# A spatial small area model for predicting average forest biomass using fine resolution geospatial data

Grayson W. White[1,2] & Andrew O. Finley[1,2]

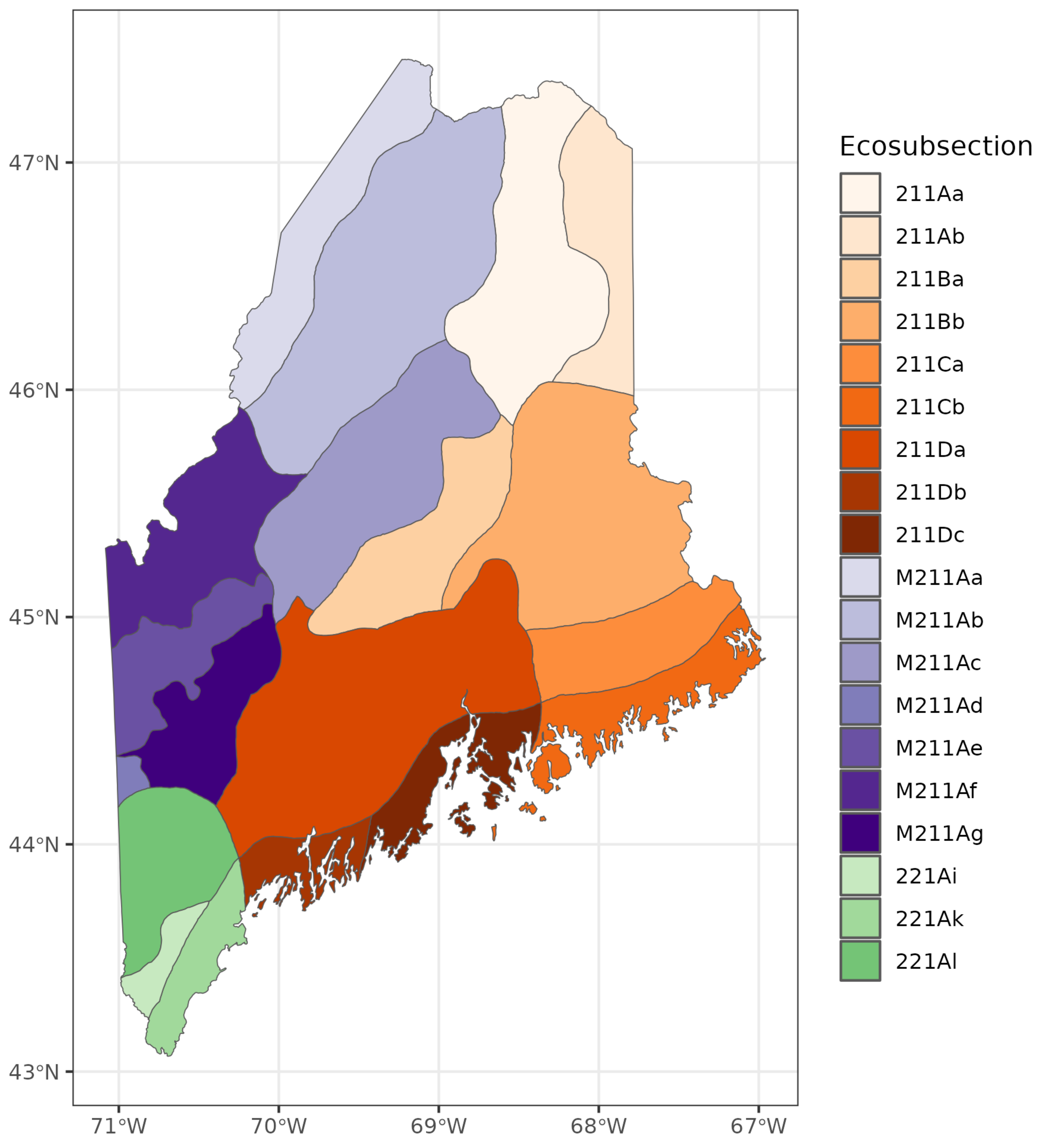
[1] Michigan State University, Department of Forestry
[2] Michigan State University, Department of Statistics & Probability

# What's the big idea?

- Combine ground truth biomass data and high resolution canopy height auxiliary data to fit unit-level models to estimate means of biomass within small areas of interest.

- Borrow strength spatially to account for spatial covariance in data.

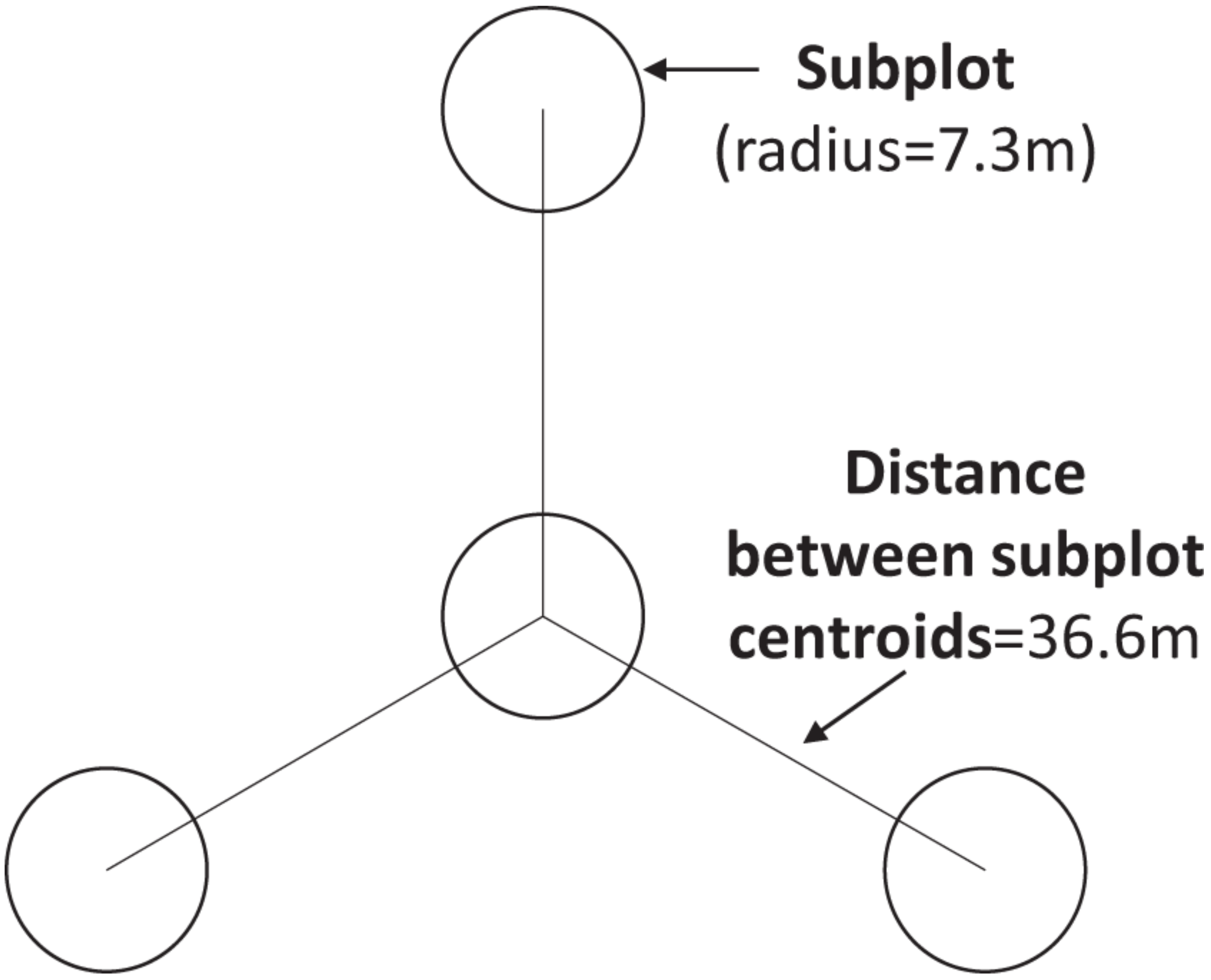
- Compare candidate spatial model to non-spatial model based estimators such as: **a)** a linear regression estimator, and **b)** a linear regression with random effects for "ecosubsection" to group the data ecologically.



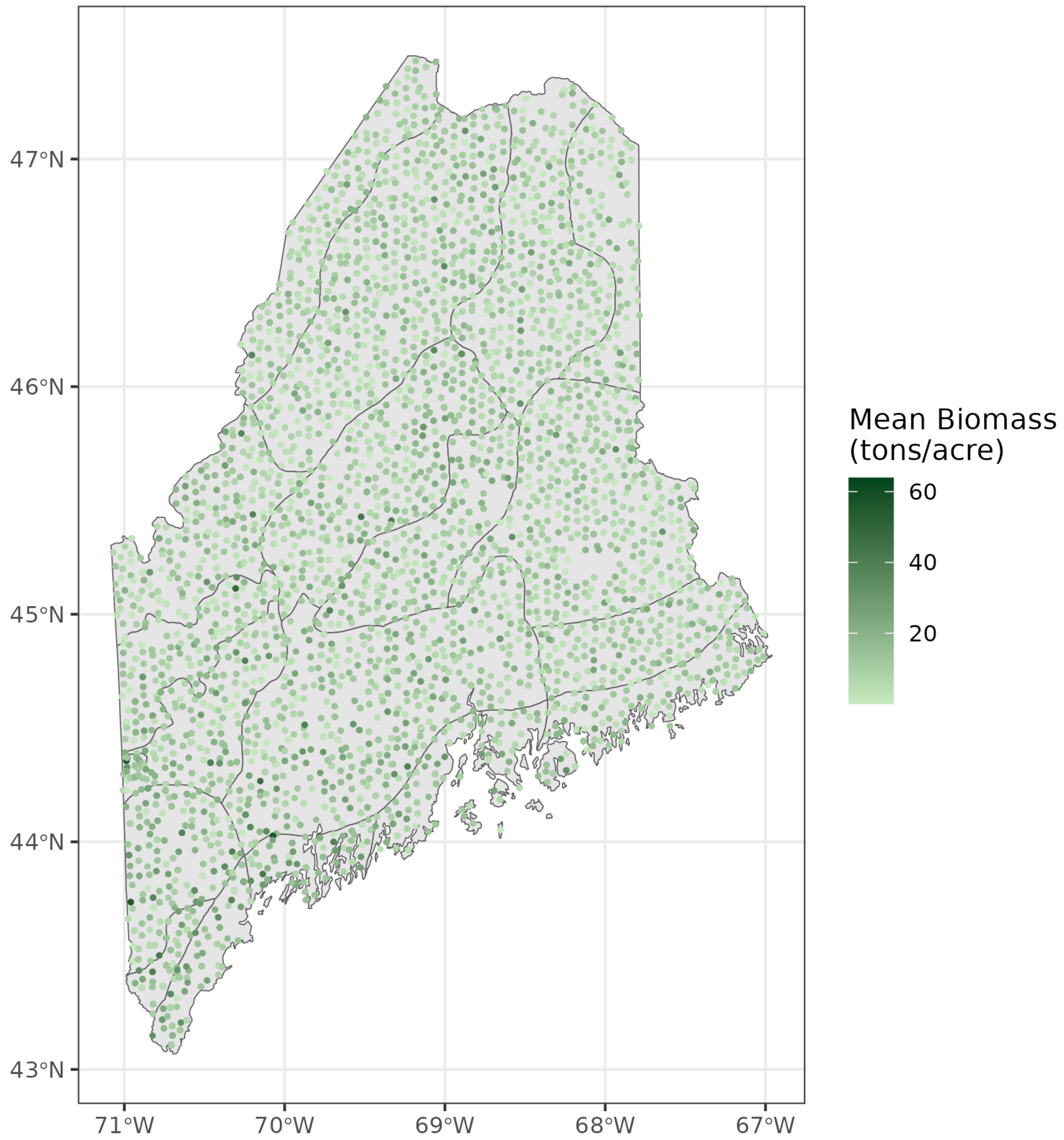
# Data

# Data: USDA Forest Service: FIA Program

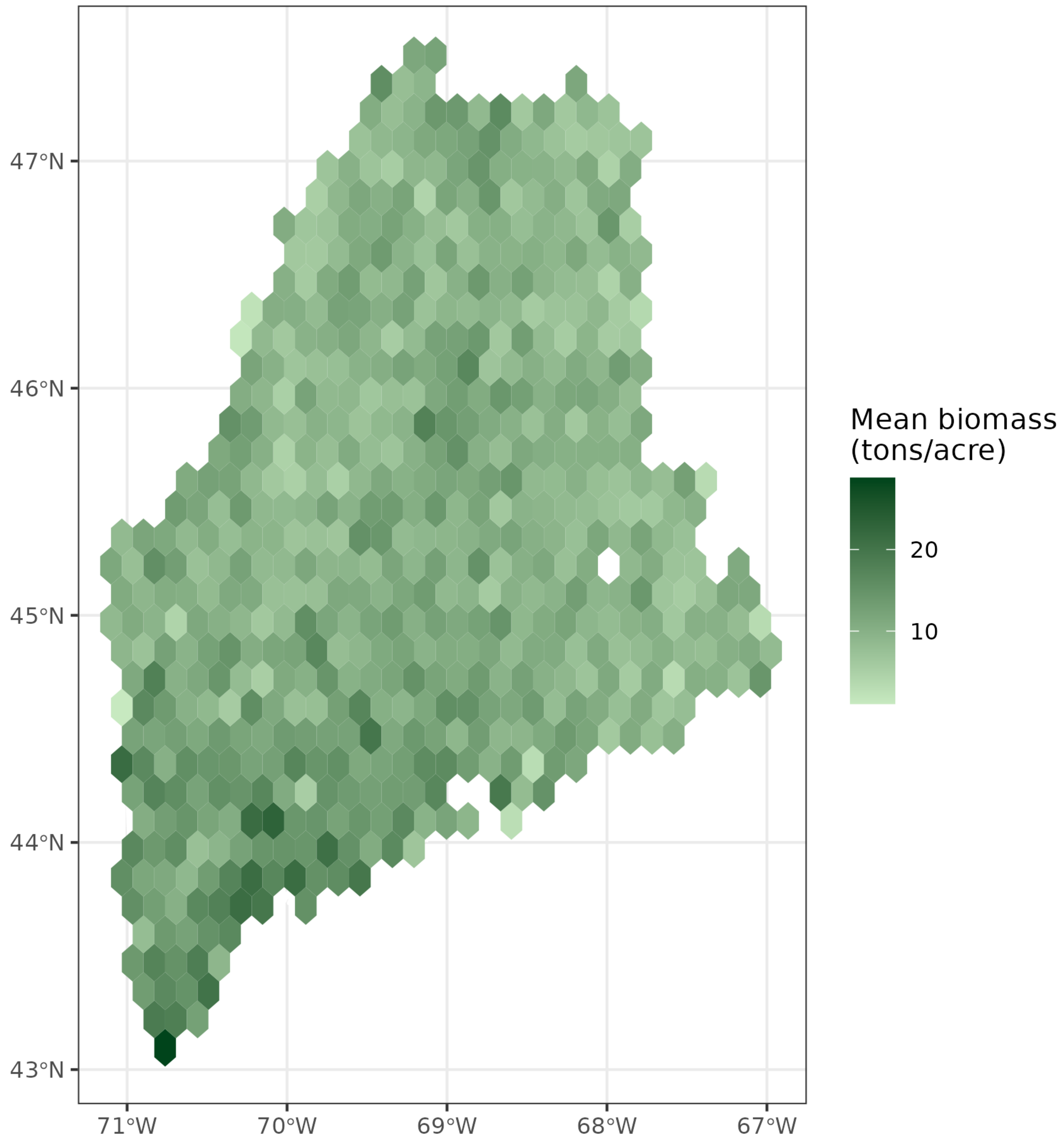
- FIA plots contain four 7.3m radius *subplots*, which are spaced 36.6m apart.
- FIA plots are spaced across the US, one plot per 6,000 acres, through a quasi-systematic sampling design.
- In the Eastern US, plots are measured once every 5 years.
- Our models will be fit at the *subplot level*.



# FIA Plots in Maine, USA



# FIA Plots in Maine, USA



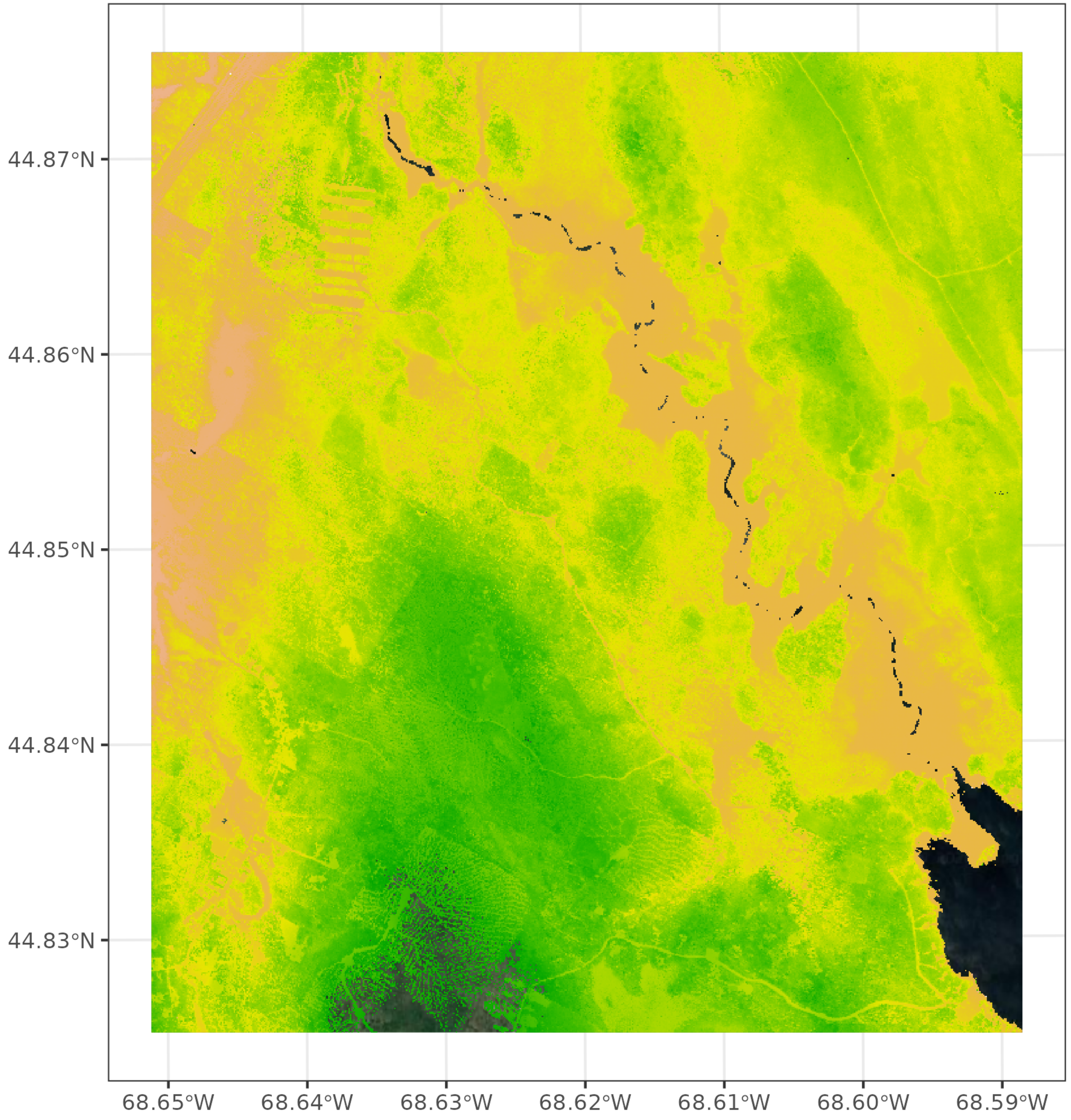
# Auxiliary Data: NAIP Satellite Imagery

- 1-meter resolution canopy height model (CHM) derived from a digital surface model (DSM) and digital terrain model (DTM).

- Collected in 2021.

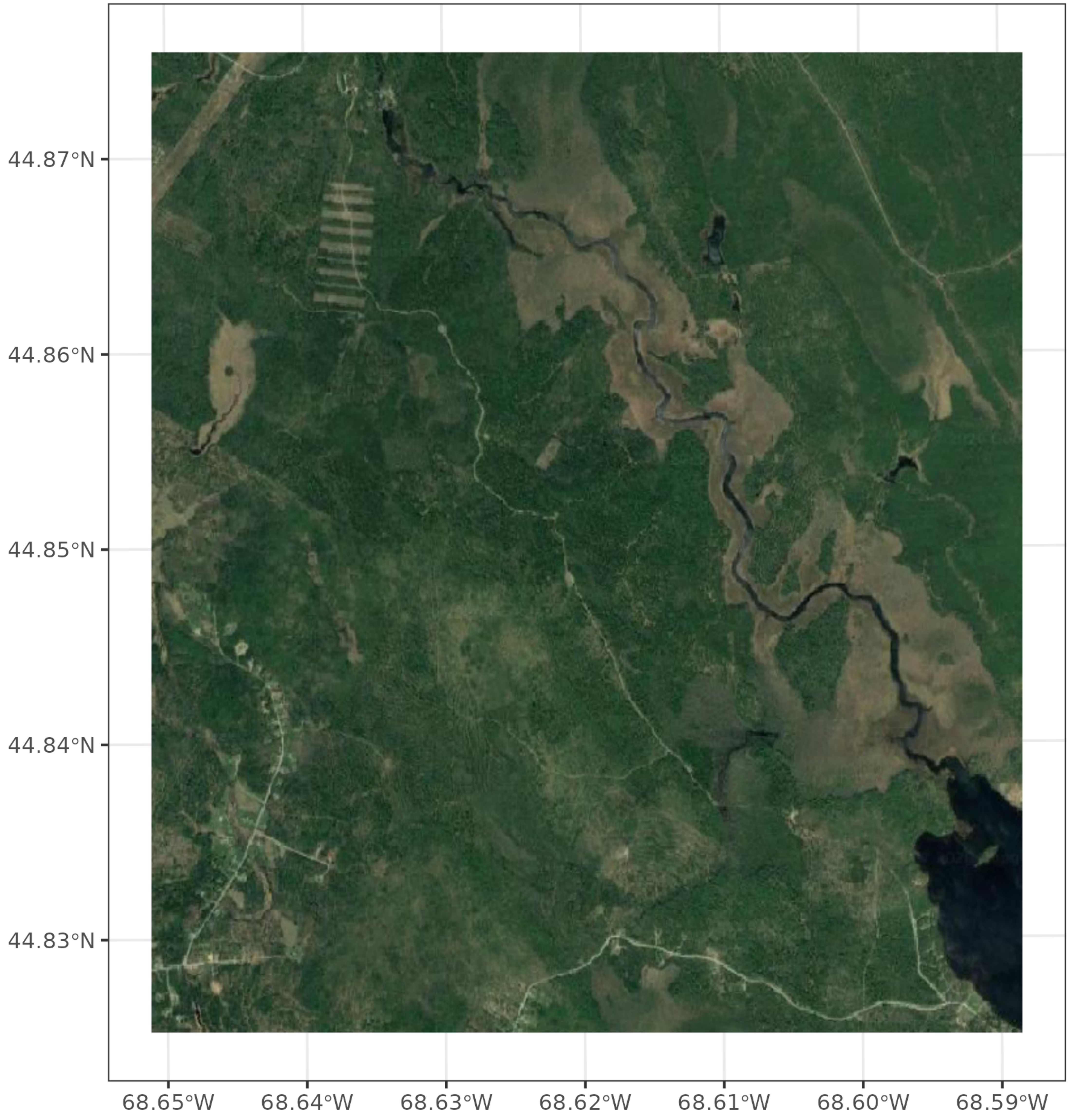
- Approximately 168 pixels per subplot.

- Derived predictors: mean canopy height, and 98th- and 2nd-quantile canopy height



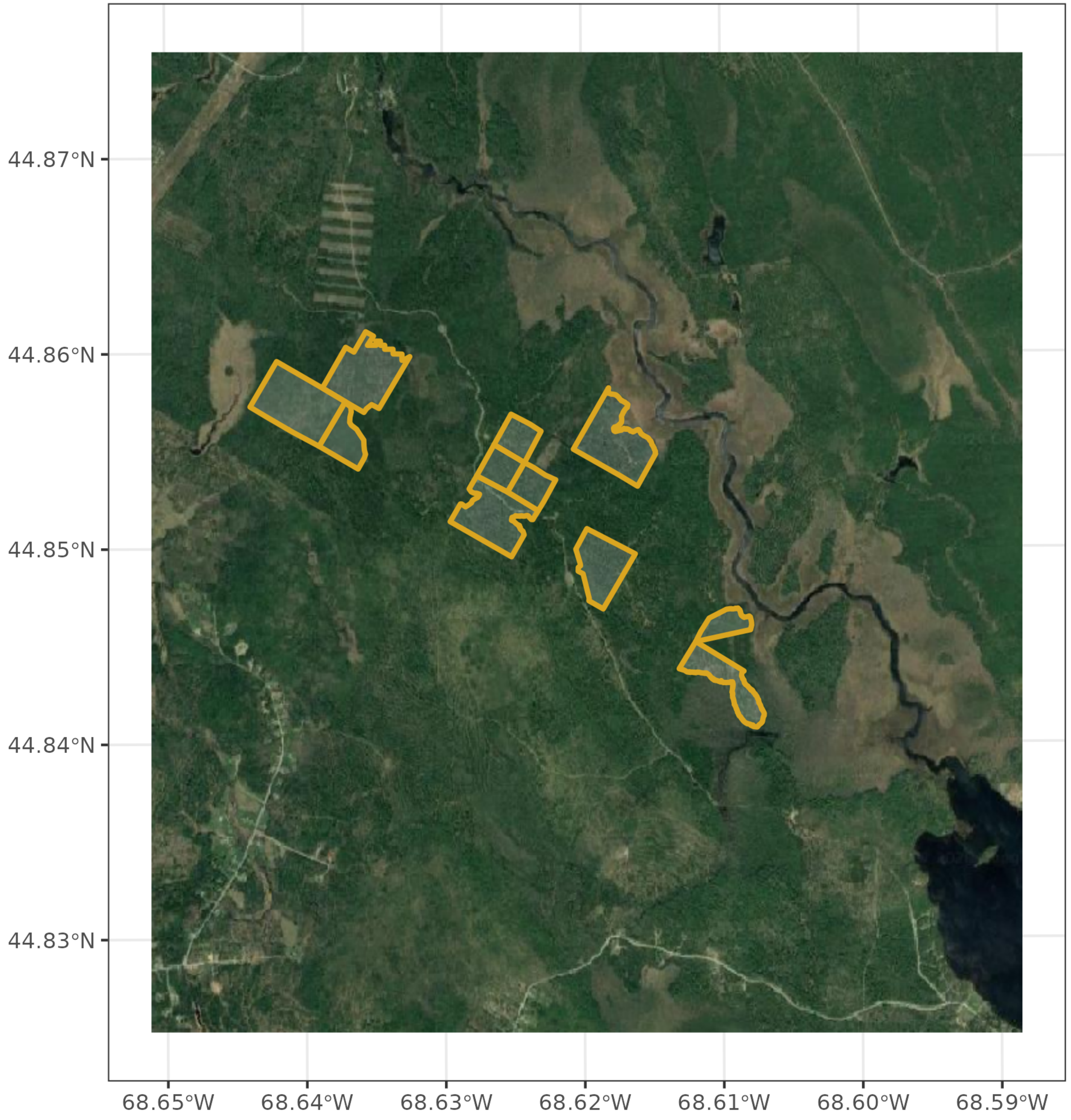
# The Penobscot Experimental Forest (PEF)

- Experimental forest near Bangor, ME, jointly managed by the USDA Forest Service and University of Maine.
- Approximately 4000 acres in total.



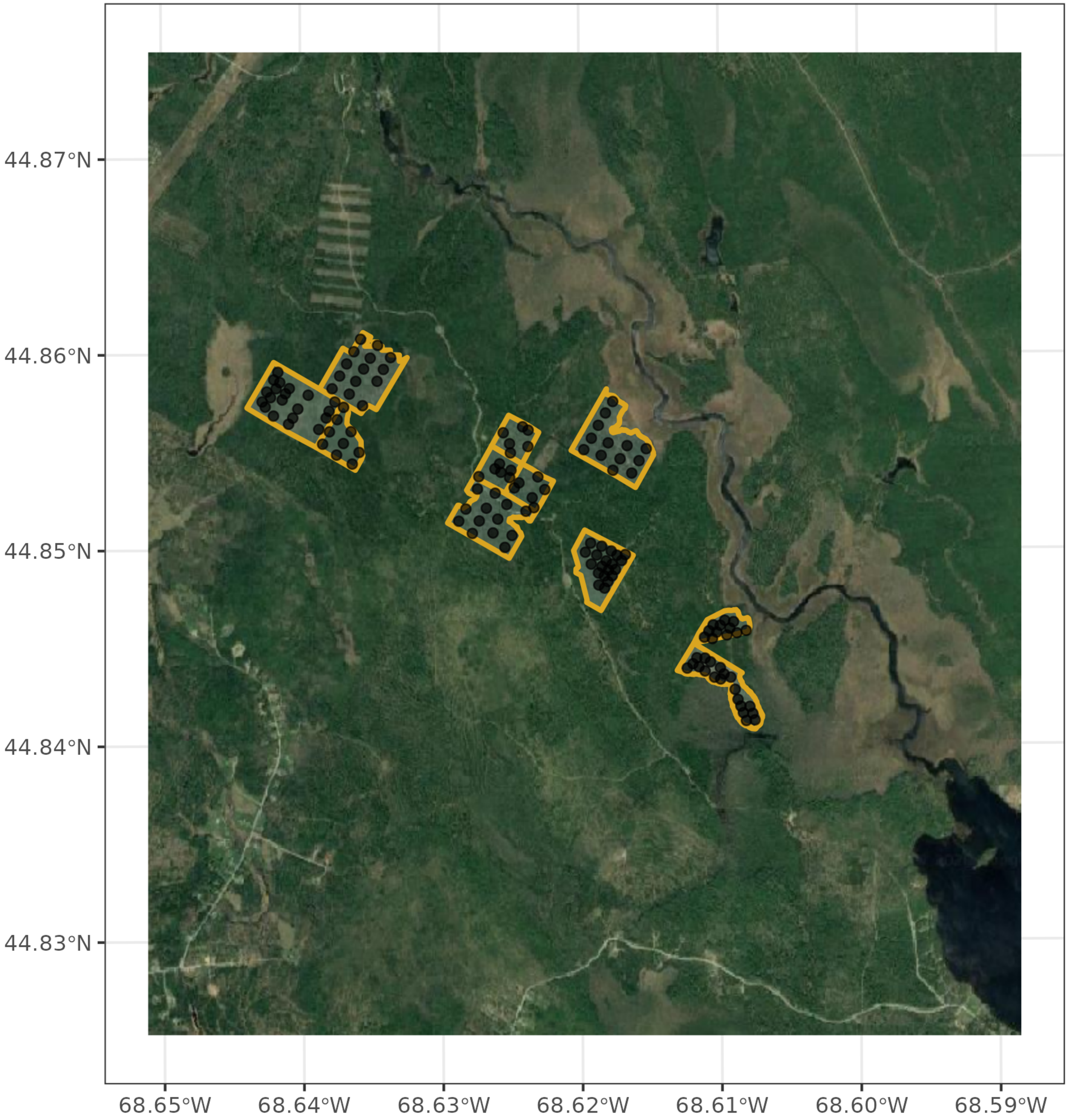
# The Penobscot Experimental Forest (PEF)

- Experimental forest near Bangor, ME, jointly managed by the USDA Forest Service and University of Maine.
- Approximately 4000 acres in total.
- Contains management units approximately 20 acres in size, 11 of which make up our small area of interest.



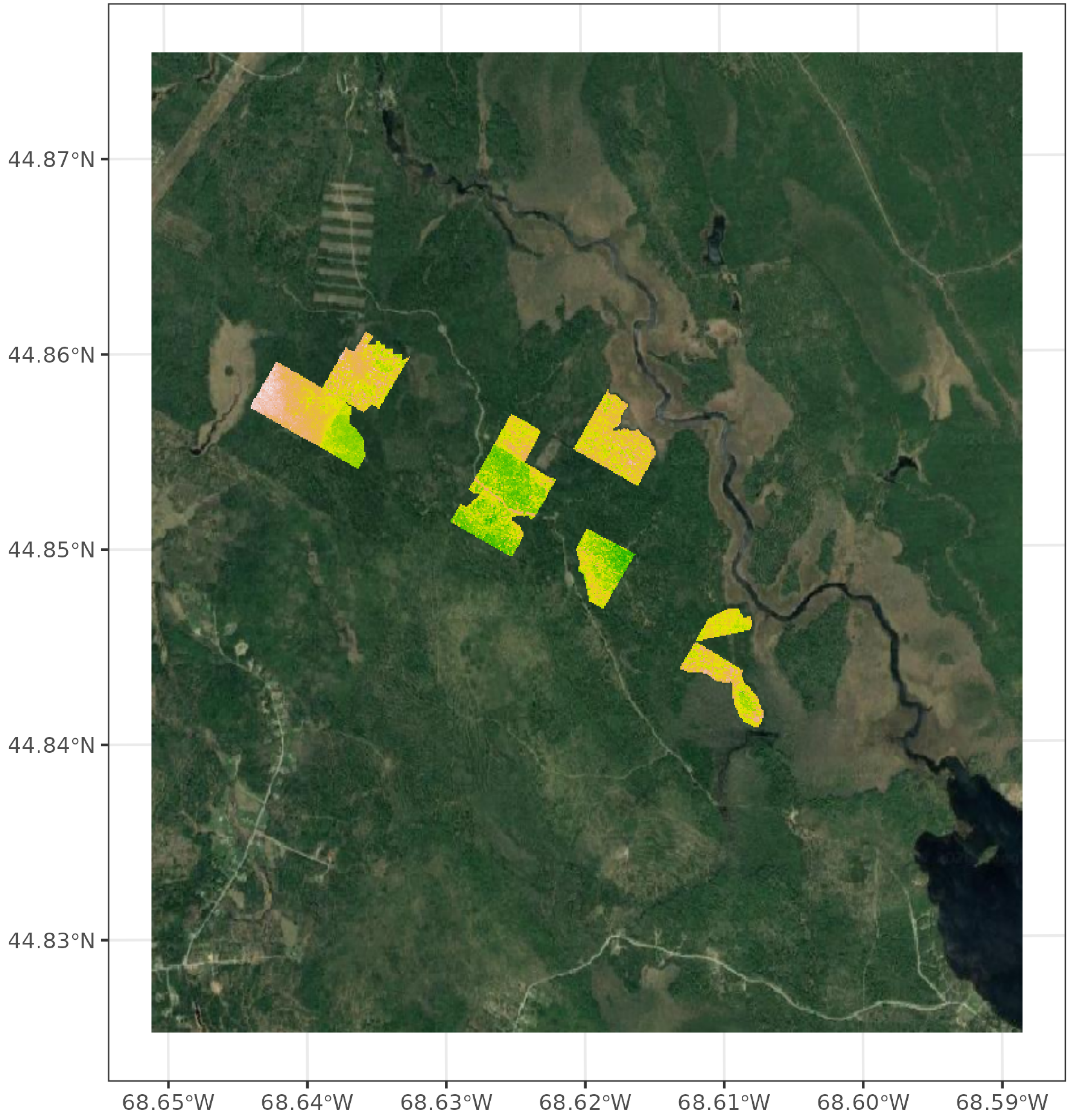
# The Penobscot Experimental Forest (PEF)

- Experimental forest near Bangor, ME, jointly managed by the USDA Forest Service and University of Maine.
- Approximately 4000 acres in total.
- Contains management units approximately 20 acres in size, 11 of which make up our small area of interest.
- A dense inventory of plots are collected in MUs, which we will use as our sense of truth.



# The Penobscot Experimental Forest (PEF)

- Experimental forest near Bangor, ME, jointly managed by the USDA Forest Service and University of Maine.
- Approximately 4000 acres in total.
- Contains management units approximately 20 acres in size, 11 of which make up our small area of interest.
- A dense inventory of plots are collected in MUs, which we will use as our sense of truth.



# Methods

# Methods: *Linear Model*

For the linear regression model, we assume a model for $y_i$, the biomass at FIA subplot $i$ of the following form:

$$\sqrt{y_i} \sim \mathrm{N}(x_i^T \beta,\ \sigma_e^2)$$

with priors

$$f(\beta) \propto 1,\ f(\sigma_e^2) \sim \mathrm{IGamma}$$

# Methods: *Linear Mixed Model*

For the linear mixed model, we assume a model for $y_{ij}$, the biomass at FIA subplot $i$ in ecosubsection $j$ of the following form:

$$\sqrt{y_{ij}} \sim \mathrm{N}(x_i^T \beta + \nu_j, \ \sigma_e^2),$$

$$\nu_j \sim \mathrm{N}(0, \ \sigma_\nu^2)$$

with priors

$$f(\beta) \propto 1, \ f(\sigma_e^2), \ f(\sigma_\nu^2) \sim \mathrm{IGamma}$$

# Methods: *Spatial Linear Mixed Model*

For the spatial linear mixed model, we assume a model for $y_i$, the biomass at generic spatial location $\mathrm{s}_i$ of the following form:

$$\sqrt{y}(\mathrm{s}_i) \sim \mathrm{N}\left(x(\mathrm{s}_i)^T\beta + w(\mathrm{s}_i),\ \tau^2\right),$$

$$w(\mathrm{s}_i) \sim \mathrm{GP}\left(0,\ \tilde{\mathrm{C}}(.,. \mid \theta)\right)$$

where $w(\mathrm{s}_i)$ is estimated through a nearest neighbor Gaussian process with the `spNNGP` R package and $\tilde{\mathrm{C}}$ is a covariance function leading to a sparse covariance matrix, constructed from C, an exponential covariance function:

$$\mathrm{C}(\mathrm{s}_1, \mathrm{s}_2 \mid \theta) = \sigma^2 \exp(-\phi||\mathrm{s}_1 - \mathrm{s}_2||)$$

with priors:

$$\phi \sim \mathrm{Unif}(3/100,\ 3/5),\ \sigma^2 \sim \mathrm{IG}(2,\ 0.5),\ \tau^2 \sim \mathrm{IG}(2,\ 0.2)$$

# Methods: *Small Area Estimation*

For each model, we aggregate posterior samples after burn-in and thinning to produce a posterior distribution for $\mu$, the mean biomass for a given small area of interest.

$$\begin{array}{cc} & \longleftarrow \mathbf{M}\ \text{MCMC samples} \longrightarrow \\ \mathbf{P}\ \text{Pixels} & \begin{bmatrix} \sqrt{\hat{y}_{1,1}} & \cdots & \sqrt{\hat{y}_{1,\mathrm{M}}} \\ \vdots & \ddots & \vdots \\ \sqrt{\hat{y}_{\mathrm{P},1}} & \cdots & \sqrt{\hat{y}_{\mathrm{P},\mathrm{M}}} \end{bmatrix} \end{array}$$

# Methods: *Small Area Estimation*

For each model, we aggregate posterior samples after burn-in and thinning to produce a posterior distribution for $\mu$, the mean biomass for a given small area of interest.

$$
\begin{array}{cc}
 & \longleftarrow \mathbf{M}\ \text{MCMC samples} \longrightarrow \\
\mathbf{P}\ \text{Pixels} & \begin{bmatrix} \sqrt{\hat{y}_{1,1}}^{\,2} & \cdots & \sqrt{\hat{y}_{1,\mathrm{M}}}^{\,2} \\ \vdots & \ddots & \vdots \\ \sqrt{\hat{y}_{\mathrm{P},1}}^{\,2} & \cdots & \sqrt{\hat{y}_{\mathrm{P},\mathrm{M}}}^{\,2} \end{bmatrix}
\end{array}
$$

# Methods: *Small Area Estimation*

For each model, we aggregate posterior samples after burn-in and thinning to produce a posterior distribution for $\mu$, the mean biomass for a given small area of interest.

$$\begin{array}{cc} & \longleftarrow \ \textbf{M}\ \text{MCMC samples} \ \longrightarrow \\ \updownarrow\ \textbf{P}\ \text{Pixels} & \begin{bmatrix} \hat{y}_{1,1} & \cdots & \hat{y}_{1,\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \hat{y}_{\mathrm{P},1} & \cdots & \hat{y}_{\mathrm{P},\mathrm{M}} \end{bmatrix} \end{array}$$

# Methods: *Small Area Estimation*

For each model, we aggregate posterior samples after burn-in and thinning to produce a posterior distribution for $\mu$, the mean biomass for a given small area of interest.

$$\begin{array}{cc} & \longleftarrow\ \mathbf{M}\ \text{MCMC samples}\ \longrightarrow \\ & \begin{array}{ccc}\text{mean} & \text{mean} & \text{mean}\end{array} \\ \mathbf{P}\ \text{Pixels} & \begin{bmatrix} \hat{y}_{1,1} & \cdots & \hat{y}_{1,\mathrm{M}} \\ \vdots & \ddots & \vdots \\ \hat{y}_{\mathrm{P},1} & \cdots & \hat{y}_{\mathrm{P},\mathrm{M}} \end{bmatrix} \end{array}$$

# Methods: *Small Area Estimation*

For each model, we aggregate posterior samples after burn-in and thinning to produce a posterior distribution for $\mu$, the mean biomass for a given small area of interest.

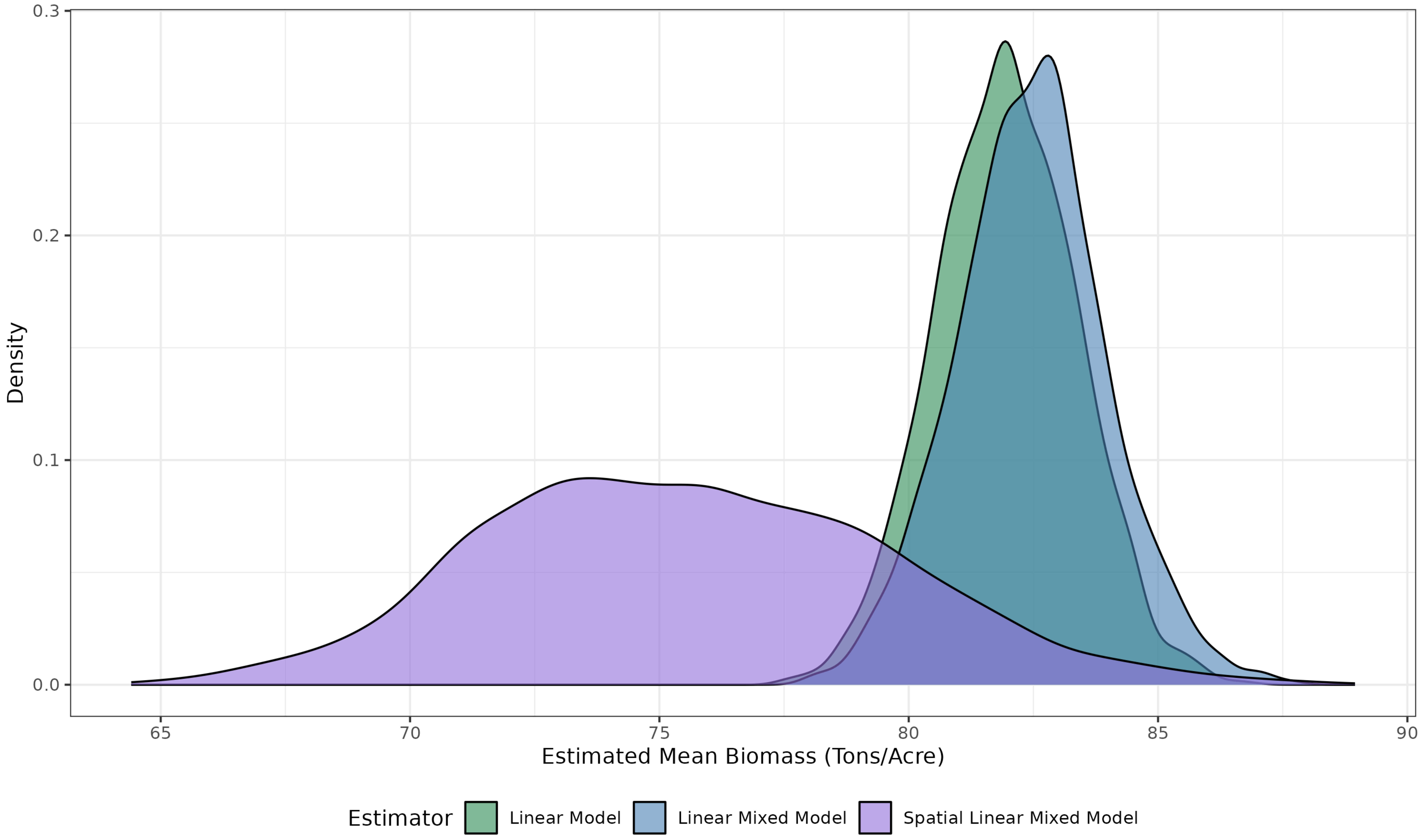
$$\xleftarrow{\qquad} \textbf{M} \text{ MCMC samples} \xrightarrow{\qquad}$$

$$\left[\hat{\mu}_1 \quad \hat{\mu}_2 \quad \cdots \quad \hat{\mu}_M\right]$$

# Results

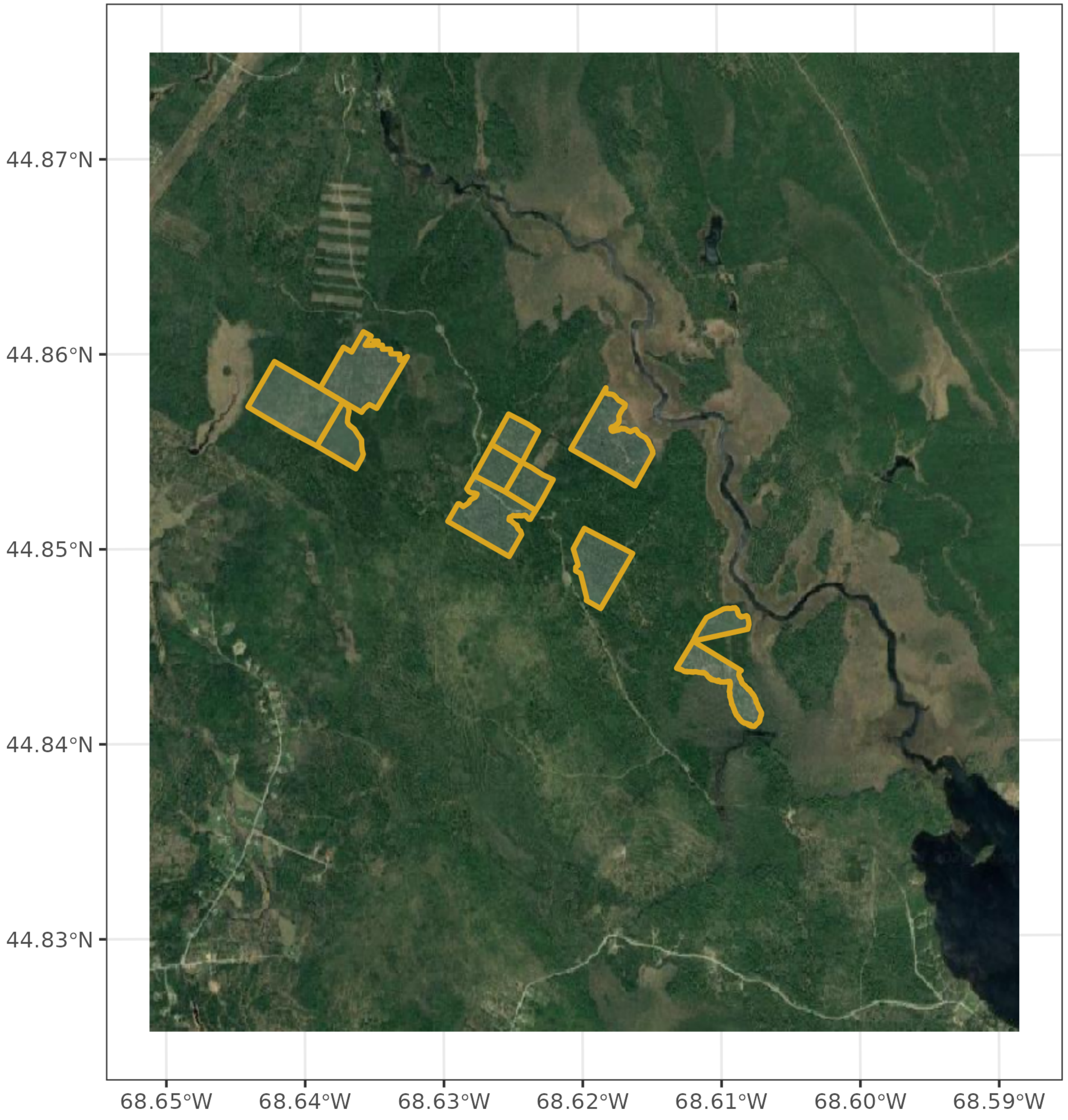
# Spatial Model Pixel Predictions in MUs

# Posterior Distributions of Mean Biomass

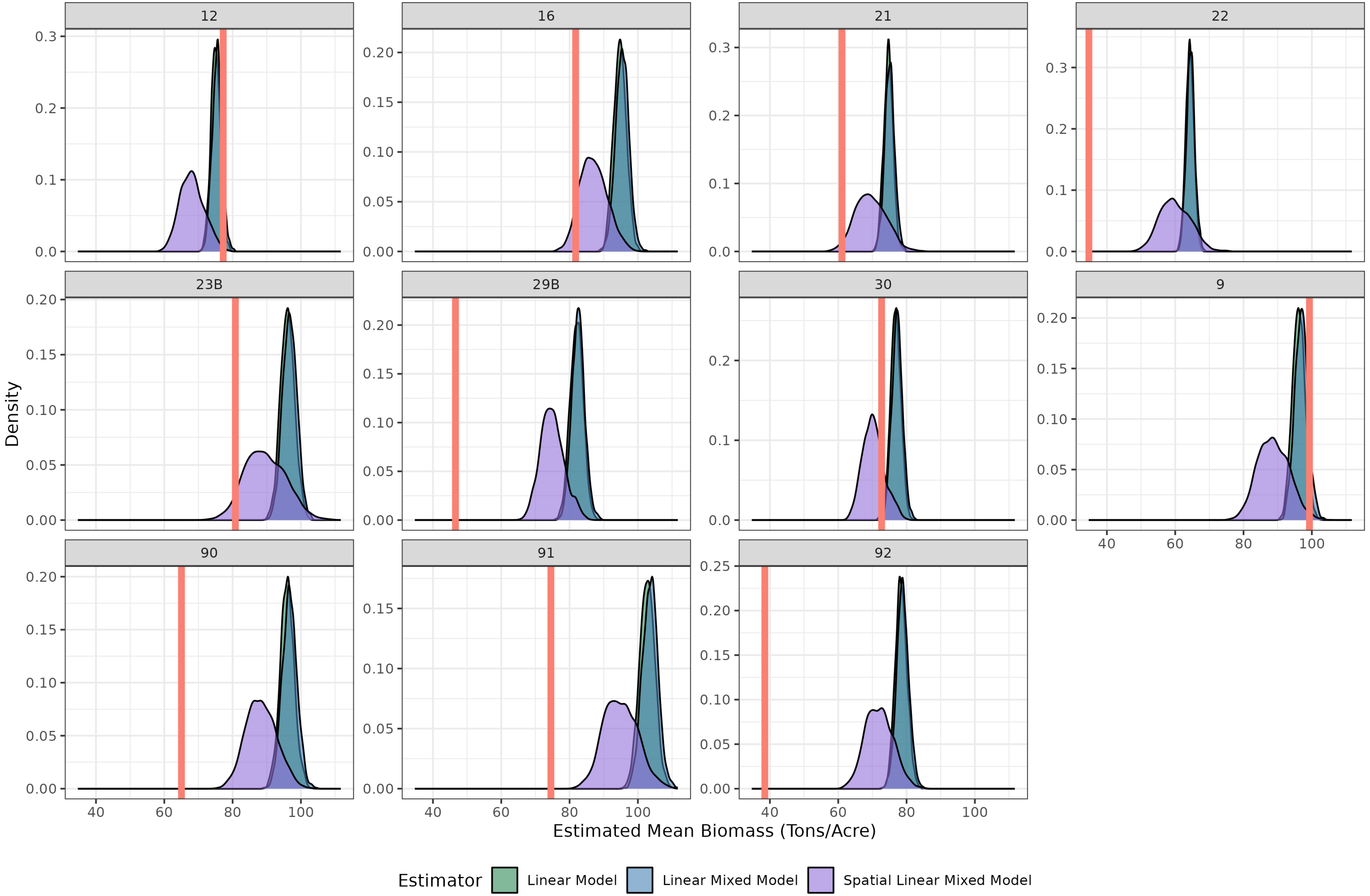


# Posterior Distributions of Mean Biomass

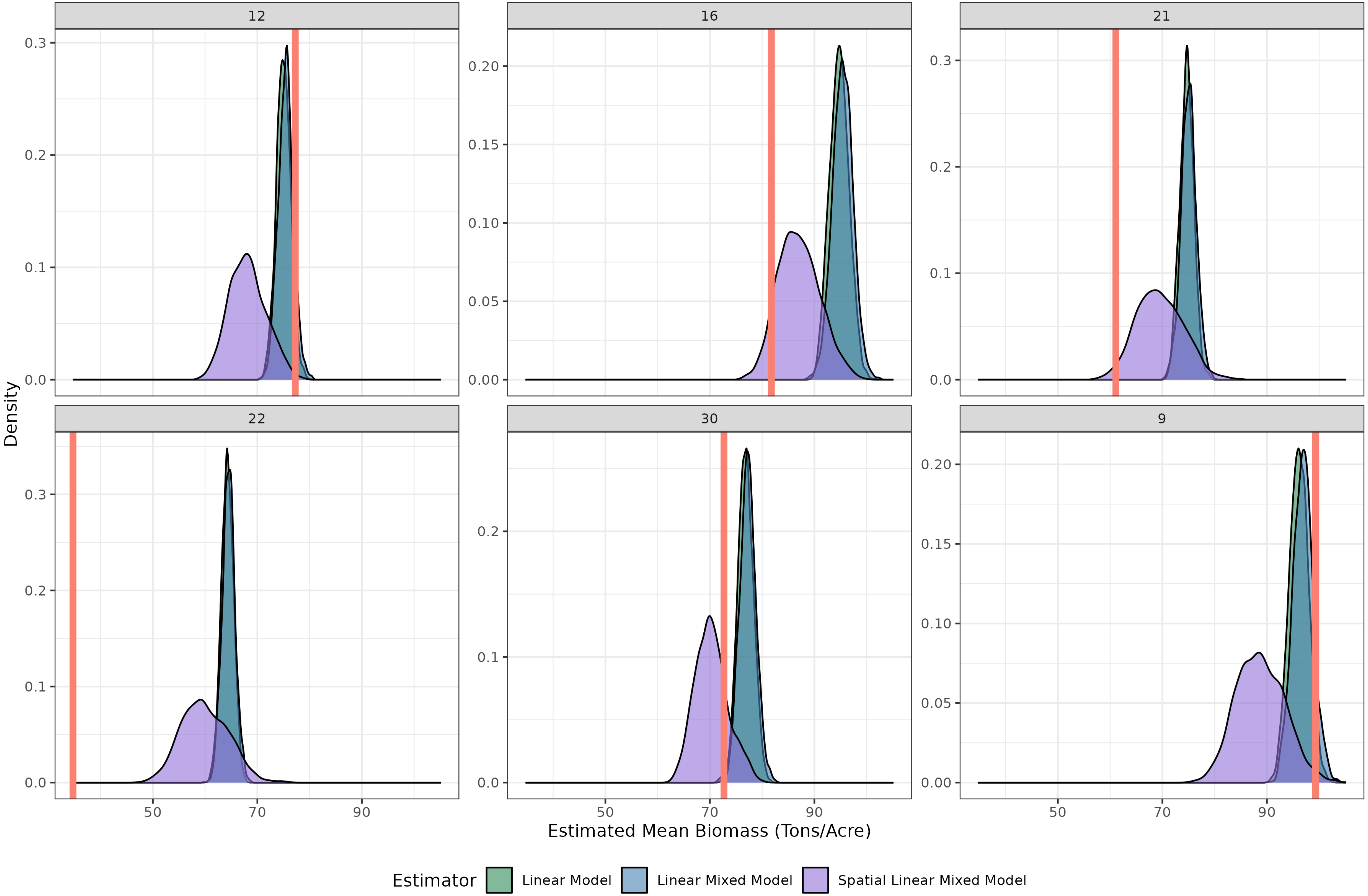
# Recall Our 11 Management Units (MUs)



# Mean Biomass by MU



# Mean Biomass by MU (Excluding 2011 Data)



# Discussion and Further Work

- Space-varying coefficient model.
- Further explorations of performance in different small areas of interest.
- Potential of deriving more predictors from this data.

Image credit: University of Maine

# References

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, *111*(514), 800-812.

Finley, A. O., Datta, A., & Banerjee, S. (2022). spNNGP R Package for Nearest Neighbor Gaussian Process Models. *Journal of Statistical Software*, *103*(5), 1–40. https://doi.org/10.18637/jss.v103.i05

Frescino TS, Moisen GG, Paterson PL, Toney JC, & White GW (2023) FIESTA: A Forest Inventory Estimation and Analysis R Package. *Ecography.* https://doi.org/10.1111/ecog.06428

# Thank you!
# Questions?

Image credit: University of Maine