

## Contents

<b>1</b>	<b>PaleON biomass modeling</b>	<b>1</b>
1.1	Biomass data . . . . .	1
1.2	Model . . . . .	1
1.2.1	Fitting the GAM . . . . .	2
1.2.2	Dividing biomass between taxa . . . . .	2
1.2.3	Drawing from the “posterior” of biomass . . . . .	2
1.2.4	Notes on the distribution of biomass among the taxa . . . .	2

## 1 PaleON biomass modeling

### 1.1 Biomass data

Per-grid-cell biomass data for the upper midwest, calculated from stem density and basal area via allometry. Simon did those calculations. For prototyping, we use only Wisconsin data (the biomass models take a while to run).

### 1.2 Model

Biomass is a positive, continuous quantity. It can be exactly zero where there are no trees growing, so we presume that the biomass follows a Tweedie distribution. An alternative would be to model biomass in two stages: presence/absence and then distribution.

We’re going to use a generalized additive model (GAM) with the sole predictor being a smooth on spatial location. R-INLA would be a great way to model biomass in a Bayesian context with the linear predictor being a GMRF, but there is no Tweedie likelihood available for R-INLA, and adding one turns out to be difficult (due to Rue’s choice of C compiler).

### 1.2.1 Fitting the GAM

There is not a generally agreed-upon method of estimating the Tweedie distribution's "power" parameter,  $p$ . But note that the Tweedie family is a type of "Exponential dispersion model", meaning that the "variance function" relating the variance to the expected value can be written  $E(y) = \mu$ ,  $\text{var}(y) = \phi\mu^\theta$  where  $\theta$  is the power parameter and  $\phi$  is the dispersion parameter.

### 1.2.2 Dividing biomass between taxa

There is concern that the variance of a sum of biomasses, modeled for individual taxa, would be greater than the variance of a single model for total biomass. I think this concern is backward, since e.g.  $x^2 + y^2 + z^2 < (x + y + z)^2$ . In any case, the plan is to make a total biomass model and divvy the total fitted biomass between the taxa based on draws from the biomass models for individual taxa.

### 1.2.3 Drawing from the "posterior" of biomass

Our goal is to calculate a distribution of total biomass, so we need to be able to make draws from the fitted model.

Consider that the smooth coefficients are drawn from a gaussian distribution with covariance. That distribution is conditional on the GAM's smoothing parameters (as always) and also on the Tweedie power parameter (specific to Tweedie models). We'd like the marginal distribution of biomass (not conditional on those fitted values) so we're going to use the parametric bootstrap to find their distributions. This part is covered in the Simon Wood mgcv book, sections 5.2.7, 5.4.2, and 4.9.3.

This will have to be done for all taxa, I think... it's going to be a slow process.

### 1.2.4 Notes on the distribution of biomass among the taxa

If the distribution of the biomass was gamma (and all taxa had the same shape parameter) then the distribution of biomass among taxa would be Dirichlet. It's unclear whether there is a friendly analog (generalization?) to the Dirichlet when the biomass follows a gamma distribution.