Assessing Carbon Sequestration Dynamics in Wisconsin Forests

A Bayesian Approach

Introduction

- A number of strategies have been proposed to address **climate change**. One that is currently being implemented is commonly referred to as **'Cap and Trade'** - involves placing an overall limit ('cap') on carbon emissions, allowing carbon emitters to pay carbon sequesterers ('trade') to ensure the overall carbon limit is met.

- If someone wants to emit more carbon than their allotted *cap*, then they need to *trade* with someone else. For example, a power plant burning natural gas and emitting CO_2 (net carbon emitting) might need to buy offsets from someone removing CO_2 from the atmosphere (net carbon sequestering).

Introduction

- Plants remove carbon from the atmosphere during the process of photosynthesis. **Trees sequester** especially large amounts of **CO_2**, using atmospheric carbon to create sugars, wood, bark, and leaves. Much of this carbon is then stored in the soil as wood and leaves decay (fossil fuels themselves come from ancient plant matter stored in soil).

- In the US, **California** has instituted a **'Cap and Trade'** policy, creating a **national market** within the US **for 'carbon credits.'** Forest owners in WI can sell credits (sometimes 'carbon offsets') for sequestered carbon, which CA companies emitting carbon then buy to meet their carbon 'cap.'

Carbon Capture in Wisconsin

- About half (48.9%) of Wisconsin's land area is covered by forest - 17.1 million acres out of 35 million acres (Forest Service, 2016).

- Of this forest, around 52% is held by private landowners (WI DNR)

- For economic reasons, several wood mills have recently closed in northern Wisconsin - carbon capture may provide alternate livelihood for forest owners.

- In order **to sell carbon credits**, forest owners need a scientifically based estimate of sequestered carbon as a **basis for pricing.**

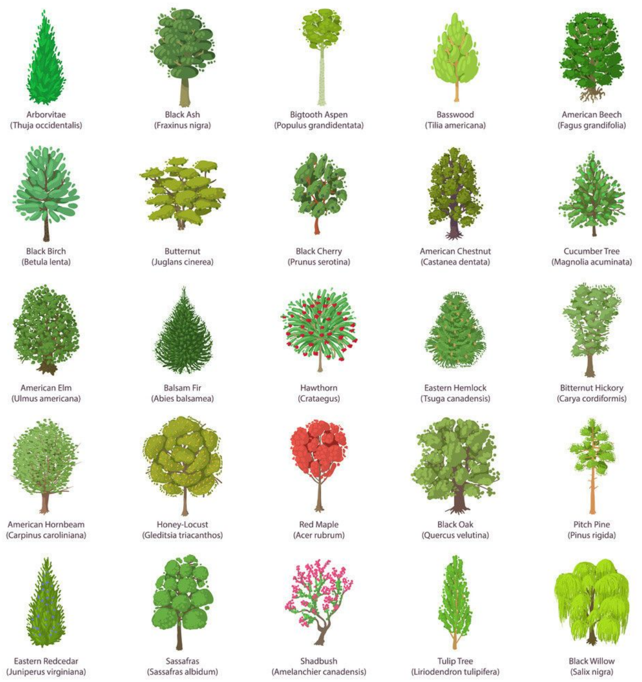
Research Question of Interest

How does the carbon capture of forests commonly grown on private land in WI differ by forest type and region?

Why is this important:

- Relative ‘Carbon Price’ of a forest depends on both forest type and region; some regions are better suited to certain forest types, due to soil, local climate, and other factors

- More generally, we want to know what types of forests should be planted in what regions to sequester carbon and combat climate change.

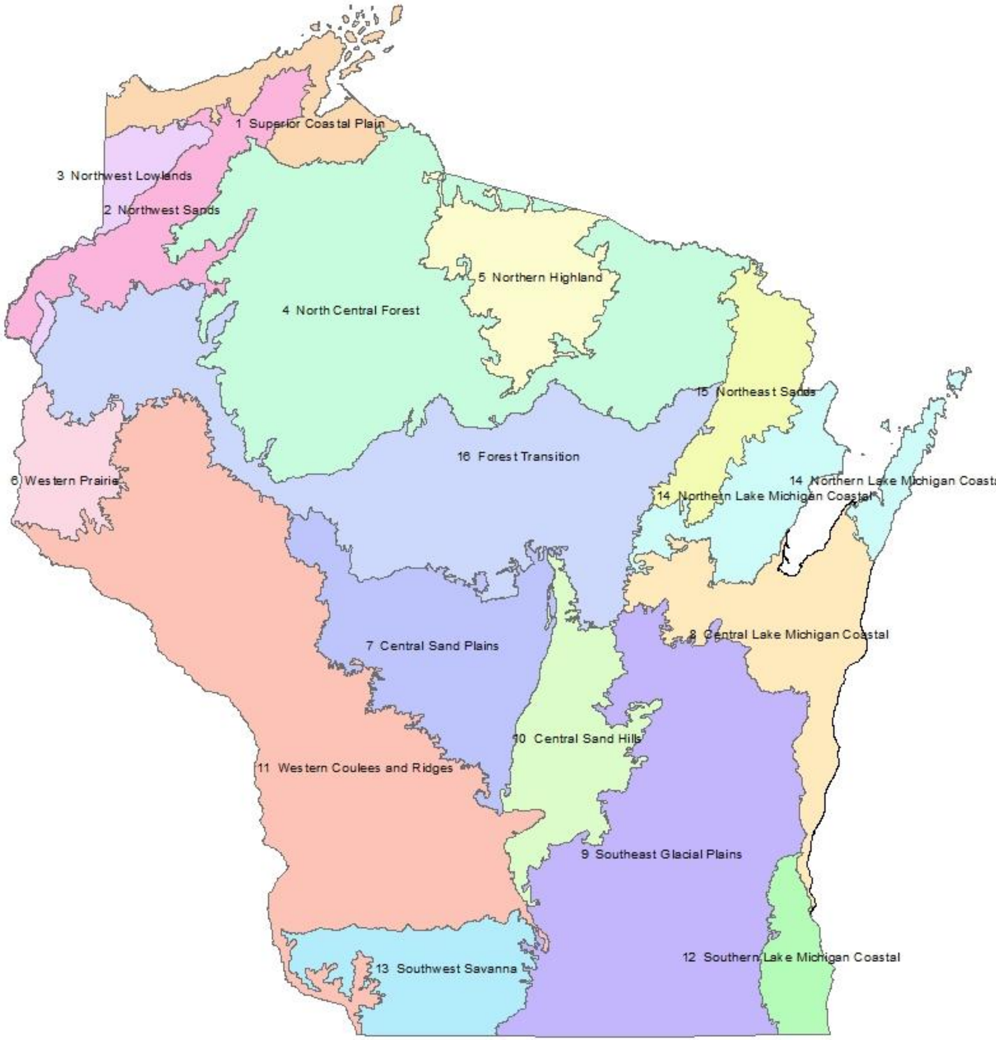


https://farmfoodfamily.com/types-of-trees/

The Data

Data

- For a given 120m x 120m plot: estimated carbon capture by photosynthesis (NPP in g C m^{-2}), estimate of forest type on that land
- **16 eco-regions**, over 21 years (2000-2021), 10 forest types:
  - Aspen, Central Hardwoods, Fir Spruce, Hemlock Hardwoods, Mixed Deciduous/Coniferous Forest, Mixed D/C Wetland, Northern Hardwoods, Oak, Pine, Red Maple



Data Sources

1. NPP estimates from University of Montana Numerical Terradynamic Simulation Group - using LANDSAT satellite imagery to estimate carbon capture through photosynthesis (accessed through Google Earth Engine)

2. Land cover data from WiscLand2 - Wisconsin Department of Natural Resources landcover database

3. Ecological zones from WI DNR eco-regions database - divides the state into 16 regions based on soil, climate, and plant communities.

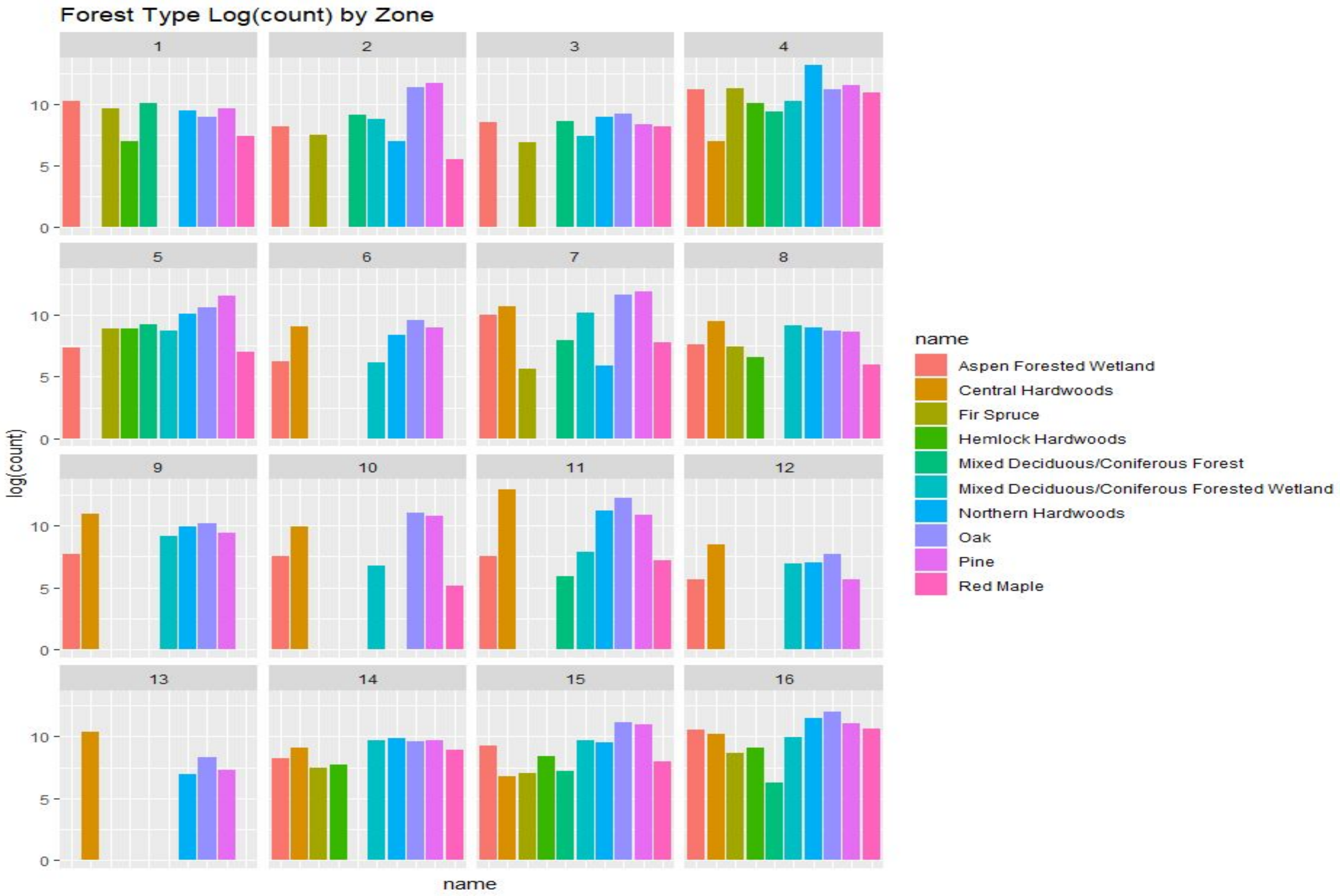
Data Cleaning & Aggregation

1. NPP Data initially at 30m x 30m resolution → aggregate to 120m x 120m using mean

2. Land Cover Data (WiscLand2) initially at 30m x 30m resolution → aggregate to 120m x 120m using mode (most common forest type out of 4 pixels, random assignment when evenly balanced)

3. Using Eco-Regions of Wisconsin: Aggregate raw NPP observations for each forest type - zone - year combination → median (less sensitive to outliers and tail behavior than mean). Drop categories with n < 150.

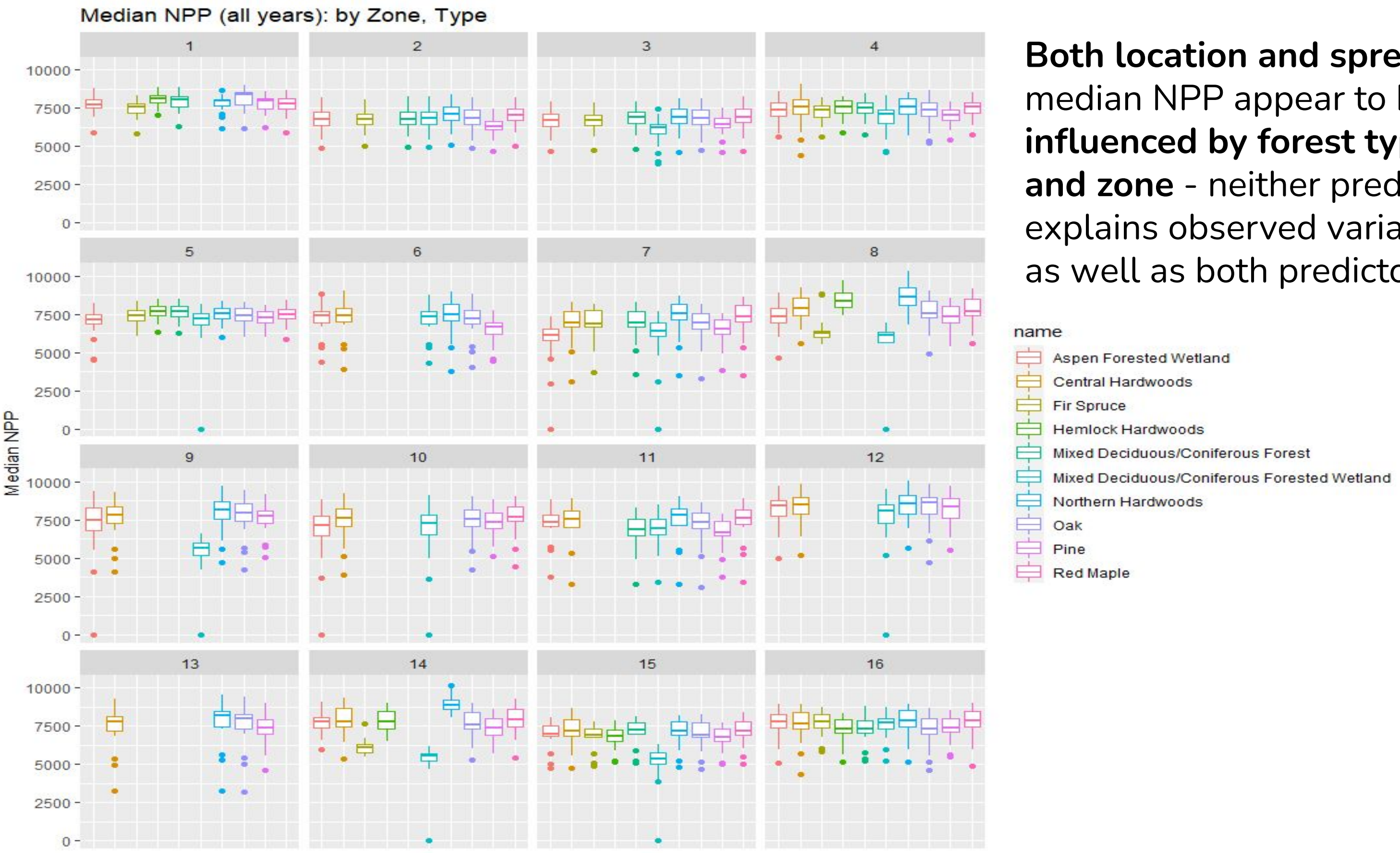
   (raster, OGR, rgdal, sp packages in R)


Forest Type Log(count) by Zone
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
log(count)
0
5
10
name
name
Aspen Forested Wetland
Central Hardwoods
Fir Spruce
Hemlock Hardwoods
Mixed Deciduous/Coniferous Forest
Mixed Deciduous/Coniferous Forested Wetland
Northern Hardwoods
Oak
Pine
Red Maple

Median NPP vs. Year: by Zone, Type
Aspen Forested Wetland
Central Hardwoods
Fir Spruce
Hemlock Hardwoods
Mixed Deciduous/Coniferous Forest
ed Deciduous/Coniferous Forested Wetla
Northern Hardwoods
Oak
Pine
Red Maple
Median NPP
10000
7500
5000
2500
0
year
0
5
10
15
20
zone
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

Our Model

Model Setting: observed heteroskedasticity in data



Both location and spread of median NPP appear to be **influenced by forest type and zone** - neither predictor explains observed variation as well as both predictors

→ Allow both median NPP (μ_{NPP}) and SD of median over time (σ_μ) to vary based on zone, type

→ Handle this simply : MLR / decomposition of variance

Treating all predictors as factors:

$$\boldsymbol{\mu_{NPP}} \sim \tau + X \begin{bmatrix} \Delta_{type} \\ \Delta_{zone} \\ \Delta_{year} \end{bmatrix} \qquad \tau := grand\ mean\ of\ MNPP\ values = 7212$$

(R^2: 0.57 adj-R^2: 0.56)

$$\boldsymbol{\sigma_{NPP}} \sim \sigma_0 * \sigma_{type} * \sigma_{zone} \leftrightarrow \log(\sigma_{NPP}) \sim \log(\sigma_0) + \log(\sigma_{type}) + \log(\sigma_{zone})$$

(R^2: 0.78 adj-R^2: 0.73)

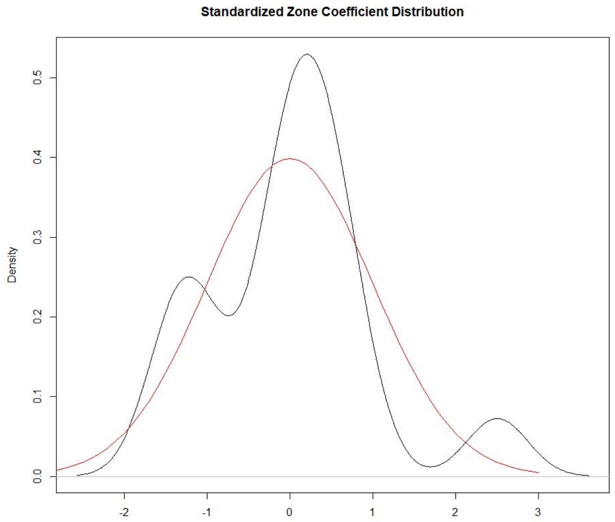
$$\log(\sigma_0) := grand\ mean\ of\ \log(\sigma_{NPP})\ values = 6.854$$

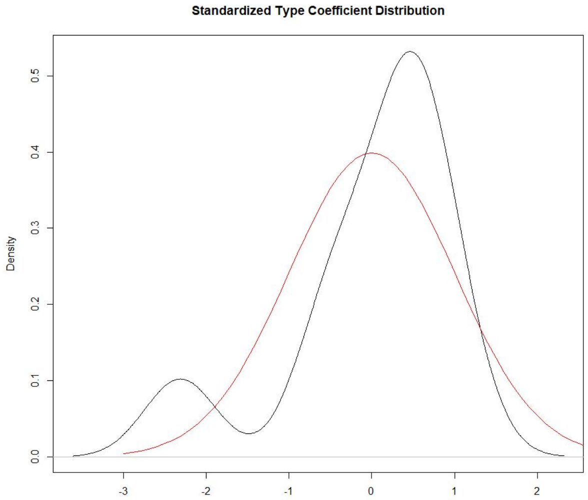
Prior Setting:

→ perform MLR on: (1) $\boldsymbol{\mu}_{NPP} \sim \tau + X \begin{bmatrix} \Delta_{type} \\ \Delta_{zone} \\ \Delta_{year} \end{bmatrix}$ (2) $\log(\sigma_{NPP}) \sim \log(\sigma_0) + \log(\sigma_{type}) + \log(\sigma_{zone})$

→ gives a sense of the spread of effects → SD for each prior

appropriate to fit standard normal prior to scaled model effects? Appears reasonable.





Model:

1 - MNPP location and spread models:

$$\boldsymbol{\mu}_{\boldsymbol{NPP}} \sim \tau + X\begin{bmatrix}\Delta_{type}\\ \Delta_{zone}\\ \Delta_{year}\end{bmatrix} \qquad \tau := grand\ mean\ of\ MNPP\ values = 7212$$

$$\boldsymbol{\sigma}_{\boldsymbol{NPP}} \sim \sigma_0 * \sigma_{type} * \sigma_{zone} \leftrightarrow \log(\sigma_{NPP}) \sim \log(\sigma_0) + \log(\sigma_{type}) + \log(\sigma_{zone})$$

$$\log(\sigma_0) := grand\ mean\ of\ \log(\sigma_{NPP})\ values = 6.854$$

2 - prior distributions:

$$\Delta_{type} \sim N(0, 420^2) \leftrightarrow \frac{\Delta_{type}}{420} \sim N(0,1)$$

$$\Delta_{zone} \sim N(0, 371^2) \leftrightarrow \frac{\Delta_{zone}}{370} \sim N(0,1)$$

$$\Delta_{year} \sim N(0, 807^2) \leftrightarrow \frac{\Delta_{year}}{807} \sim N(0,1)$$

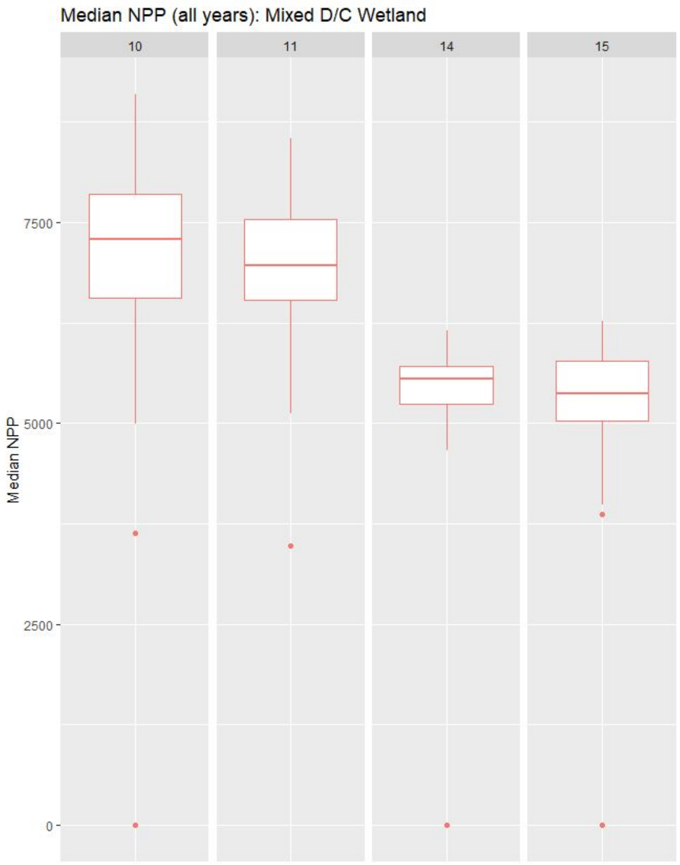
$$\log(\sigma_{type}) \sim N(0,1)$$

$$log(\sigma_{zone}) \sim N(0,1)$$

3 - observed value generating distribution:

$$Median\ NPP \sim N(\mu_{NPP}, \sigma_{NPP}^2)$$

Intuition Pump:



Location and spread of median NPP values appear to be affected by forest type and zone

← Boxplots of median across years for one forest type (mixed deciduous/coniferous wetland), 4 zones

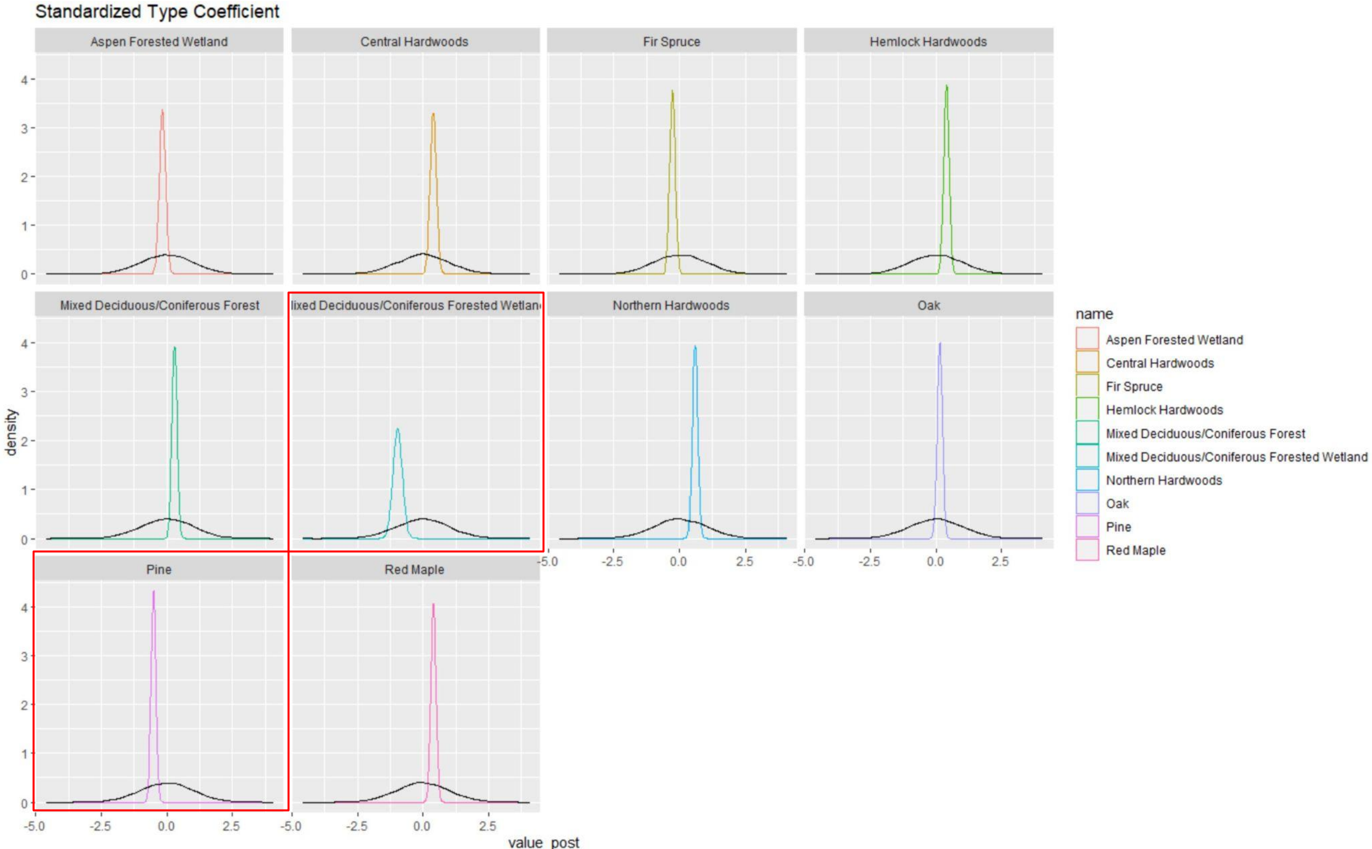
We set:

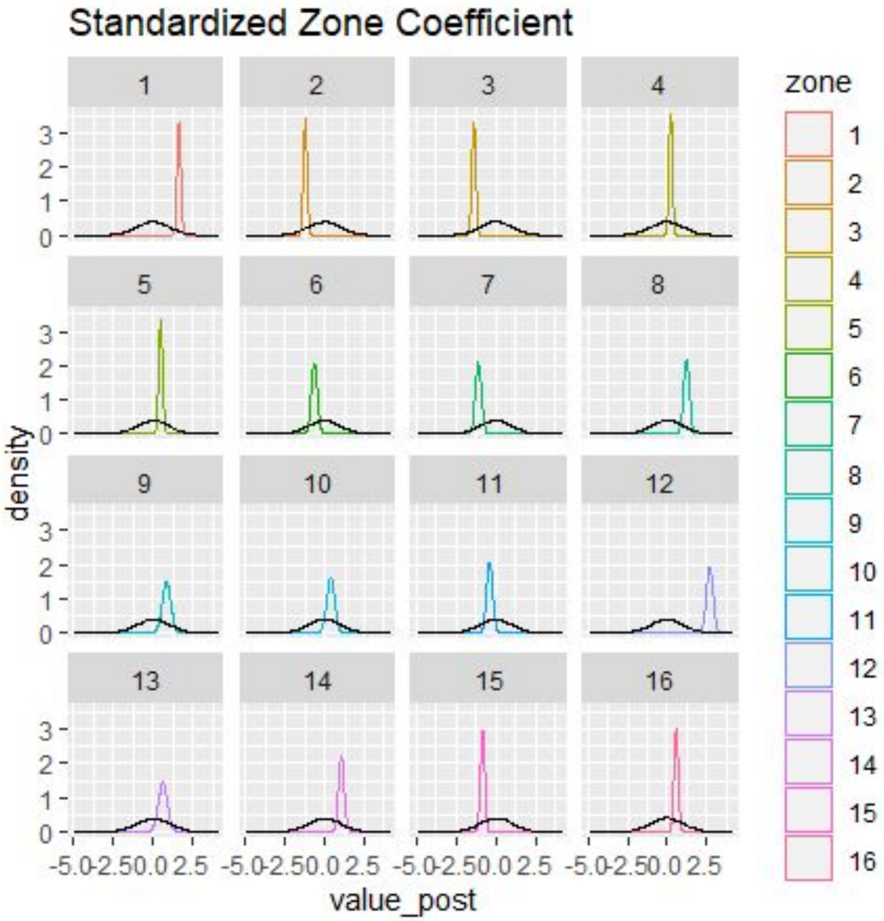
$\boldsymbol{\mu} = \mathrm{f}(\boldsymbol{\beta}_{\mu}, \text{zone, type, year})$
$\boldsymbol{\sigma}_{\mu} = \mathrm{f}(\boldsymbol{\beta}_{\sigma}, \text{zone, type})$

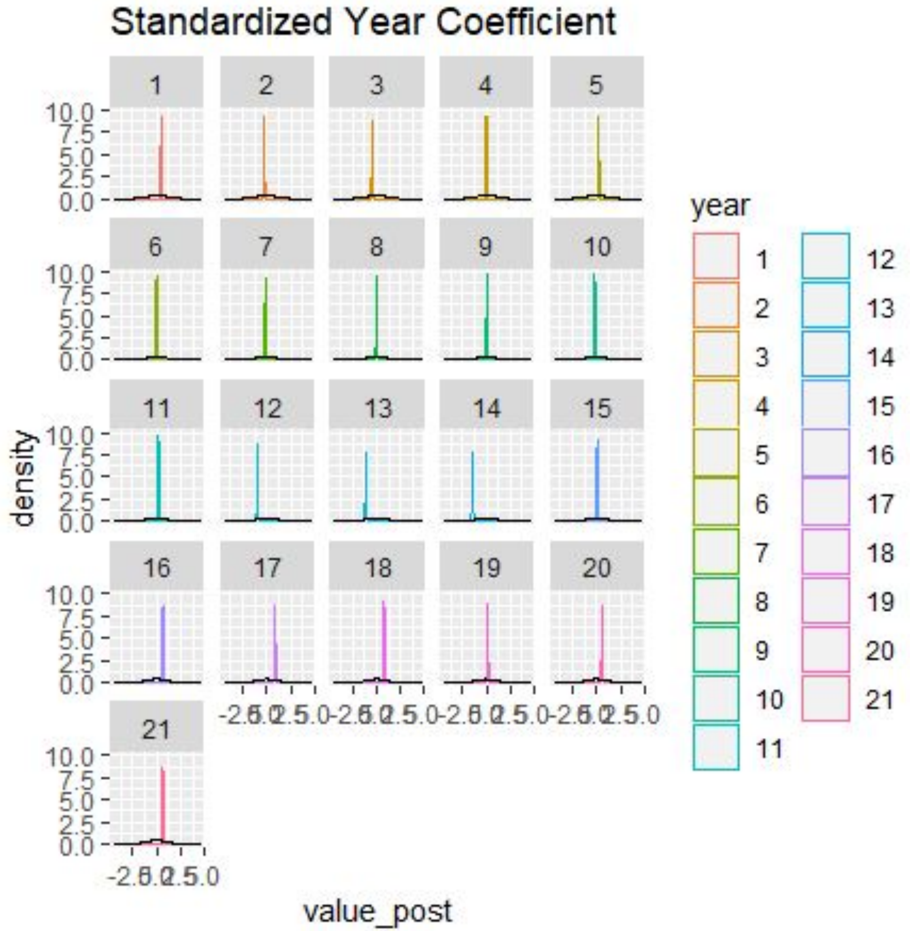
Median NPP ~ $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}_{\mu})$

Posterior Summaries

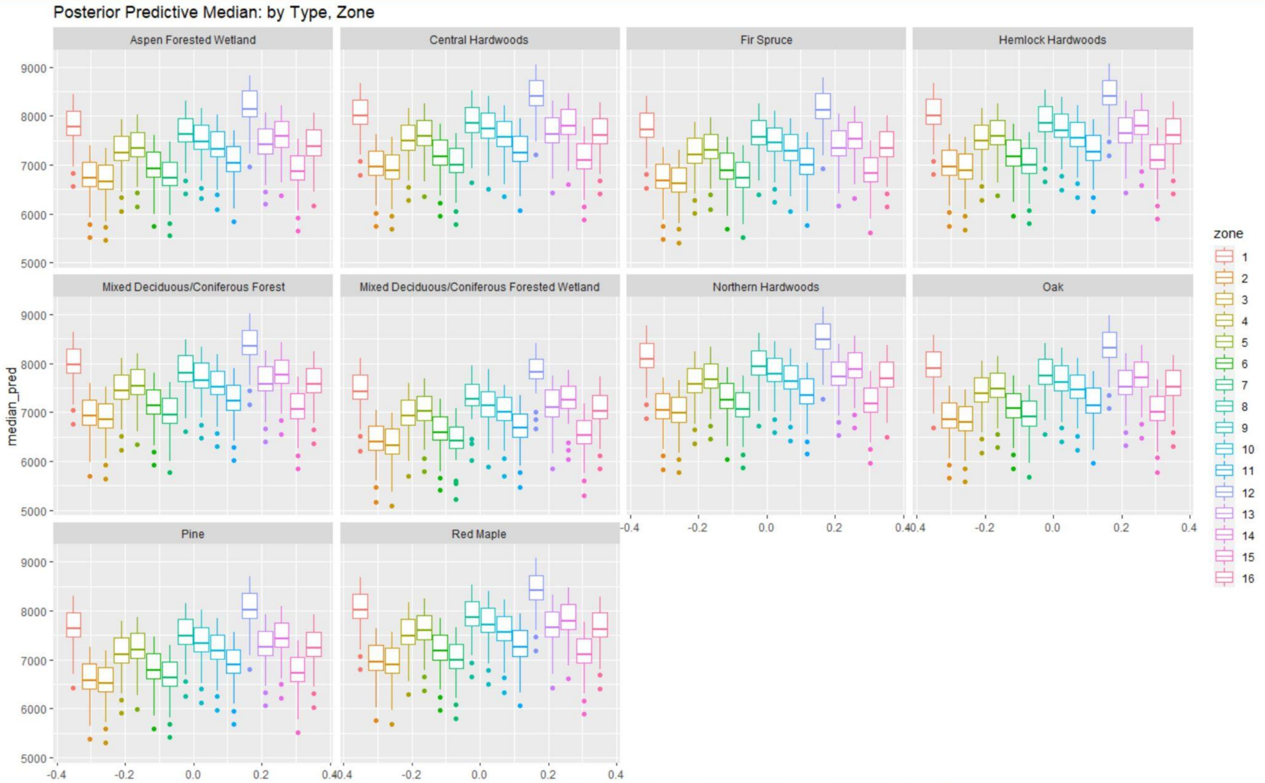
Prior vs Posterior for estimated parameter




Standardized Zone Coefficient
density
value_post
zone

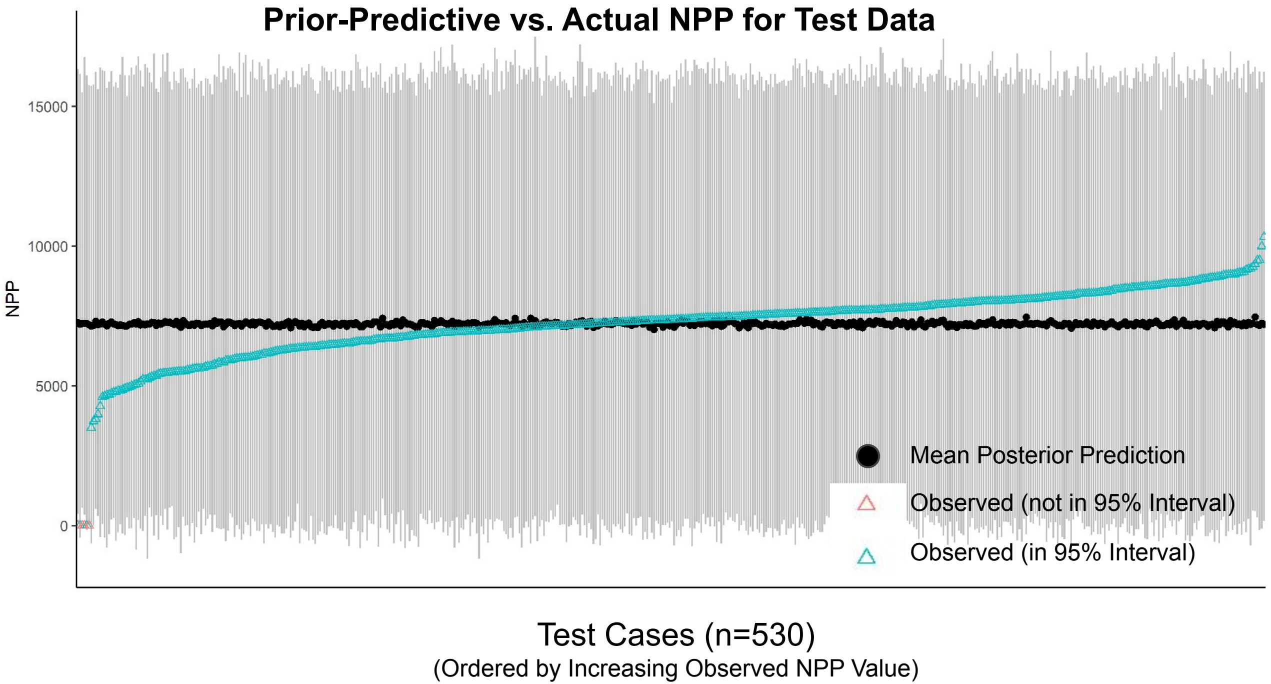

Standardized Year Coefficient
density
value_post
year

Posterior Predictive



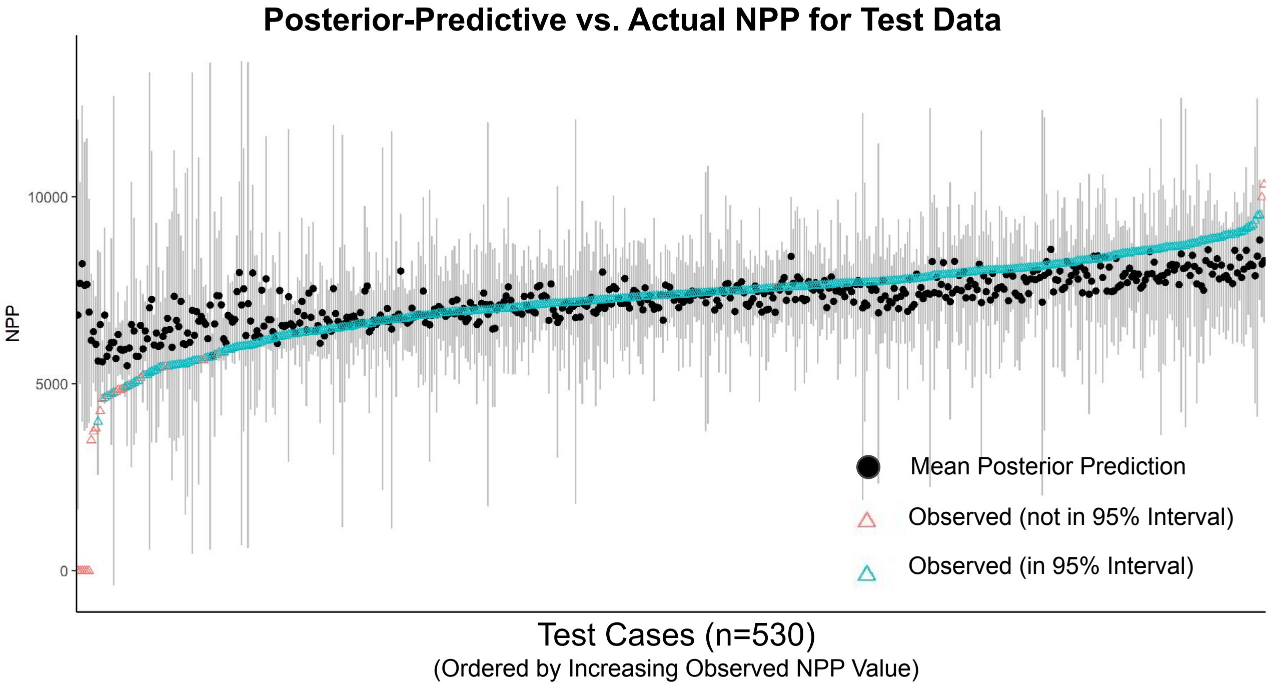
Given forest type and eco-region, we can estimate the median of carbon capture by photosynthesis (NPP in g C m^{-2}) in the future.

Prior-Predictive Model Testing



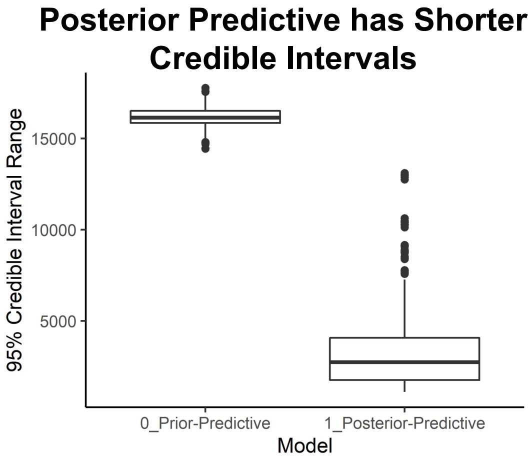
- We ***randomly*** partitioned the input dataset into 2 parts, with one comprising of 80% of data.
- The larger partition was used for parameter imputation and modeling.
- The smaller partition was used for testing prior and posterior predictive predictions.

Posterior-Predictive Model Testing



- ~95% (505 / 530) of observed median NPP per test case fell in the 95% credible interval of the posterior-prediction.

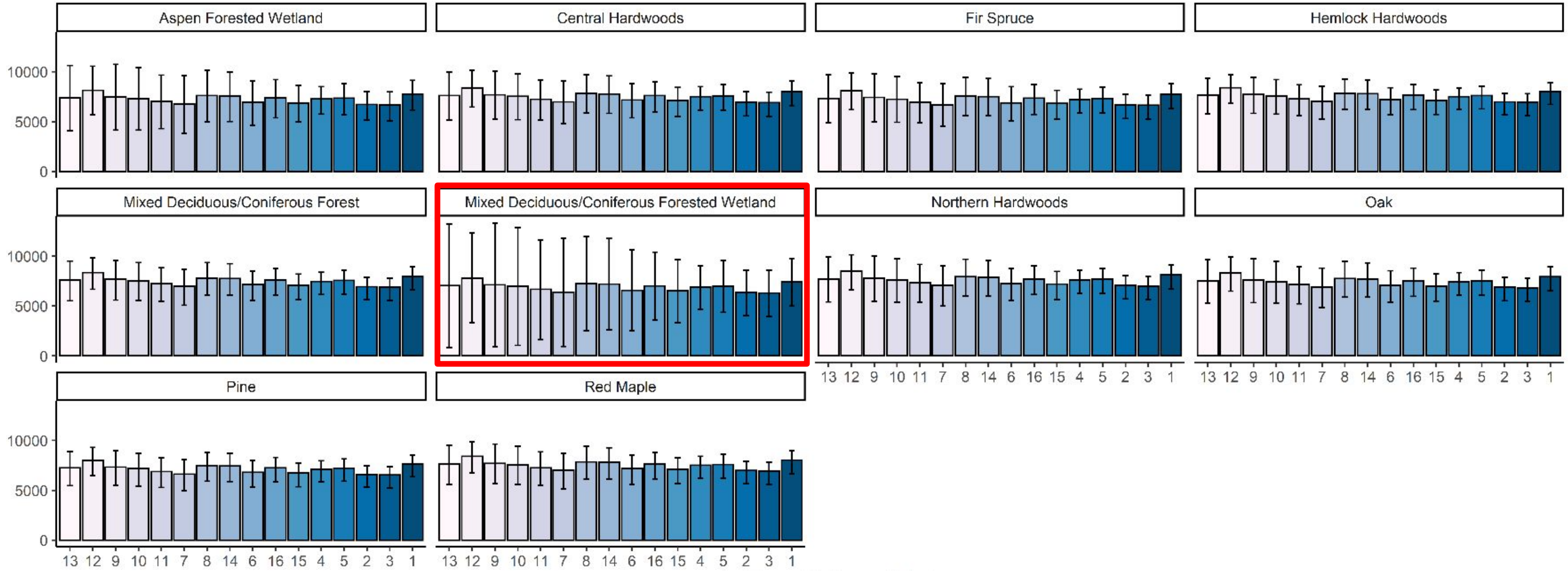
- ~60% (15 / 25) of test cases which were outliers, were related to the years of 2012 & 2013
  - "***The most intense period of drought occurred the week of July 24, 2012***" - National Integrated Drought Information System



Results

No major differences in NPP between northern and southern regions

Median NPP from Posterior Prediction Across All Years
(Bars Show 95% Interval)

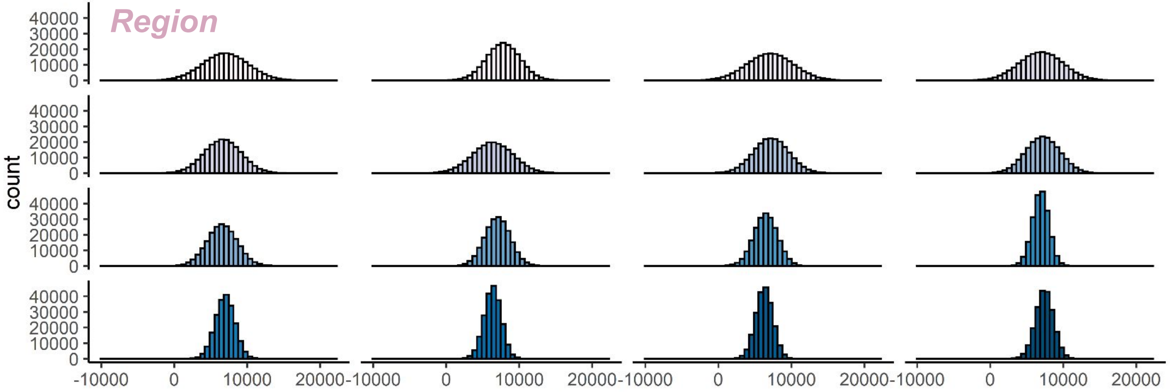


Regions/Zone IDs in Order of South → North

Posterior-predictive tends to be more variable in southern regions for Mixed Deciduous / Coniferous Forested Wetlands

This is not explained by availability of data, suggesting variability of certain trees might actually differ between southern and northern regions due to biological reasons such as climate differences along the gradient. Droughts have affected Southern WI most severely.

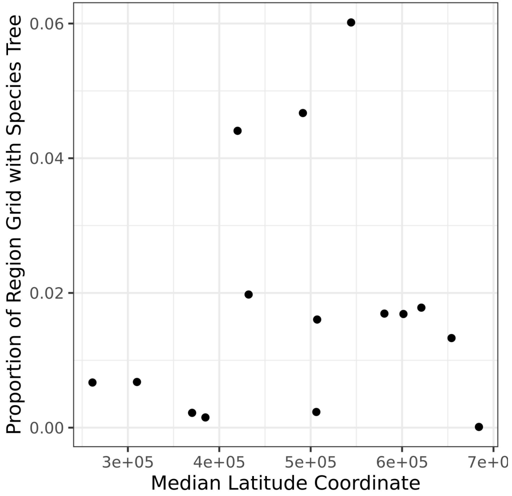
Southernmost Region



Northernmost Region

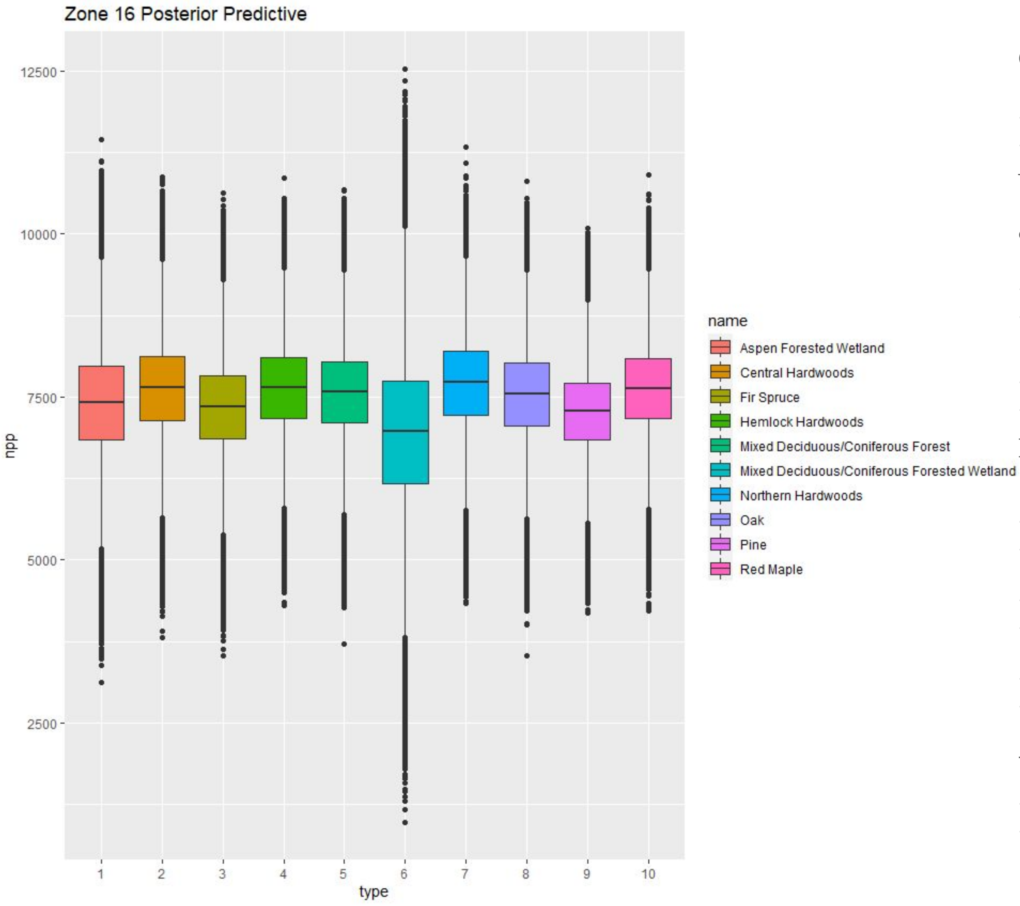
Mixed Deciduous / Coniferous Forested Wetlands Posterior Predictions Across All Years

No Clear Relationship Between Region Latitude and Proportion of Region Categorized as Mixed Deciduous / Coniferous Forested Wetland



South → North

Pricing Differentials



We construct posterior distributions and credible intervals for all zones and species.

Eventual carbon price likely depends on relative weighting of $\boldsymbol{\mu}$ vs $\boldsymbol{\sigma}$; posterior distributions allow consideration of both

E.g. Posterior Predictions for Zone 16 (left):

Northern Hardwoods (type 7) appear to capture more carbon than Pine (type 9)

Posterior 90% Credible Intervals

NH: (6457, 8858)

Pine: (6134, 8278)

Welch's t-test → 90% CI:
NH > Pine by (435, 438)

Model Criticism

- **Time series data** is often highly **auto-correlated** (current years related to observations of past years) - proceeding from a **time-series analysis** perspective rather than a **mixed effects model** might produce different/ useful results.

- **Forest type** and carbon sequestration are **connected to underlying factors** like soil and history of land management. We are only making inference about **existing forests**, **not potential future forests**. Drawing causative conclusions from this data is dangerous without further analysis (“you could plant more northern hardwoods in zone 16 and expect to sequester *X* g Carbon m^{-2}”).

- **Data** easily available **about precipitation/drought, wind** - could be factored into a more robust analysis which accounts for drought effects and more directly examines **carbon-capture resilience to wind, drought, extreme precipitation**.

- We **aggregated data to a median** (to make our large dataset manageable for Stan n ~ 2500 vs 78.5 million) - with more time and computing resources we could **fit a model to our full dataset, or fit separate models on subsets of the data and then create ensemble model**, which would account better for full variability observed in data, and provide a more realistic basis for carbon pricing of individual forest plots.

Conclusions

- Because **year-to-year variability is** an **important** facet of carbon sequestration, we construct a **multiple linear regression** based **Bayesian model** with **built-in heteroskedasticity** to investigate **forest type, zone, and year effects** on carbon sequestration across Wisconsin.

- Measured with an 80-20 training-testing split, this model appears **fairly accurate**. **96% of test cases** had predicted median NPP values falling **in the 95% credible interval of their respective posterior predictions**. Drought years were severe outliers → one main takeaway: **model only reflects as much variability as exists in data**.

- This model **summarizes variability seen in the past in posterior distributions**, which can be used to predict future carbon sequestration. But, **if expected future conditions are different from past conditions**, then **perhaps the model should be trained on a subset of past data** representing expected future conditions.

- Posterior distributions from our model both provide inference on past behavior and predict future carbon sequestration, but **specific pricing scheme would depend on regulatory and consumer tolerance for variability**.