Modeling PalEON biomass

Wesley Brooks

UW-Madison

May 28, 2013

Outline

Goal

- Produce a model of per-species biomass at time of settlement
- Complicated by the presence of zeros

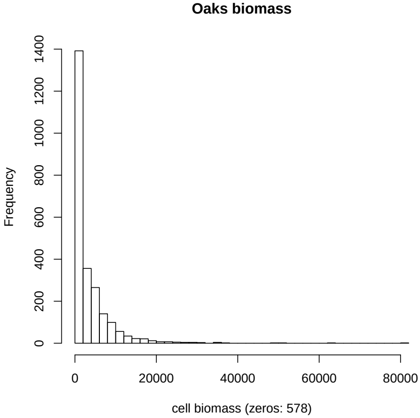
1. Data
2. Modeling
3. Model structure
   - Two-stage
   - One-stage
4. Spatial fitting
   - Independent grid cells
   - Spline models
   - Bayesian hierarchical model
5. Recap

Most common taxa

Taxon	Biomass
Oaks	7900000
Pine	6900000
Hemlock	6500000
Birches	6400000
Maple	4900000
Basswood	1700000
Elms	1600000
Tamarack	1600000
Cedar	1500000

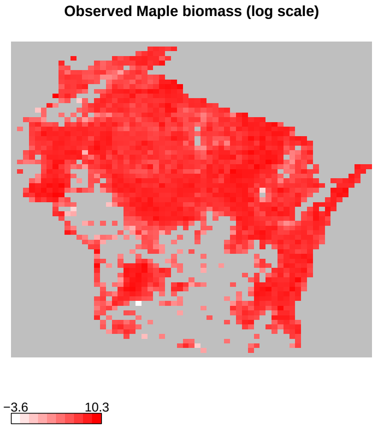
Oaks

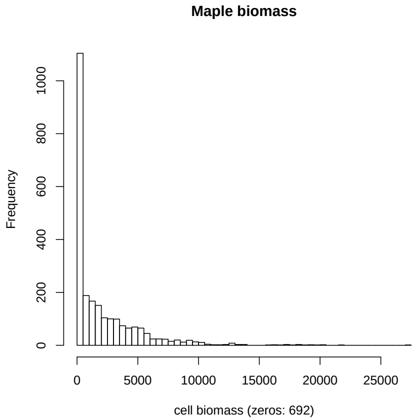
Observed Oaks biomass (log scale)
−0.9
11.3

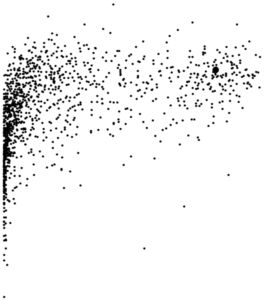


Oaks biomass
Frequency
0
200
400
600
800
1000
1200
1400
0
20000
40000
60000
80000
cell biomass (zeros: 578)

Pine


Observed Maple biomass (log scale)
−3.6
10.3


Maple biomass
Frequency
0
200
400
600
800
1000
0
5000
10000
15000
20000
25000
cell biomass (zeros: 692)



1. Data
2. **Modeling**
3. Model structure
   - Two-stage
   - One-stage
4. Spatial fitting
   - Independent grid cells
   - Spline models
   - Bayesian hierarchical model
5. Recap

Modeling grid

Model		Spatial fit: Indep.	Splines	GMRF
	One-stage (Tweedie)			
	Two-stage (Bernoulli-Gamma)			

- Model types:
  - Tweedie: single stage model that accounts for exact zeros
  - Binomial-Gamma: binomial stage for presence/absence and gamma stage for biomass, conditional on presence
- Spatial fitting methods:
  - Independent: ignore spatial structure, treat observations as conditionally independent
  - Splines: smoothing spline for the spatial effect
  - GMRF: spatial random effect (fit with the INLA algorithm)

Defining terms

Let:

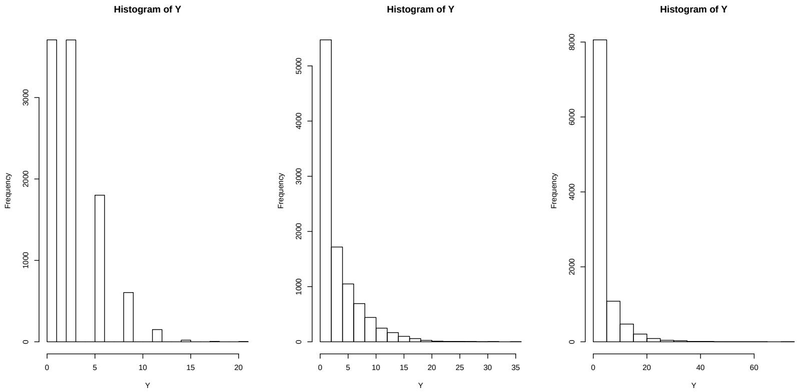
- s index location, k index taxon
- $Y_{k,s}$ denote the biomass of taxon k in cell s
- $p_{k,s}$ denote the composition fraction of taxon k in cell s
- γ_s denote the overall stem density in grid cell s

Two-stage models

- First stage: zero/non-zero
  - Logistic regression
  - $Z_s \sim \text{Bernoulli}(\zeta_s)$
  - $\text{logit}(\zeta_s) = f(\cdots)$
- Second stage: distribution of positive biomass
  - $Y_s | Z_s = 1 \sim \text{Gamma}(\alpha_s, \beta_s)$
  - $\text{E}\left(Y_s | Z_s = 1\right) = \mu_s = g(\eta_s)$
  - $\eta_s = h(\cdots)$

One-stage models: the Tweedie family

- Tweedie-family distributions have a point mass at zero as well as a continuous positive distribution.
- Parameter θ controls the mixture, from $\theta = 1$ (Poisson) to $\theta = 2$ (Gamma)



Conceptualizing the Tweedie distribution as a Gamma-Poisson mixture with parameters α, β, and λ:

- Draw $N \sim \text{Poisson}(\lambda)$
- Now make N iid draws: $V_\ell \sim \text{Gamma}(\alpha, \beta)$
- $Y = \sum_{\ell=1}^{N} V_\ell$

Tweedie model

With θ given, the Tweedie distribution is in the exponential family.

- $EY = \mu$
- $\text{var}(Y) = \phi\mu^{\theta}$
- ϕ is a scale parameter
- $P(Y = 0) = \exp\left(-\phi^{-1}\frac{\mu^{2-\theta}}{2-\theta}\right)$
  - $P(Y = 0) \uparrow$ as $\mu \to 0$ (as latent predictor $\eta \to -\infty$)
  - $P(Y = 0) \uparrow$ as $\phi \uparrow$
  - $P(Y = 0) \uparrow$ as $\theta \to 1$
- But we first need θ
  - Find $\hat{\theta}$ so that the model's deviance residuals match the assumed variance function.

1. Data
2. Modeling
3. Model structure
   - Two-stage
   - One-stage
4. **Spatial fitting**
   - Independent grid cells
   - Spline models
   - Bayesian hierarchical model
5. Recap

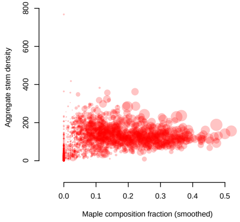
Independent grid cells

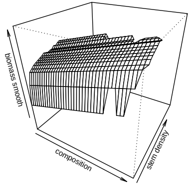
Ignoring the spatial structure, model the biomass as a function of composition and stem density

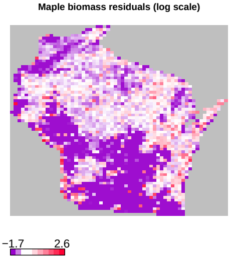
$$
\begin{aligned}
\eta_s &= f(p_{k,s}, \gamma_s) \\
\mu_s &= EY_s = \exp(\eta_s) \\
Y_s &\sim \text{Tweedie}(\mu_s, \theta)
\end{aligned}
$$

- $s(\cdot)$ is a smoothing spline.

The pictures are from a one-stage model for Maple:

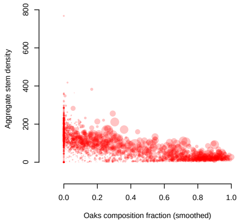


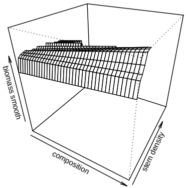


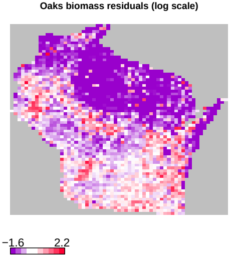


Independent grid cells

Pictures here are from a one-stage model for Oak.







Spline models

One way to account for spatial patterns is by using splines to model a spatial effect. Essentially the splines are a smooth function of latitude and longitude, and are fit using the same software as was used for the smooth functions of composition and stem density in the independent grid cell models.

$$\eta_s = f(p_{k,s}, \gamma_s) + g(s_x, s_y)$$
$$\mu_s = EY_s = \exp(\eta_s)$$
$$Y_s \sim \text{Tweedie}(\mu_s, \theta)$$