

Modeling PaleON biomass

Wesley Brooks

1. Introduction

Our objective is a model of the biomass of each species in each grid cell, where the only parameters used to calibrate the models are location and species composition. It is important to come up with an estimated probability that a given cell will have no biomass of a given species, and also to model the variance of the biomass estimate.

2. Philosophy

Jun and I discussed the interpretation of a model for biomass. At the time I saw the effort being directed toward getting a model that tells us, based on our survey of the forests at time of settlement, how much biomass there was on the landscape at time of settlement. On the other hand, Jun says that the goal is to model the process that gives rise to biomass on the landscape, observed through the survey at time of settlement.

As a practical matter, one implication is that grid cells where there were small but nonzero biomass observations are treated differently. In my view, the fact that, say, a spruce tree was seen in grid cell 517 is enough to say that there is zero probability of there being no spruce biomass in cell 517. In Jun's view, we could easily imagine the same process populating the landscape in such a way that, randomly, cell 517 has no spruce trees. So there's a positive probability of zero spruce in cell 517.

3. Models

There are two possible directions here: a two-stage hierarchical model, where in stage one we randomly decide whether a given cell has nonzero biomass, and then if so stage two randomly sets the biomass; alternatively, a single-stage model with zero-inflation, maybe one that could be tuned like a `glm` (e.g. a Tweedie model).

3.1. Single-stage (Tweedie model)

Tweedie family of distributions for $\theta \in (0, 1)$:

$$\begin{aligned}
 N &\sim \text{Poisson}(\lambda) \\
 Z_i &\sim \text{Gamma}(\alpha, \tau) \\
 Y &= \sum_{i=1}^N Z_i \\
 \lambda &= \phi^{-1} \frac{\mu^{2-\theta}}{2-\theta} \\
 \alpha &= \frac{2-\theta}{\theta-1} \\
 \tau &= \phi(\theta-1)\mu^{\theta-1} \\
 \eta(s) &= f(\mathbf{x}(s)) = g(\mu(s)) \\
 \mu(s) &= g^{-1}(\eta(s))
 \end{aligned}$$

Where $g(\cdot)$ is the link function, $\mathbf{x}(s)$ is the vector of covariates (x, y coordinates and composition fraction) at location s .

4. Figures

[Figure 1 about here.]

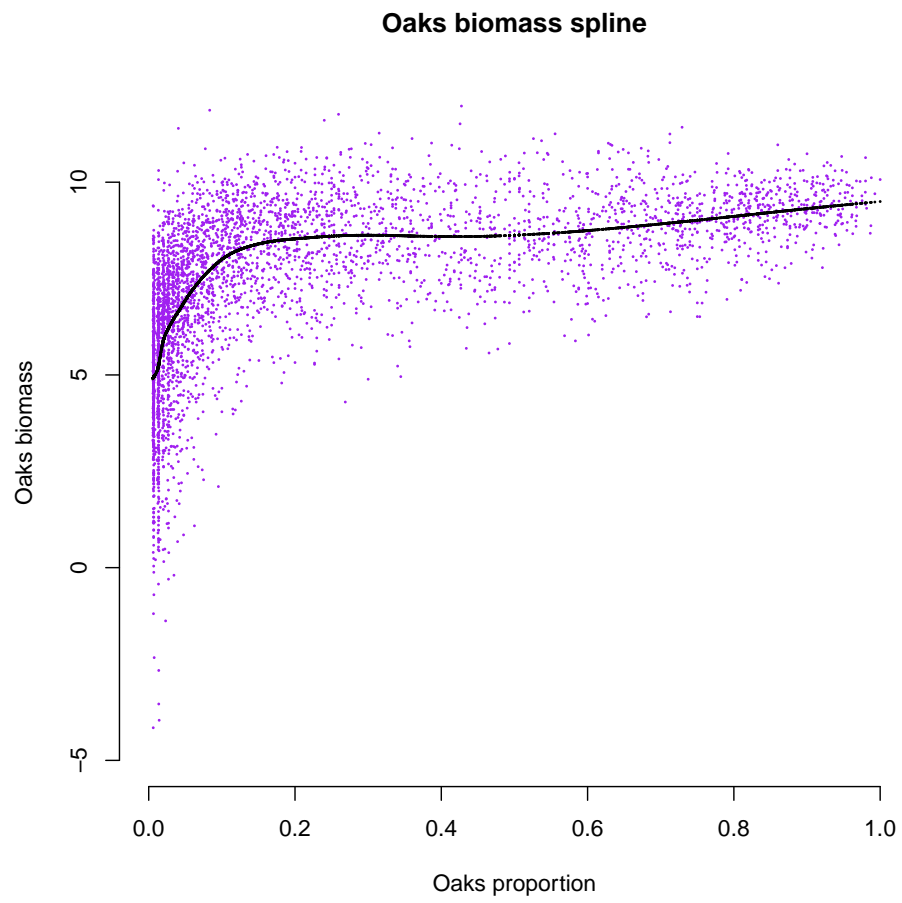
[Figure 2 about here.]

[Figure 3 about here.]

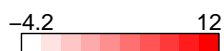
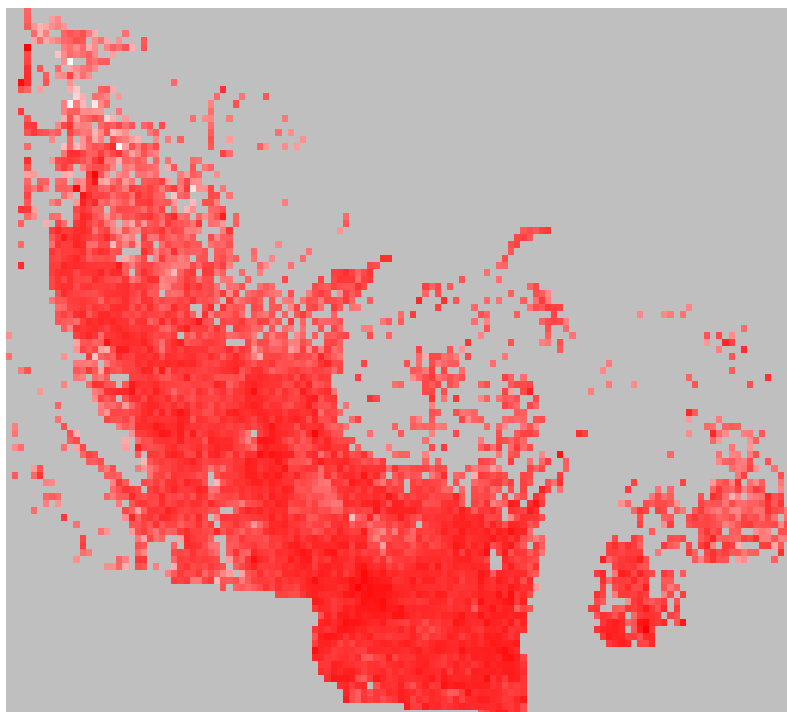
[Figure 4 about here.]

5. References

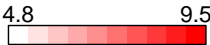
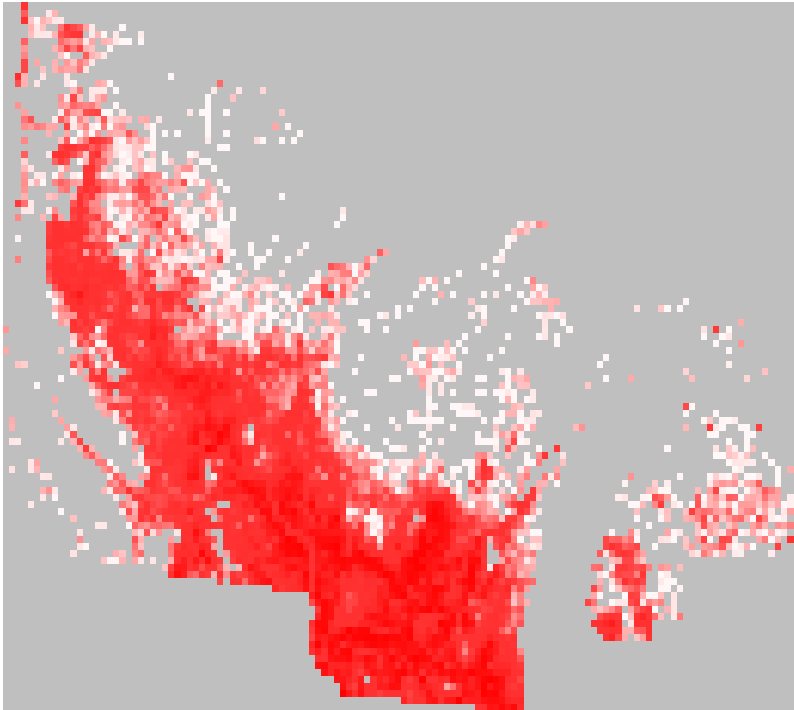
List of Figures



Oaks observed biomass (log scale)



Oaks fitted biomass (log scale)



Oaks biomass residual (log scale)

