s}$  is the correlation parameter, and  $\mathbf{R}(\rho_{\eta^s}) = (\mathbf{D} - \rho_{\eta^s} \mathbf{W})^{-1}$  is the  $J \times J$  correlation matrix reflecting a conditional autoregressive (CAR) spatial structure, see, e.g., Banerjee et al. (2004).

## Chapter 2: Spatial random effects

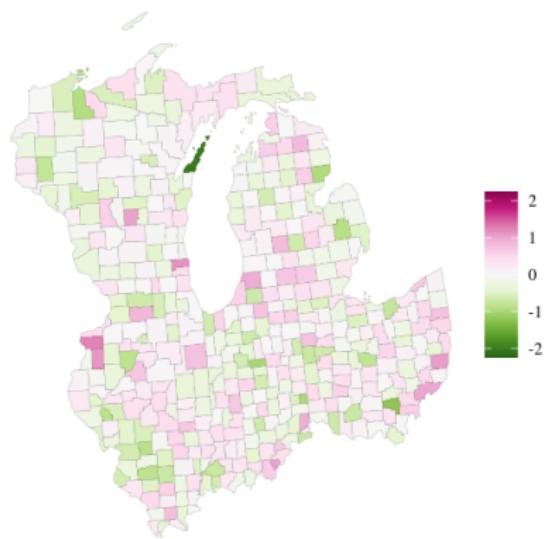


Figure 5:  $\sigma_{\eta^s}^2 = 1$  and  $\rho_{\eta^s} = 0.2$ .

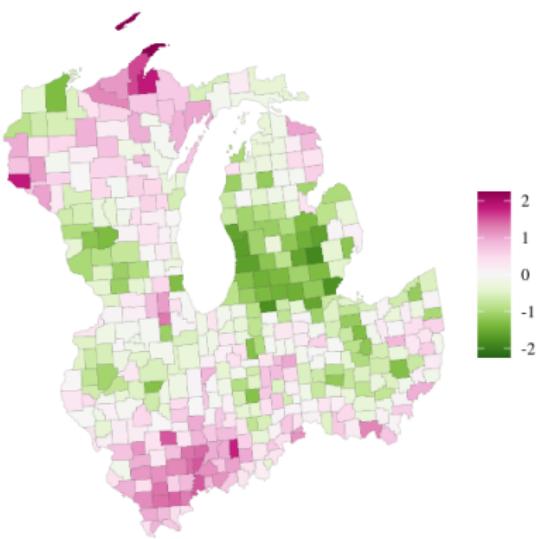


Figure 6:  $\sigma_{\eta^s}^2 = 1$  and  $\rho_{\eta^s} = 0.99$ .

## Chapter 2: Temporal random effects

When collecting all  $N = JT$  space and time observations (stacked by county), we define a  $N \times 1$  vector of temporal random effects as

$$\boldsymbol{\eta}^t \sim MVN \left( \mathbf{0}, \sigma_{\eta^t}^2 \mathbf{I} \otimes \mathbf{A}(\alpha_{\eta^t}) \right), \quad (6)$$

where  $\sigma_{\eta^t}^2$  is a scalar variance,  $\mathbf{I}$  is a  $J \times J$  identity matrix,  $\otimes$  is the Kronecker product operator, and  $\mathbf{A}(\alpha_{\eta^t})$  is a  $T \times T$  first order autoregressive correlation matrix with temporal correlation parameter  $\alpha_{\eta^t}$ .

This specification allows for county-specific temporal effects, but does not accommodate spatial association between counties.

# Chapter 2: Temporal random effects

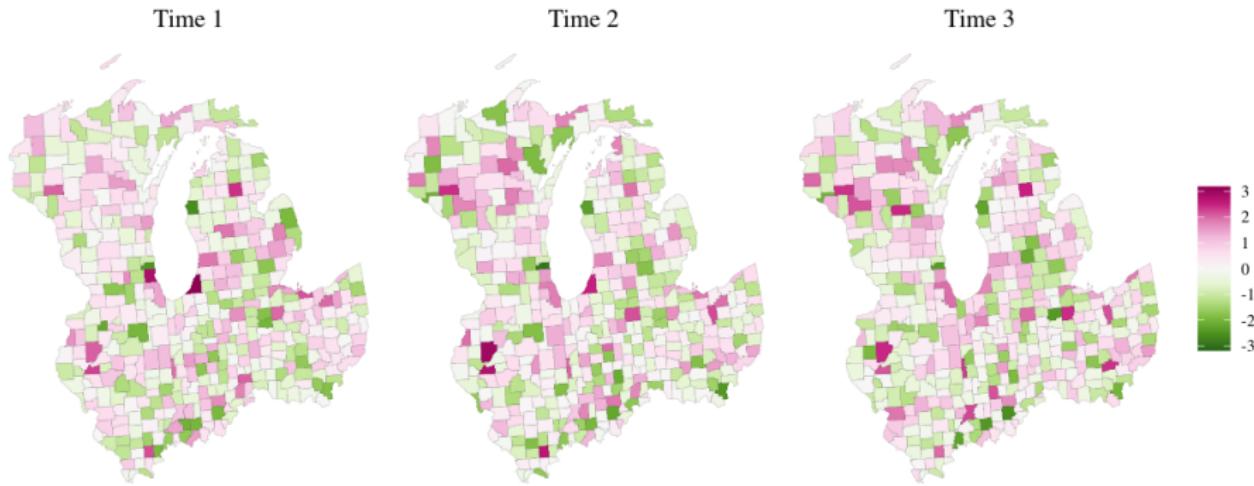


Figure 7:  $\sigma_{\eta^t}^2 = 1$  and  $\alpha_{\eta^t} = 0.8$ .

## Chapter 2: Spatio-temporal random effects

Finally, we define a  $N \times 1$  vector of spatial-temporal random effects as

$$\boldsymbol{\eta}^{st} \sim MVN \left( \mathbf{0}, \sigma_{\eta^{st}}^2 \mathbf{R}(\rho_{\eta^{st}}) \otimes \mathbf{A}(\alpha_{\eta^{st}}) \right), \quad (7)$$

where  $\sigma_{\eta^{st}}^2$  is a scalar variance, and all other terms were defined previously.

This specification allows for a county-specific spatial effect to evolve over time.

# Chapter 2: Spatio-temporal random effects

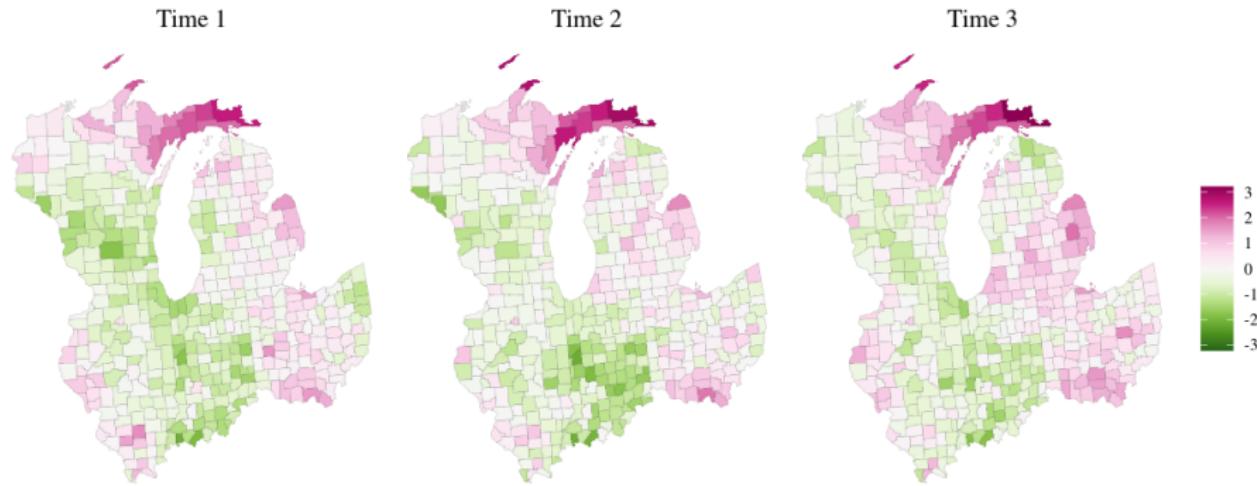


Figure 8:  $\sigma_{\eta^{st}}^2 = 1$ ,  $\rho_{\eta^{st}} = 0.99$ , and  $\alpha_{\eta^{st}} = 0.8$ .

## Chapter 2: Candidate models

In subsequent analyses, we consider the following candidate models for  $\mu_{j,t}$ .

Each model uses a single predictor variable  $x_{TCC,j,t}$ , which is percent tree canopy cover (TCC) per county and year (Housman et al., 2023).

$$\text{Full model: } \mu_{j,t} = \beta_0 + \eta_{0,j,t}^{st} + x_{TCC,j,t}\beta_{TCC} + \tilde{x}_{TCC,j,t}\eta_{TCC,j}^s + \epsilon_{j,t}$$

$$\text{Submodel 1: } \mu_{j,t} = \beta_0 + \eta_{0,j,t}^{st} + x_{TCC,j,t}\beta_{TCC} + \epsilon_{j,t}$$

$$\text{Submodel 2: } \mu_{j,t} = \beta_0 + \eta_{0,j,t}^t + x_{TCC,j,t}\beta_{TCC} + \epsilon_{j,t}$$

## Chapter 2: Full Model

To complete the Bayesian model specification, we assign prior distributions to the model parameters. For the full model, the joint posterior distribution for all parameters is proportional to

$$\begin{aligned} & \prod_{j=1}^J \prod_{t=1}^T N\left(\hat{\mu}_{j,t} \mid \mu_{j,t}, \sigma_{j,t}^2\right) \times \prod_{j=1}^J \prod_{t=1}^T N\left(\mu_{j,t} \mid \beta_0 + \eta_{0,j,t}^{st} + \sum_{k=1}^p x_{k,j,t} \beta_k + \sum_{k=1}^q \tilde{x}_{k,j,t} \eta_{k,j}^s, \sigma_\epsilon^2\right) \times \\ & \prod_{k=0}^p N\left(\beta_k \mid \mu_\beta, \sigma_\beta^2\right) \times \prod_{j=1}^J \prod_{t=1}^T IG\left(\sigma_{j,t}^2 \mid \frac{n_{j,t}}{2}, \frac{(n_{j,t}-1) \hat{\sigma}_{j,t}^2}{2}\right) \times IG\left(\sigma_\epsilon^2 \mid a_\epsilon, b_\epsilon\right) \times \\ & MVN\left(\boldsymbol{\eta}_0^{st} \mid \mathbf{0}, \sigma_{\eta^{st}}^2 \mathbf{R}(\rho_{\eta^{st}}) \otimes \mathbf{A}(\alpha_{\eta^{st}})\right) \times \\ & IG\left(\sigma_{\eta_0^{st}}^2 \mid a_{\eta_0^{st}}, b_{\eta_0^{st}}\right) \times U\left(\rho_{\eta_0^{st}} \mid a_\rho, b_\rho\right) \times U\left(\alpha_{\eta_0^{st}} \mid a_\alpha, b_\alpha\right) \times \\ & \prod_{k=1}^q MVN\left(\boldsymbol{\eta}_k^s \mid \mathbf{0}, \sigma_{\eta_k^s}^2 \mathbf{R}(\rho_{\eta_k^s})\right) \times \prod_{k=1}^q IG\left(\sigma_{\eta_k^s}^2 \mid a_{\eta^s}, b_{\eta^s}\right) \times \prod_{k=1}^q U\left(\rho_{\eta_k^s} \mid a_\rho, b_\rho\right). \end{aligned} \tag{8}$$

## Chapter 2: Simulation Study

- A single population is generated using fixed and known values for parameters, then estimates for parameters are computed from each of a large number of independent samples, i.e.,  $R$  replicates, taken from the population.
- Estimates are then compared with population parameters using a set of measures that assess the estimators' bias, accuracy, and precision.

## Chapter 2: Simulation Study

- To mimic qualities of the observed annual county-level data, we simulated a population comprising 7,809,952 point-referenced population units laid out in a 1-by-1 (km) regular grid over the CONUS land area.
- Specifically, the outcome  $y_t(\ell)$  at generic population unit location  $\ell$  and time  $t$  is given by

$$y_t(\ell) = \zeta_0 + u_t(\ell) + \zeta_1 v_{TCC,t}(\ell) + \epsilon_t(\ell), \quad \epsilon_t(\ell) \stackrel{\text{iid}}{\sim} N(0, \sigma_y^2),$$

$$u_t(\ell) = u_{t-1}(\ell) + w_t(\ell), \quad u_0(\ell) = 0,$$

$$w_t(\ell) \stackrel{\text{ind}}{\sim} GP(0, C(\cdot, \gamma)), \quad t = 1, 2, \dots, T,$$

- $C(\cdot, \gamma) = \sigma_w^2 \exp(-\gamma ||\ell - \ell'||)$

# Chapter 2: Simulation Study

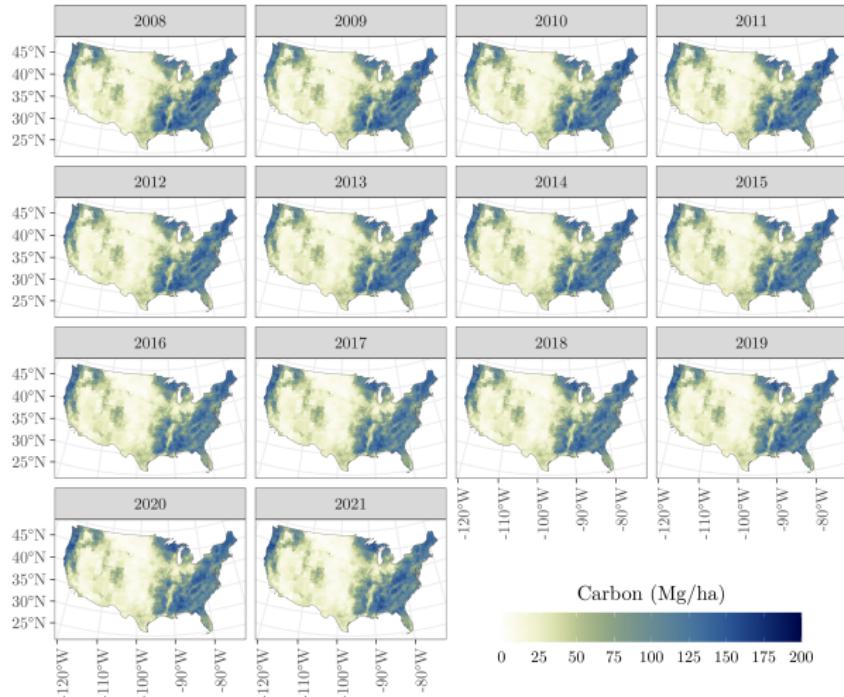


Figure 9: Simulated population forest carbon density  $\mu_{\text{true},t,j}$  (Mg/ha).

## Chapter 2: Simulation Study

Given  $\mu_{true,j,t}$  and  $R$  estimates from each estimator,

$$\text{Bias}_{j,t,l} = \frac{\sum_{r=1}^R (\mu_{j,t,l,r} - \mu_{true,j,t})}{R}, \quad (9)$$

$$\text{RMSE}_{j,t,l} = \sqrt{\frac{\sum_{r=1}^R (\mu_{j,t,l,r} - \mu_{true,j,t})^2}{R}}. \quad (10)$$

$$\text{Coverage}_{j,t,l} = \frac{\sum_{r=1}^R I(\mu_{j,t,l,r}^L \leq \mu_{true,j,t} \leq \mu_{j,t,l,r}^U)}{R}, \quad (11)$$

$$\text{Width}_{j,t,l} = \frac{\sum_{r=1}^R (\mu_{j,t,l,r}^U - \mu_{j,t,l,r}^L)}{R}. \quad (12)$$

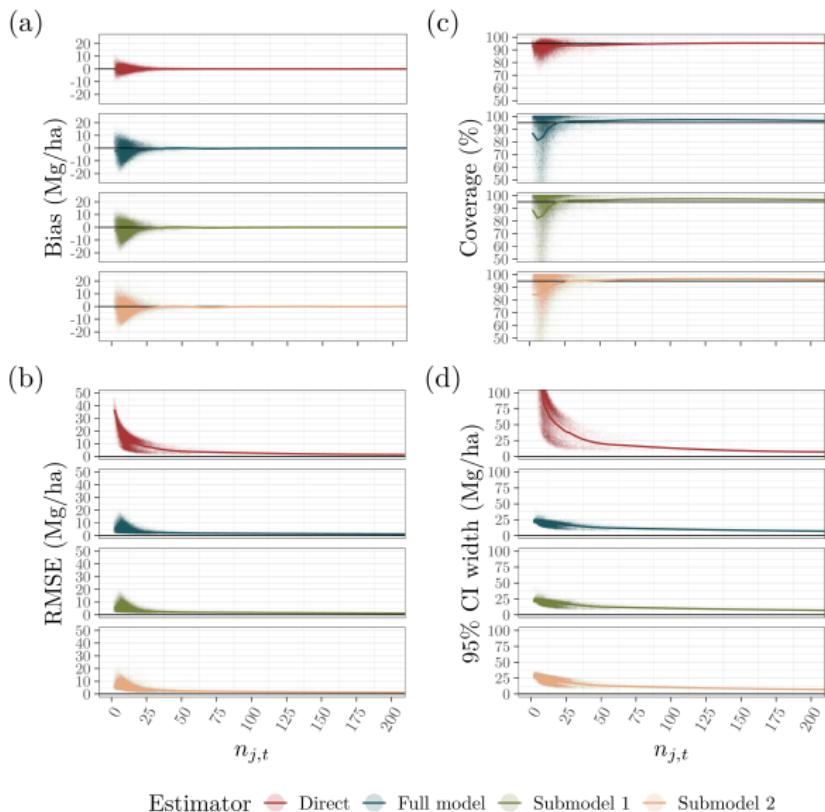


Figure 10: Point values represent averages over  $R$  replicates for a county and year. A loess line is added to indicate trends across sample size ( $n_{j,t}$ ).

## Chapter 2: FIA analysis

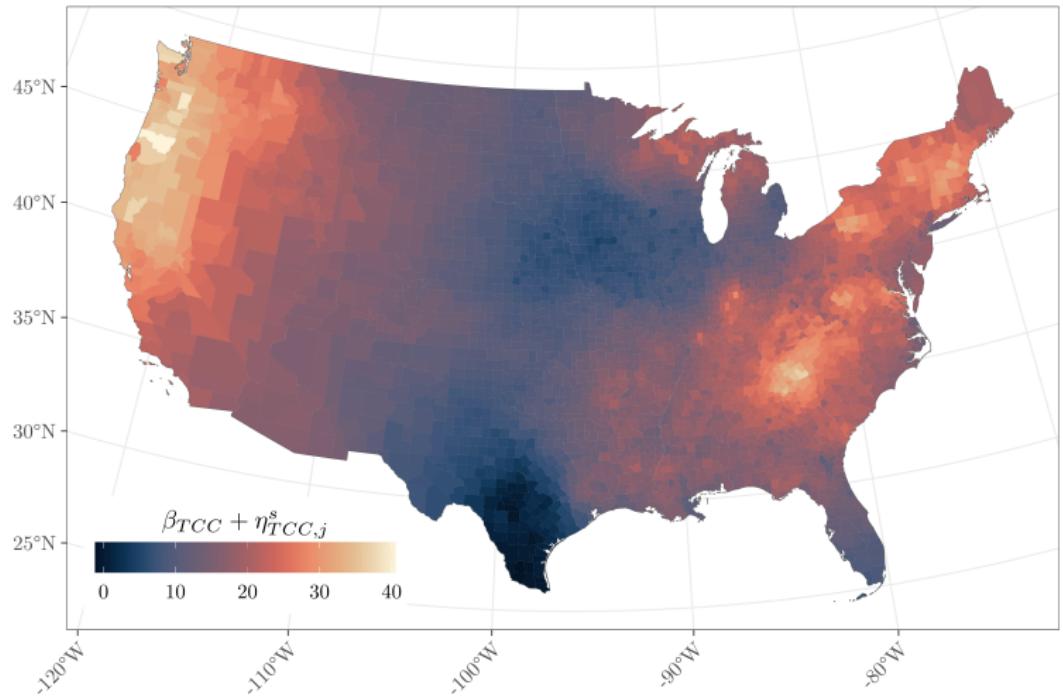


Figure 11: Full model's  $\beta_{TCC} + \eta_{TCC,j}^s$  (i.e., strong evidence of a nonstationary relationship between TCC and carbon density). TCC was scaled and centered.

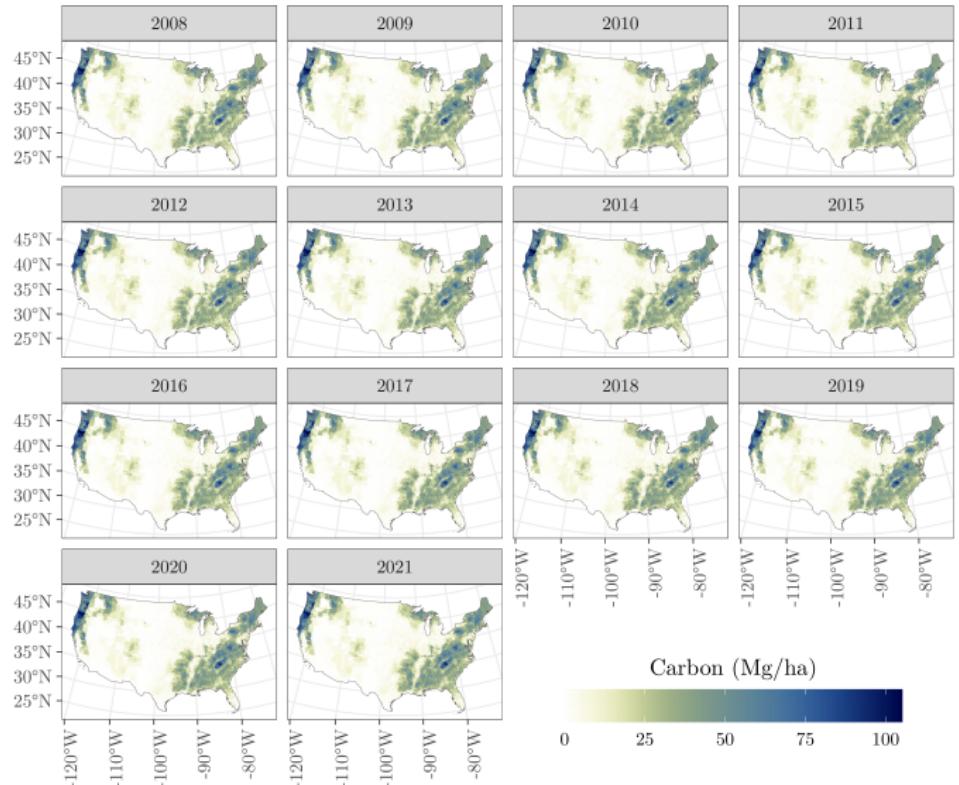


Figure 12: Full model's  $\mu_{j,t}$  posterior mean.

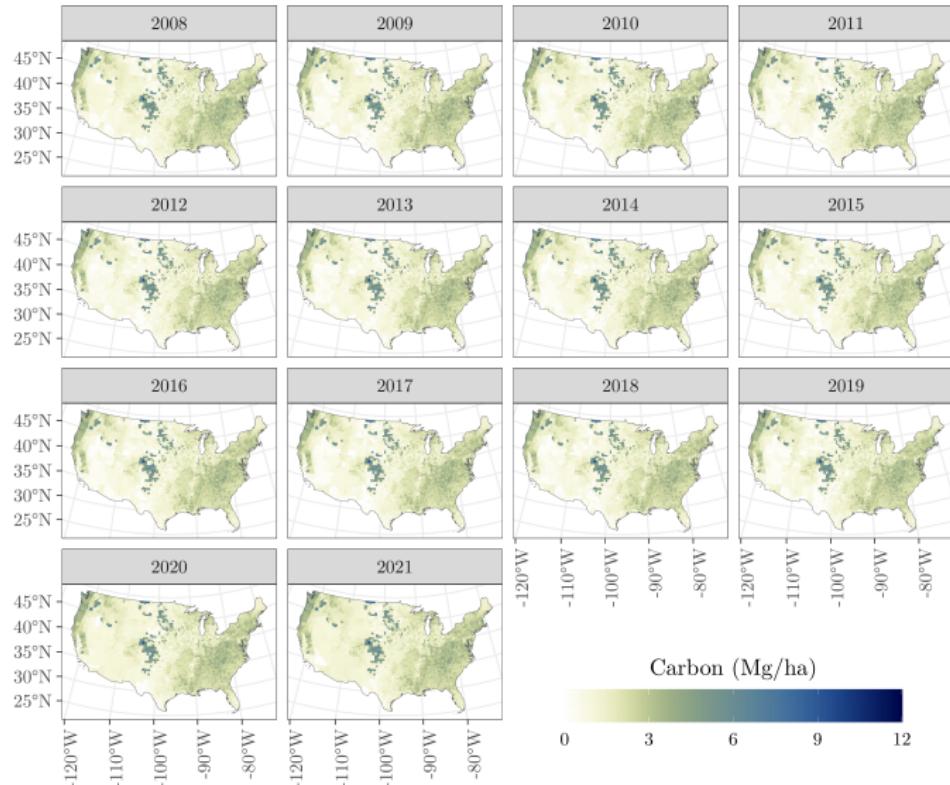


Figure 13: Full model's  $\mu_{j,t}$  posterior standard deviation.

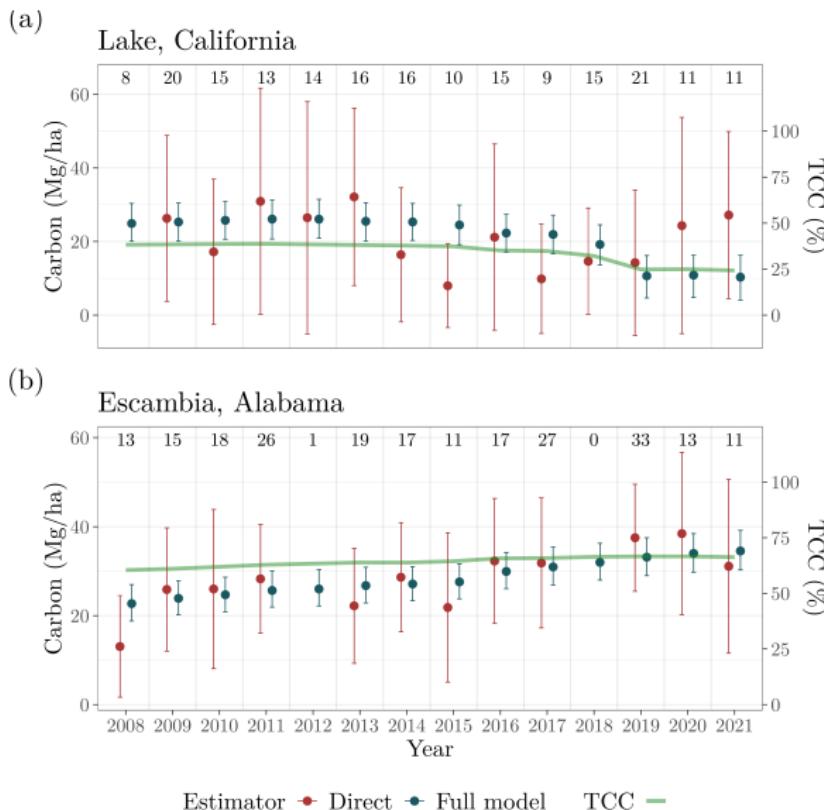


Figure 14: Direct and Full model estimates of  $\hat{\mu}_{j,t}$  and  $\mu_{j,t}$ , respectively, for a few county  $j$ 's and  $t = 1, \dots, T$ , along with associated 95% confidence and credible intervals.

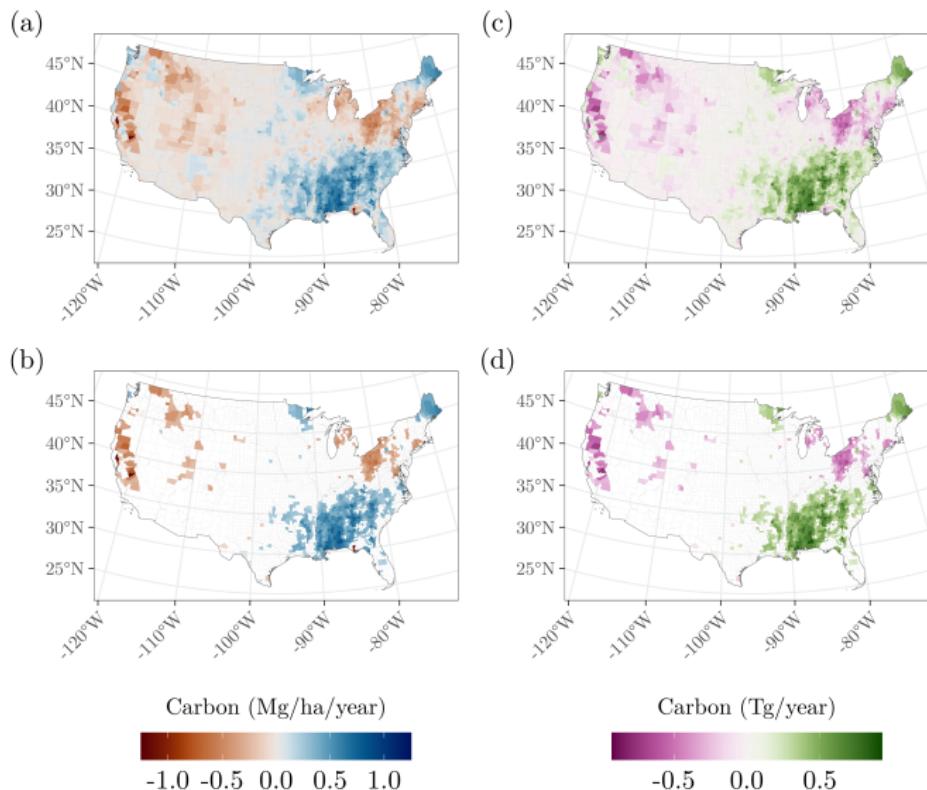
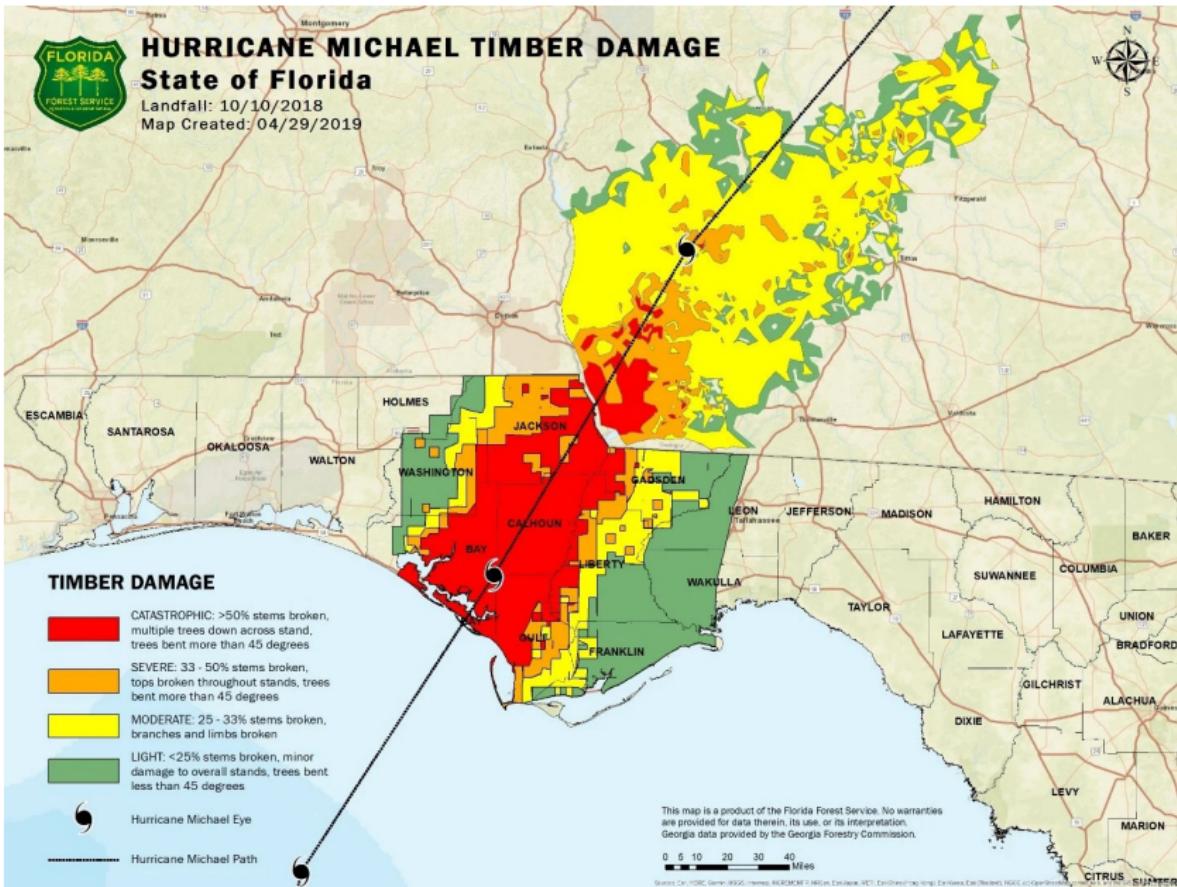
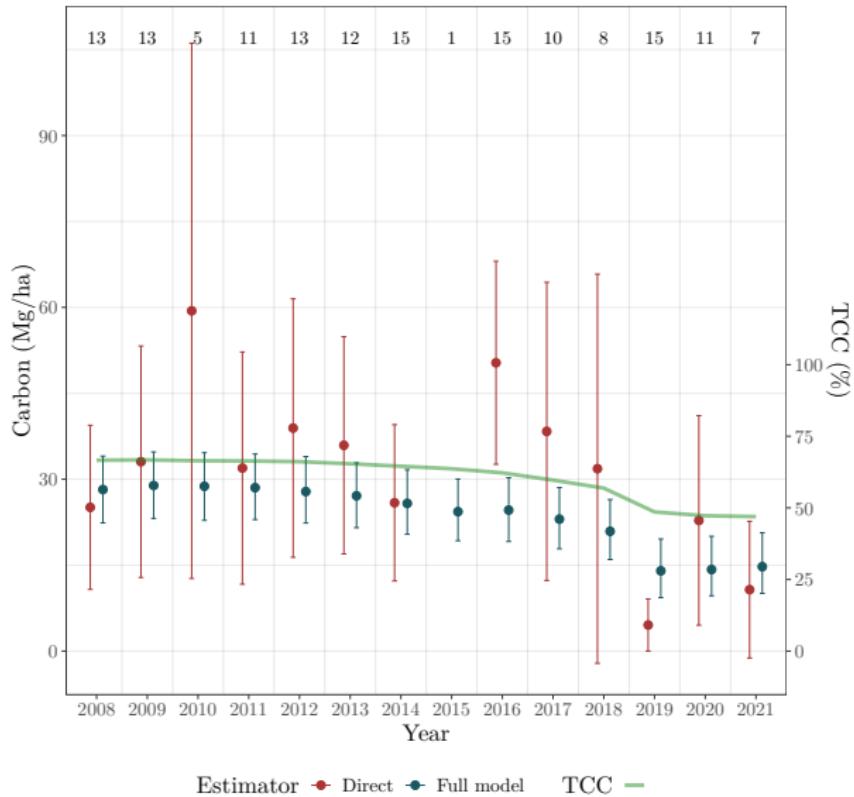


Figure 15: Posterior mean of linear trend and change in carbon over 2008-2021.



# Chapter 2: FIA analysis

Calhoun, Florida



Estimator • Direct • Full model      TCC —

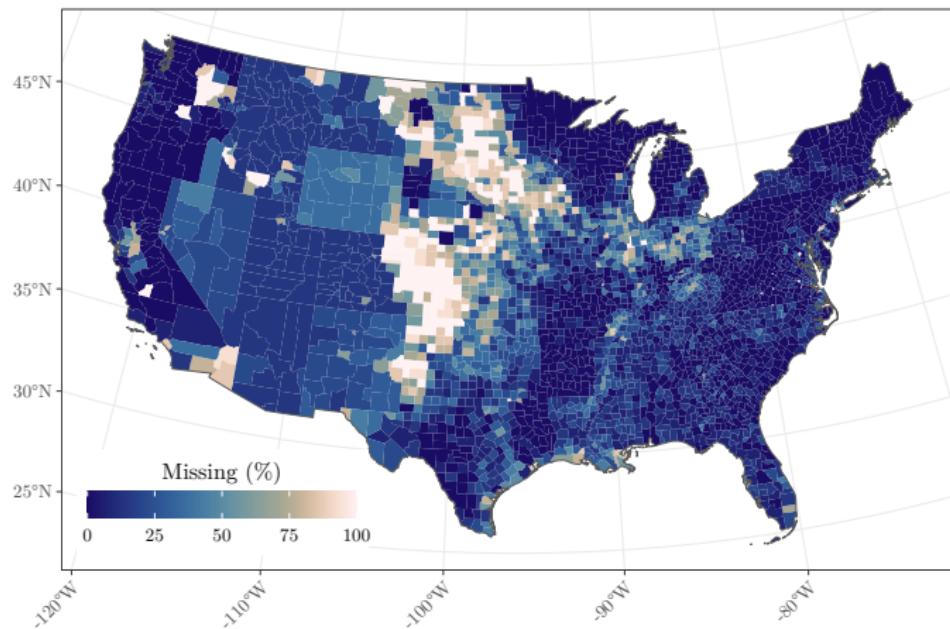
## Chapter 2: FIA analysis



Figure 16: Timber damage from Hurricane Michael. Photo courtesy of Jarek Nowak, Ph.D., Florida Forest Service.

## Chapter 2: Missing Direct Estimates

- When  $n_{j,t} \in \{0, 1\}$ , direct estimates (1) and/or (2) will be missing.
- Direct estimate (2) is also missing when individual measurements in area  $j$  at time  $t$  are identical.



## Chapter 3: Goal

We propose an improved 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.

This model more efficiently leverages sparse NFI data.

## Chapter 3: Model

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

$$y_{i,j,t} = \underbrace{\mathbf{x}_{j,t}^T \boldsymbol{\beta}_t + \tilde{\mathbf{x}}_{j,t}^T \boldsymbol{\eta}_j + u_{j,t}}_{\mu_{j,t}} + \varepsilon_{i,j,t}, \quad i = 1, \dots, n_{j,t}, \quad (13)$$

where

- $\varepsilon_{i,j,t} \stackrel{\text{ind}}{\sim} N(0, \sigma_t^2)$ ,
- $\boldsymbol{\beta}_t$  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.

## Chapter 3: Temporally-varying Regression Coefficients

$\beta_t$  is modeled dynamically as

$$\beta_t = \beta_{t-1} + \xi_t, \text{ with} \quad (14)$$

$$\xi_t \stackrel{\text{ind}}{\sim} MVN(\mathbf{0}, \Sigma_\xi), \quad t = 1, \dots, T, \quad (15)$$

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  $\Sigma_\xi$ .

## Chapter 3: Space-varying Regression Coefficients

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

$$\boldsymbol{\eta}_q^* \sim MVN \left( \mathbf{0}, \tau_{\eta,q}^2 \mathbf{R}(\rho_{\eta,q}) \right), \quad q = 1, \dots, Q, \quad (16)$$

where

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

## Chapter 3: Dynamic Spatio-temporal Intercept

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

$$u_{j,t} = u_{j,t-1} + \omega_{j,t}, \quad (17)$$

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

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

$$\boldsymbol{\omega}_t \sim MVN \left( \mathbf{0}, \tau_{\omega,t}^2 \mathbf{R}(\rho_\omega) \right). \quad (18)$$

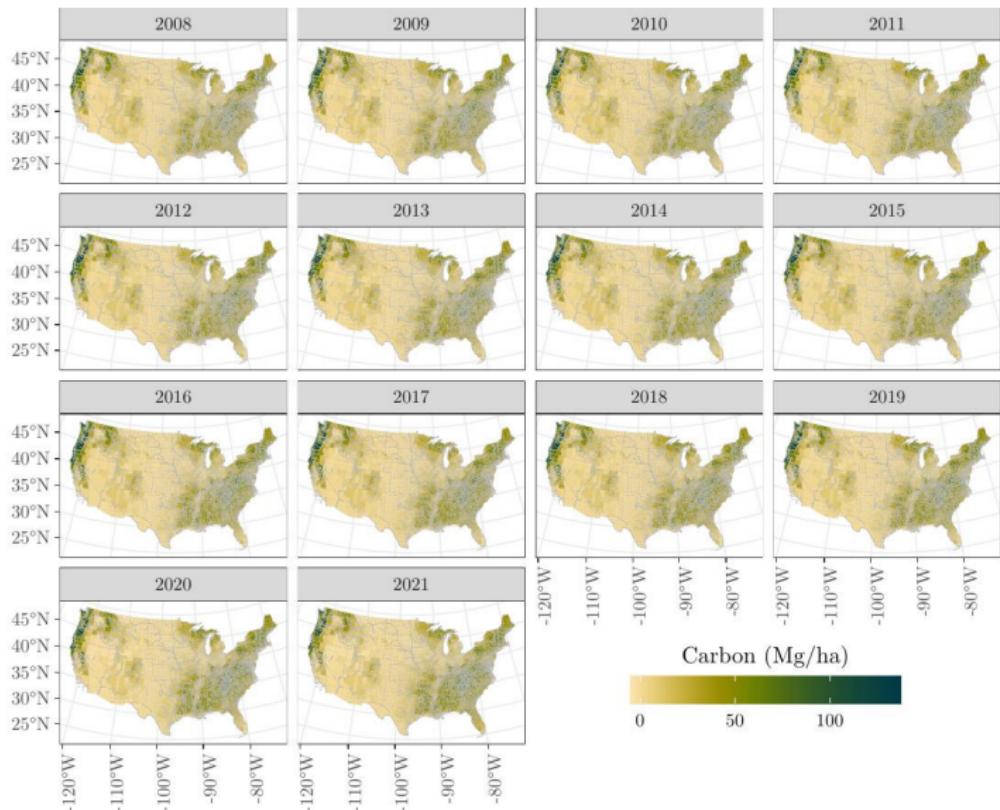


Figure 17: Posterior mean values of live forest carbon density ( $\mu_{j,t}$ ).

## Tuolumne, California

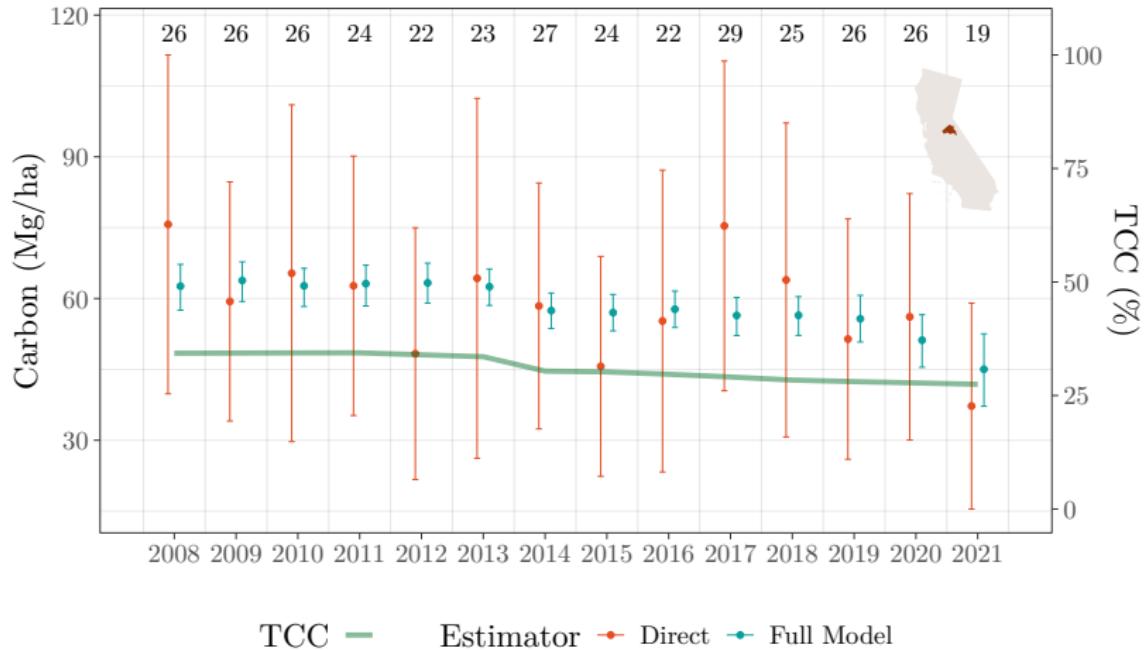


Figure 18: Posterior mean and 95% credible intervals of LFCD ( $\mu_{j,t}$ ) for Tuolumne County, California, compared to direct estimate means ( $\hat{\mu}_{j,t}$ ) and 95% confidence intervals over time. Top row displays sample sizes ( $n_{j,t}$ ).

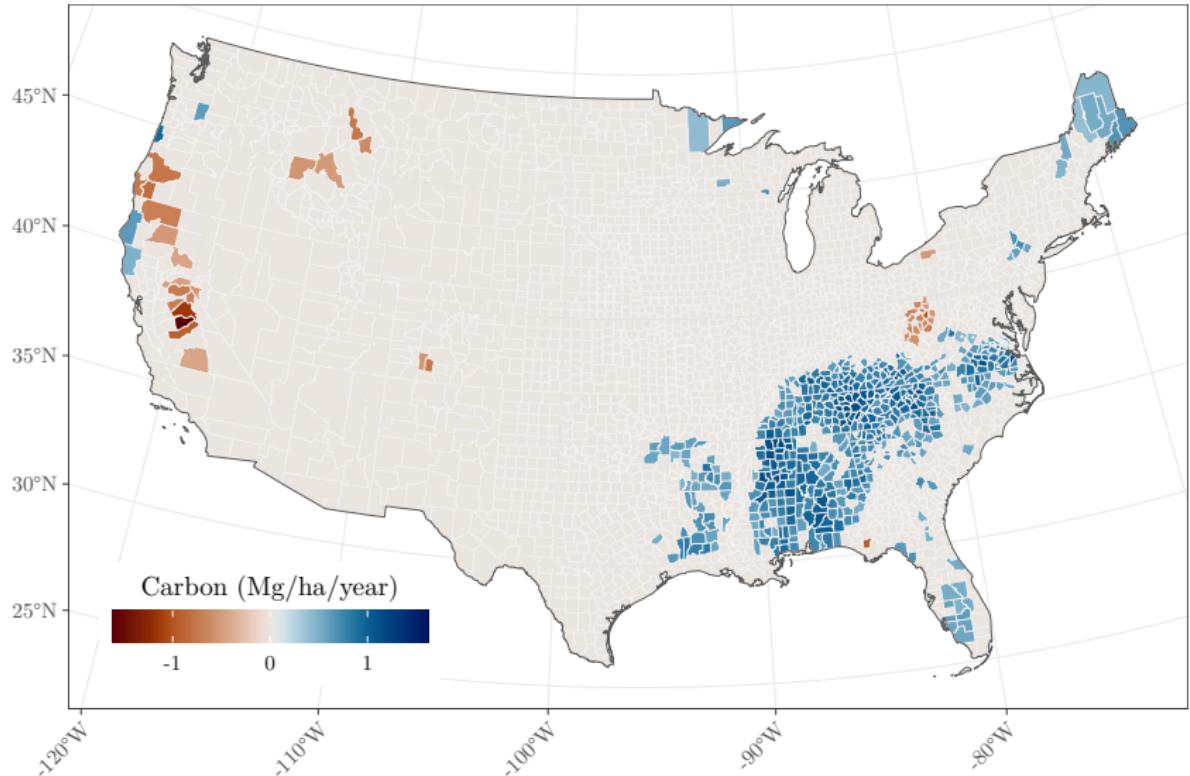


Figure 19: Significant live forest carbon density trends (Mg/ha/year).

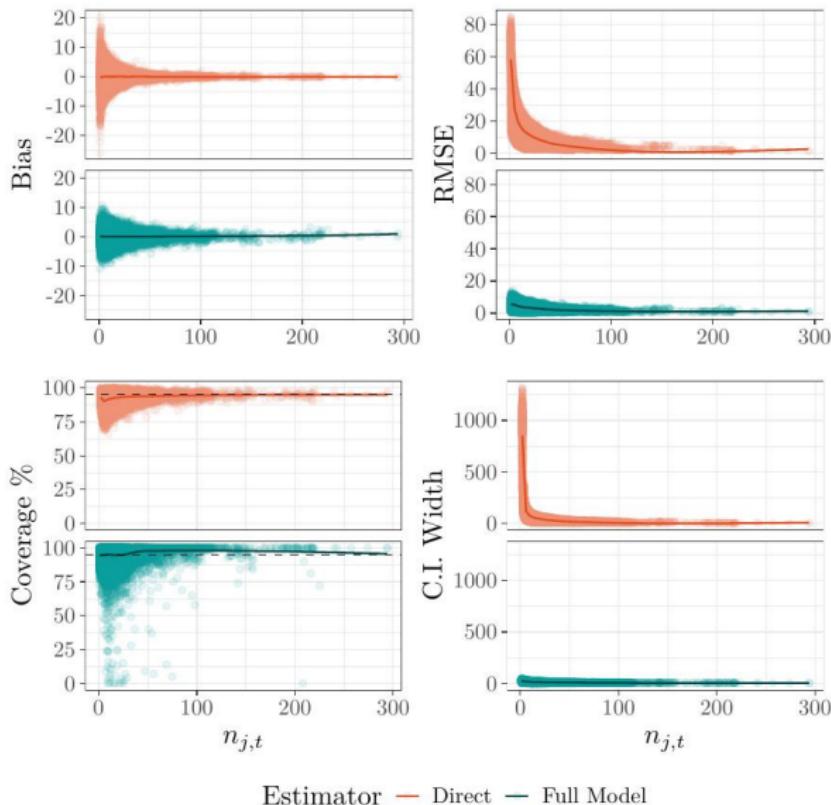


Figure 20: Average measures of bias, root mean square error (RMSE), coverage percentage, and coverage interval widths for the model and direct estimator arranged according to sample size  $n_{j,t}$ .

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

## Chapter 4: Introduction

- Ohio, Pennsylvania, and West Virginia have experienced significant natural gas energy development over the Marcellus and Utica shale formations in recent years.
- Energy production will be the primary driver of land use change in the US through 2040 (Trainor et al., 2016; McDonald et al., 2009)

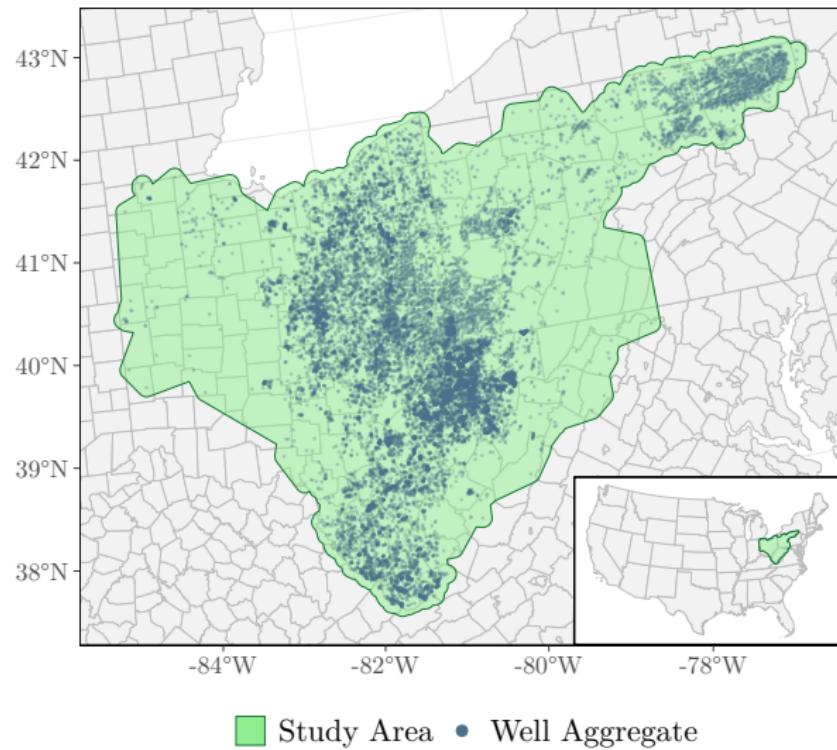


Courtesy of Ohio River Valley Institute.

## Chapter 4: Goal and Data

- Apply a zero-inflated spatio-temporal latent Gaussian SAE model to estimate pixel-level forest carbon changes at natural gas well sites (May and Finley, 2025).
- Implement a dual component covariance structure to allow variation to be modeled at different scales.
- Publicly available permitted natural gas well locations between the years of 2008-2021 were obtained from the Ohio Department of Natural Resources, the Pennsylvania Department of Environmental Protection, and the West Virginia Department of Environmental Protection.
- 99,345 FIA plot measurements to predict the presence of forest between 2008-2021.
- 56,881 forested FIA plot measurements to estimate forest carbon conditional on forest presence.

# Chapter 4: Study Area



## Chapter 4: Model

The proposed model for location  $\mathbf{s}$  at time  $t$  is

$$p(\mathbf{s}, t) \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\mu_p(\mathbf{s}, t)), \quad (19)$$

$$\text{logit}(\mu_p(\mathbf{s}, t)) = \eta_p + x(\mathbf{s}, t)\beta_p, \quad (20)$$

where  $\eta_p$  is an intercept term,  $\beta_p$  is a regression coefficient corresponding to TCC  $x(\mathbf{s}, t)$ . The forest carbon  $y(\mathbf{s}, t)$  is then modeled as

$$y(\mathbf{s}, t) = \eta_y + x(\mathbf{s}, t)\beta_y + w_y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t), \quad (21)$$

where  $\varepsilon(\mathbf{s}, t) \stackrel{\text{iid}}{\sim} \text{Normal}(0, \tau^2)$  is a normally distributed random error term,  $\eta_y$  is an intercept term,  $\beta_y$  is a regression coefficient corresponding to TCC  $x(\mathbf{s}, t)$ , and  $w_y(\mathbf{s}, t)$  is a spatio-temporal random intercept following a mean zero Gaussian process. A final equation for forest carbon is then written as  $y(\mathbf{s}, t) \times p(\mathbf{s}, t)$

## Chapter 4: Disturbance

- To assess instances of predicted forest disturbance at a location  $\mathbf{s}$ , a threshold criteria is implemented based on estimated values of  $\mu_p(\mathbf{s}, t)$ .
- for location  $\mathbf{s}$  and time  $t$ , the probability of disturbance is calculated as  $d(\mathbf{s}, t) = \mu_p(\mathbf{s}, t - 1) \times (1 - \mu_p(\mathbf{s}, t))$ , which represents the probability that location  $\mathbf{s}$  was forested at time  $t - 1$  and not forested at time  $t$ .
- A location  $\mathbf{s}$  is determined to have been disturbed if  $d(\mathbf{s}, t) > 0.5$  for some  $t > 1$ .

## Chapter 4: Opportunity Cost

- To assess opportunity costs of disturbed well aggregate areas, an Autoregressive Moving Average (ARMA) model is fit using TCC values from undisturbed well areas to estimate carbon values for disturbed pixels.
- TCC values from 10,000 undisturbed locations were used to fit the ARMA model, which was then used to predict undisturbed TCC values  $\tilde{x}(\mathbf{s}, t)$  for disturbed locations  $\mathbf{s}$  for all  $t > 1$ .
- These predicted TCC values  $\tilde{x}(\mathbf{s}, t)$  represent the undisturbed scenario for location  $\mathbf{s}$ , which is used to calculate  $\tilde{z}(\mathbf{s}, t)$  using  $\tilde{x}(\mathbf{s}, t)$  in place of  $x(\mathbf{s}, t)$ .

# Chapter 4: Results

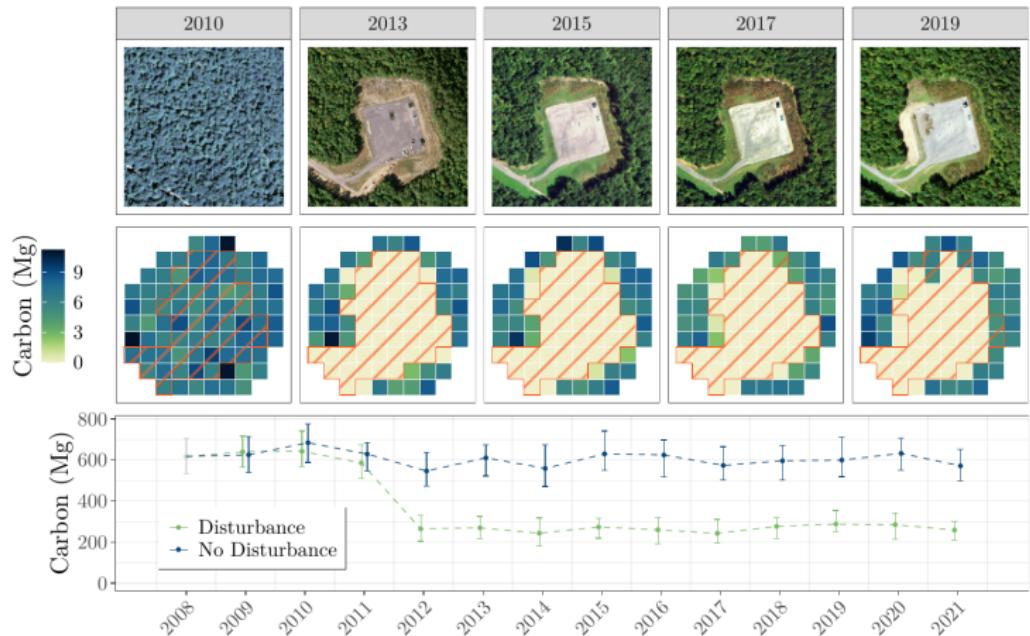


Figure 21: Posterior median pixel-level carbon estimates for a well site in Tioga County, Pennsylvania. Disturbed pixels are denoted with an orange stripe pattern.

# Chapter 4: Results

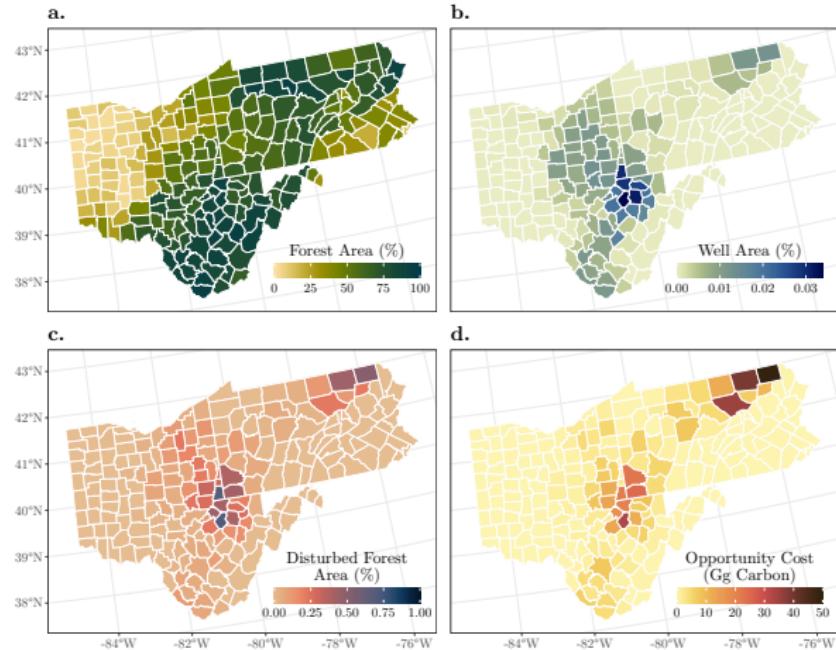


Figure 22: County-level posterior median estimates of percent forest area, percent well area, percent forest area disturbed and total opportunity costs.

# Chapter 4: Results

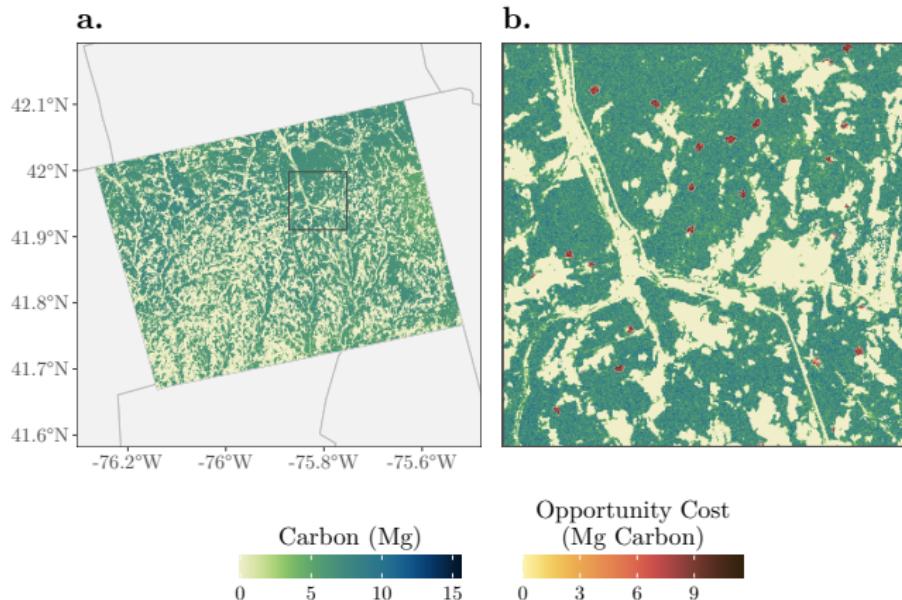


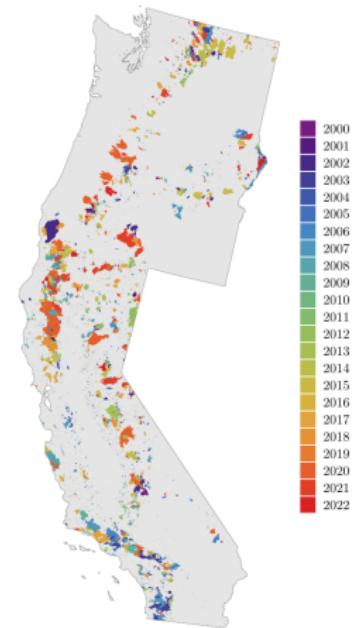
Figure 23: Baseline pixel-level posterior median forest carbon estimates for Susquehanna County, Pennsylvania, with posterior median opportunity costs mapped for disturbed pixel locations.

## Chapter 4: Discussion

- We estimate 10,854 ha of forest land were disturbed.
- Natural gas development resulted in 542,675 Mg ( $\pm$  4,275) of forest carbon loss.
- The opportunity cost associated with these disturbances is estimated to be 575,246 Mg ( $\pm$  30,774).

# Chapter 5: Introduction

- Wildfires are a main driver of forest change and disturbance in Washington, Oregon, and California (Marlon et al., 2012).
- Increased fuels and favorable conditions due to climate change continue to exacerbate the threat of catastrophic wildfires.
- Severe fires have had substantial impacts on forest biomass loss and deforestation.



## Chapter 5: Goal and Data

- Apply the same zero-inflated spatio-temporal latent Gaussian SAE model to estimate pixel-level forest carbon changes associated with West Coast wildfires (May and Finley, 2025).
- 6,750 wildland fires greater than one acre in size occurring between 2000 and 2022 on NFS lands.
- Annual 30 m percent TCC raster images between 1999-2023.
- 90,481 plot-level FIA forest presence measurements are used to fit the forest non-forest model.
- 43,930 forested FIA plot measurements to estimate forest AGB conditional on forest presence.

## Chapter 5: Derived Quantities

- Instances of deforestation are observed when  $d(\mathbf{s}, t) = p(\mathbf{s}, t - 1) - p(\mathbf{s}, t + 1) = 1$ , or in other words, when location  $\mathbf{s}$  is observed to have transitioned from forest to non-forest.
- AGB loss for a deforested location  $\mathbf{s}$  experiencing wildfire at time  $t$  is calculated as  $z(\mathbf{s}, t - 1) - z(\mathbf{s}, t + 1) = z(\mathbf{s}, t - 1)$ , since we necessarily have  $z(\mathbf{s}, t + 1) = 0$  for deforested location  $\mathbf{s}$ .
- Wildfire severity is calculated as the average AGB loss per unit area deforested.

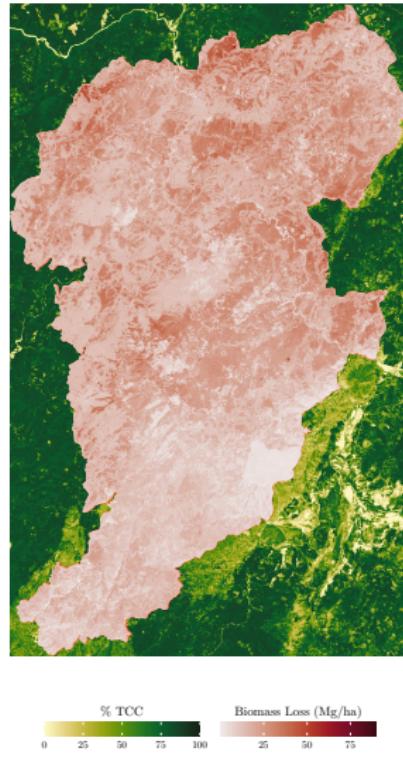


Figure 24: Pixel-level posterior mean total AGB loss for the Biscuit fire, which burned in southwestern Oregon in 2002.



Figure 25: Pixel-level posterior mean AGB loss for deforested areas in the Doe fire, which burned in California in 2020.

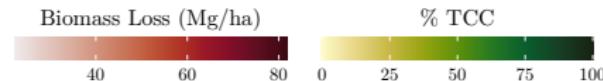
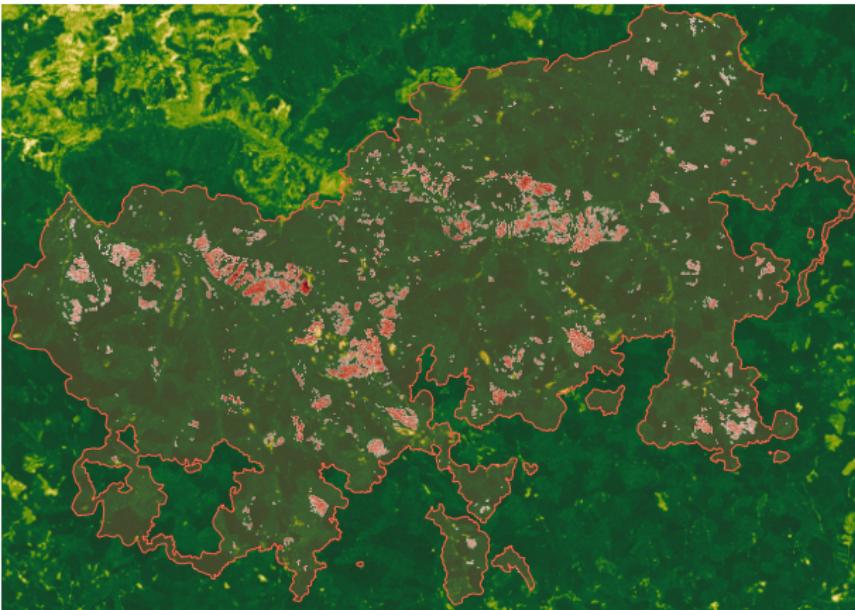
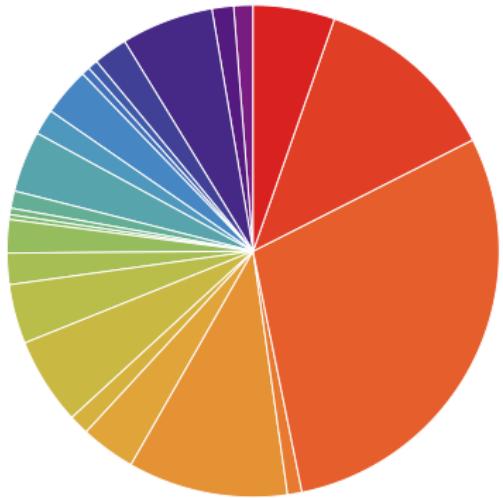


Figure 26: Pixel-level posterior mean AGB loss for deforested areas in the Gales fire, which burned in the Willamette National Forest in Oregon.

Total Biomass Loss



Deforestation



Figure 27: Annual comparison for total AGB loss and deforestation.

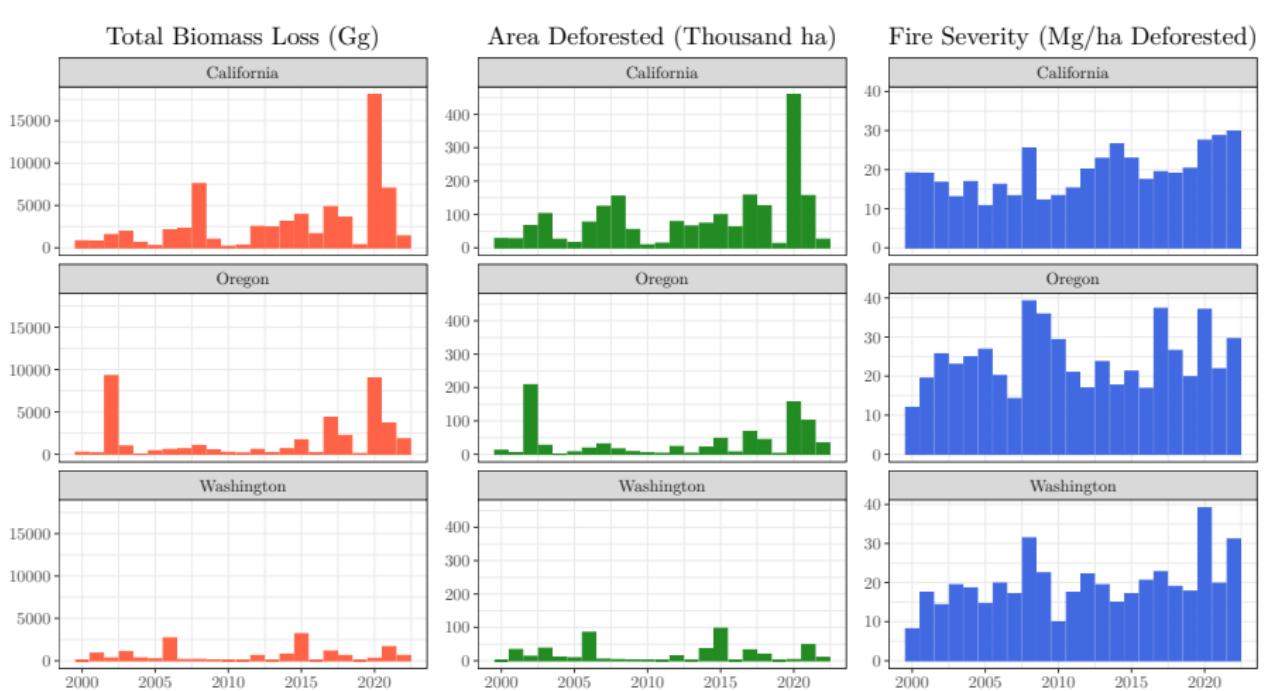


Figure 28: State summaries of posterior median total AGB loss, area deforested, and wildfire severity between 2000-2022 for Washington, Oregon and California.

## Chapter 5: Discussion

- Observed instances of deforestation associated with all of the 6,750 fires between 2000-2022 on NFS lands in Washington, Oregon and California.
- In total, an estimated 121,694,807 Mg (120,897,802, 122,667,215) of AGB were lost in wildfire areas.
- 3,318,165 ha (3,313,818, 3,322,016) of forest lands were estimated to be deforested in wildfire areas.
- Deforestation was associated with 75,602,505 Mg (75,111,471, 76,233,315) of AGB loss, resulting in an average wildfire severity of 22.79 Mg/ha deforested (22.66, 22.95)

# Conclusion: Summary

- In Chapter 2, we extend the traditional Fay-Herriot SAE model to the spatio-temporal setting to estimate mean annual forest carbon for US counties over a 14 year study period.
- In Chapter 3, we develop an improved spatio-temporal SAE model which better leverages available NFI data and bypasses the use of direct estimates.
- In Chapter 4, we apply a zero-inflated spatio-temporal latent Gaussian SAE model to quantify pixel-level impacts of natural gas development on forest carbon.
- In Chapter 5, we apply the same model to more than 20 years of wildfire history on the West Coast to quantify deforestation, AGB loss, and fire severity.

## Conclusion: Next Steps

- Environmental risk assessment and mitigation efforts related to wildfires, air and water quality, and biodiversity loss represent important avenues for future spatio-temporal SAE model development and application.
- Opportunities for collaboration with industry and government partners in risk quantification efforts bring a number of unique perspectives, datasets, and goals which will require additional model development.
- Wildfire risk in the electric utility industry:
  - ▶ Area-level spatio-temporal SAE models such as those developed in Chapters 2 and 3 would allow utilities to better leverage available data related to equipment failures and ignitions to quantify area-level risk across transmission and distribution systems.
  - ▶ Fine-resolution spatio-temporal SAE models such as those applied in Chapters 4 and 5 would provide opportunities to assess risk at the level of individual spans.

# References I

- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*, volume 101. Chapman & Hall/CRC Monographs on Statistical and Applied Probability;.
- Coulston, J. W., Green, P. C., Radtke, P. J., Prisley, S. P., Brooks, E. B., Thomas, V. A., Wynne, R. H., and Burkhart, H. E. (2021). Enhancing the precision of broad-scale forestland removals estimates with small area estimation techniques. *Forestry: An International Journal of Forest Research*, 94(3):427–441.
- Finley, A. O., Andersen, H.-E., Babcock, C., Cook, B. D., Morton, D. C., and Banerjee, S. (2024). Models to support forest inventory and small area estimation using sparsely sampled lidar: A case study involving g-light lidar in tanana, alaska. *Journal of Agricultural, Biological and Environmental Statistics*.
- Hou, Z., Domke, G. M., Russell, M. B., Coulston, J. W., Nelson, M. D., Xu, Q., and McRoberts, R. E. (2021). Updating annual state-and county-level forest inventory estimates with data assimilation and FIA data. *Forest Ecology and Management*, 483:118777.

## References II

- Housman, I., Schleeweis, K., Heyer, J., Ruefenacht, B., Bender, S., Megown, K., Goetz, W., and Bogle, S. (2023). National land cover database tree canopy cover methods v2021.4. GTAC-10268-RPT1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Geospatial Technology and Applications Center.
- Lister, A. J., Andersen, H., Frescino, T., Gatziolis, D., Healey, S., Heath, L. S., Liknes, G. C., McRoberts, R. E., Moisen, G. G., Nelson, M., et al. (2020). Use of Remote Sensing Data to Improve the Efficiency of National Forest Inventories: A Case Study from the United States National Forest Inventory. *Forests*, 11(12):1364.
- Marlon, J. R., Bartlein, P. J., Gavin, D. G., Long, C. J., Anderson, R. S., Briles, C. E., Brown, K. J., Colombaroli, D., Hallett, D. J., Power, M. J., Scharf, E. A., and Walsh, M. K. (2012). Long-term perspective on wildfires in the western usa. *Proceedings of the National Academy of Sciences*, 109(9):E535–E543.
- May, P. B. and Finley, A. O. (2025). Spatial-temporal prediction of forest attributes using latent gaussian models and inventory data.

## References III

- McDonald, R. I., Fargione, J., Kiesecker, J., Miller, W. M., and Powell, J. (2009). Energy sprawl or energy efficiency: Climate policy impacts on natural habitat for the united states of america. *PLOS ONE*, 4(8):1–11.
- Mcroberts, R., Wendt, D., and Liknes, G. (2005). Stratified estimation of forest inventory variables using spatially summarized stratifications. *Silva Fennica*, 39.
- Schroeder, T. A., Healey, S. P., Moisen, G. G., Frescino, T. S., Cohen, W. B., Huang, C., Kennedy, R. E., and Yang, Z. (2014). Improving estimates of forest disturbance by combining observations from Landsat time series with US Forest Service Forest Inventory and Analysis data. *Remote Sensing of Environment*, 154:61–73.
- Shannon, E. S., Finley, A. O., Domke, G. M., May, P. B., Andersen, H.-E., Gaines III, G. C., and Banerjee, S. (2024). Toward spatio-temporal models to support national-scale forest carbon monitoring and reporting. *Environmental Research Letters*, 20(1):014052.
- Trainor, A. M., McDonald, R. I., and Fargione, J. (2016). Energy sprawl is the largest driver of land use change in united states. *PLOS ONE*, 11(9):1–16.