Spatial smoothing of zero-inflated abundance data

Wesley Brooks

Department of Statistics
University of Wisconsin–Madison

.

Introduction

Motivation and data

- PalEON project aims to model the historical climate
- including the amount of carbon that was locked up in trees in the US
  - based on pollen counts in sediment cores
  - requires correlating settlement-era tree biomass with the pollen counts from that time
- Measurements of the settlement-era tree biomass are from the 19th-century Public Land Survey (PLS)
- Area: Minnesota, Wisconsin, northern Michigan

Here we are fitting a spline-based model to spatially smooth the PLS biomass observations and produce draws from the fitted distribution of biomass.

Introduction

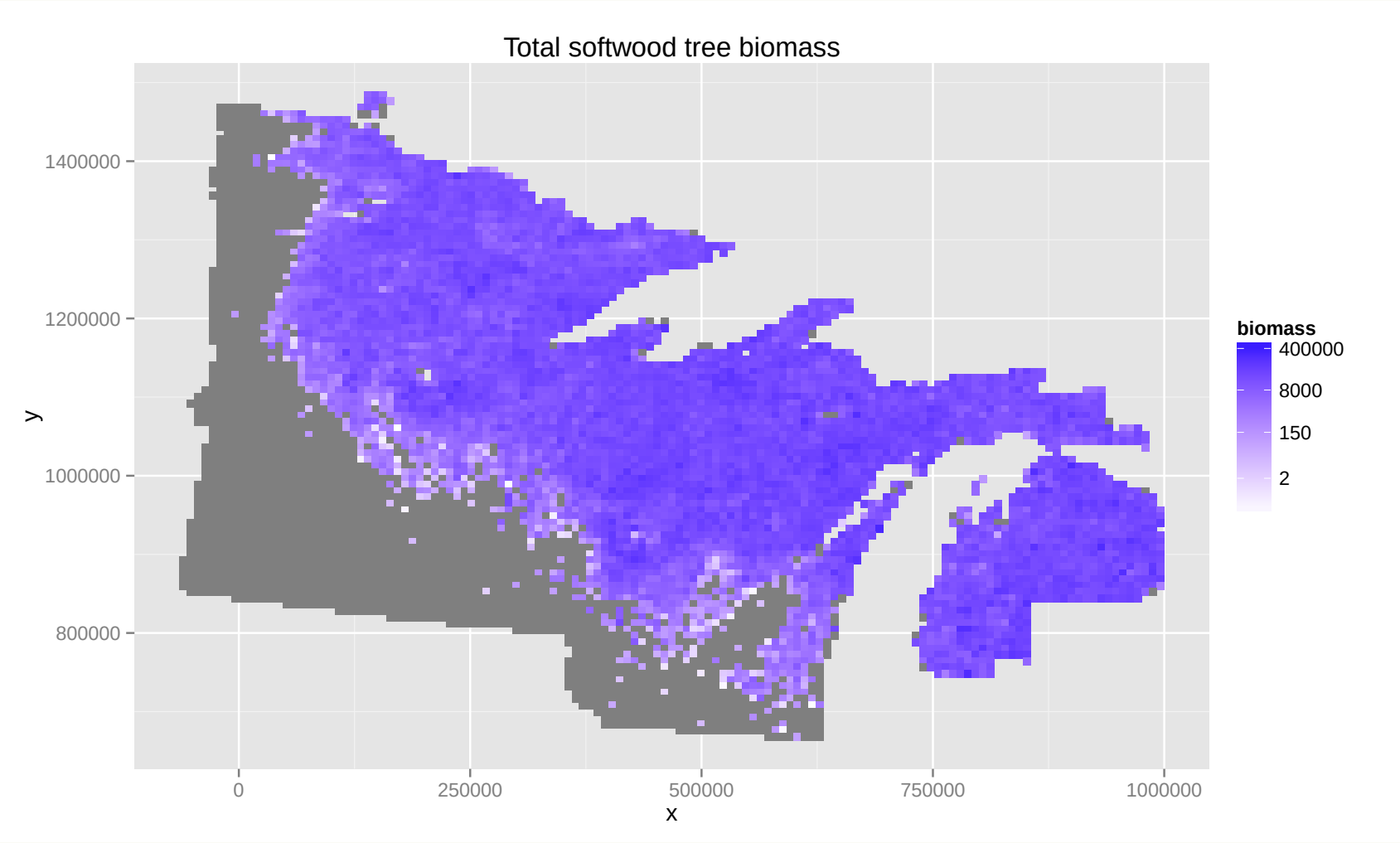


Figure: Observed softwood biomass in the upper midwest.

Introduction

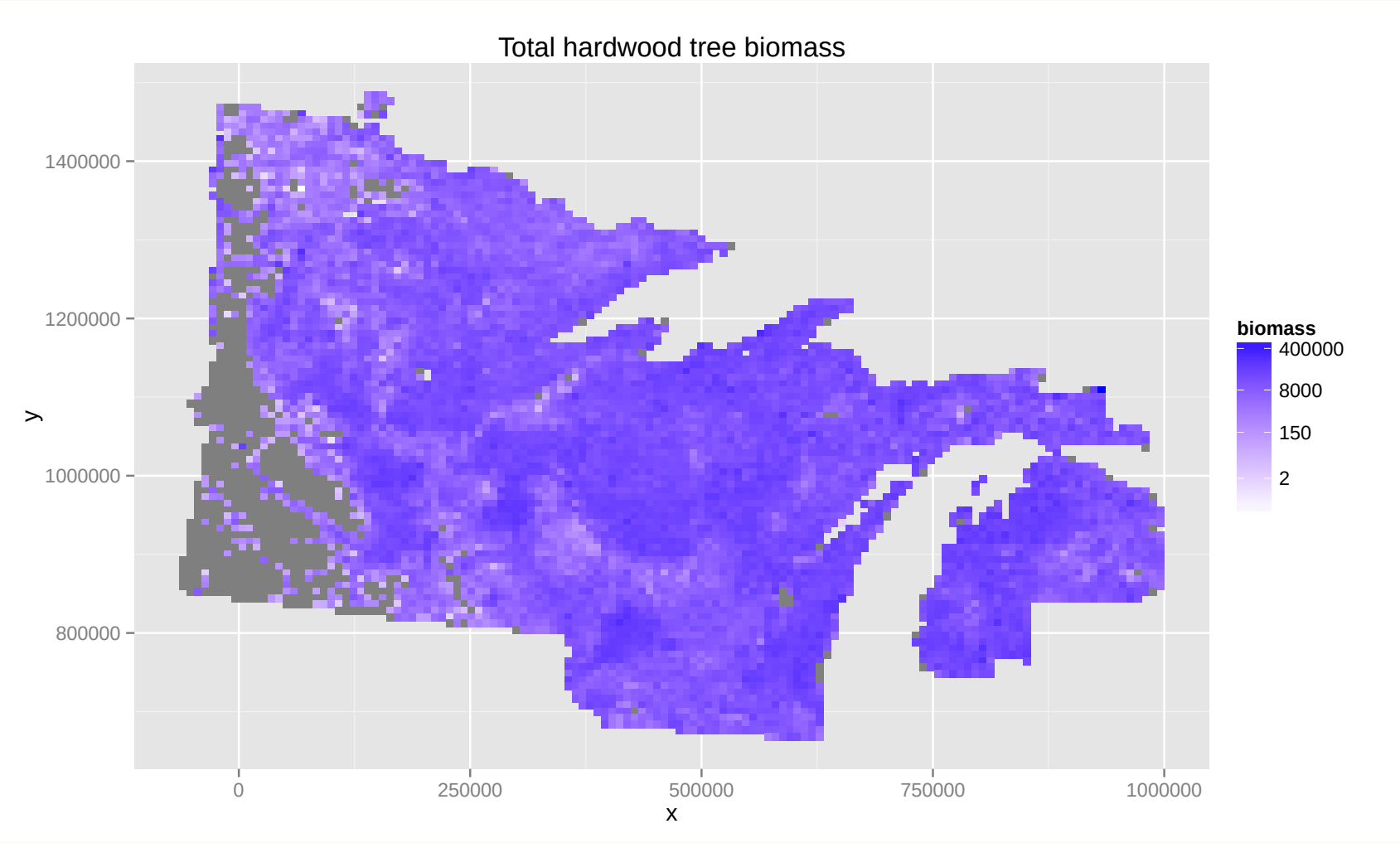


Figure: observed hardwood biomass in the upper midwest.

Model

Model

Tweedie distribution

- Poisson-Gamma mixture
- Response Y
- E$Y = \mu$

$$Y = \sum_{i=1}^{N} Z_i$$

$$N \sim \text{Poisson}(\lambda)$$

$$Z_i \sim \text{Gamma}(\alpha, \tau)$$

Model

Tweedie distribution

Poisson mean λ is a function of Tweedie parameters μ, θ, ϕ:

- $\lambda = \phi^{-1} \frac{\mu^{2-\theta}}{2-\theta}$
  - Power parameter $\theta \in (1, 2)$
  - Dispersion parameter $\phi > 0$

Gamma shape and scale are functions of the Tweedie parameters:

- $\alpha = \frac{2-\theta}{\theta-1}$
- $\tau = \phi(\theta - 1)\mu^{\theta-1}$

Model

Generalized additive model

- Used a generalized additive model (GAM)
- Covariates are spatial coordinates $s = (\text{lat}, \text{long})$
- Used 500 knots for model of Wisconsin
- Used 1000 knots for tri-state model

Model

Drawing from the model output

- GAM smoothing parameter and MLE $\hat{\theta}$ are point estimates
  - Modeled biomass distribution is conditional on these estimates
- Use parametric bootstrap to draw from marginal distribution of biomass
- Modeled after section 5.4.2 of **Wood-2006**
  - Here the smoothing parameter and $\hat{\theta}$ are estimated jointly for each bootstrap replication

Results

Results

Results are in two forms:

- Per-cell
- Total

The large area with zero softwood biomass seems to cause problems for estimating the GAM

Results

Tri-state softwoods

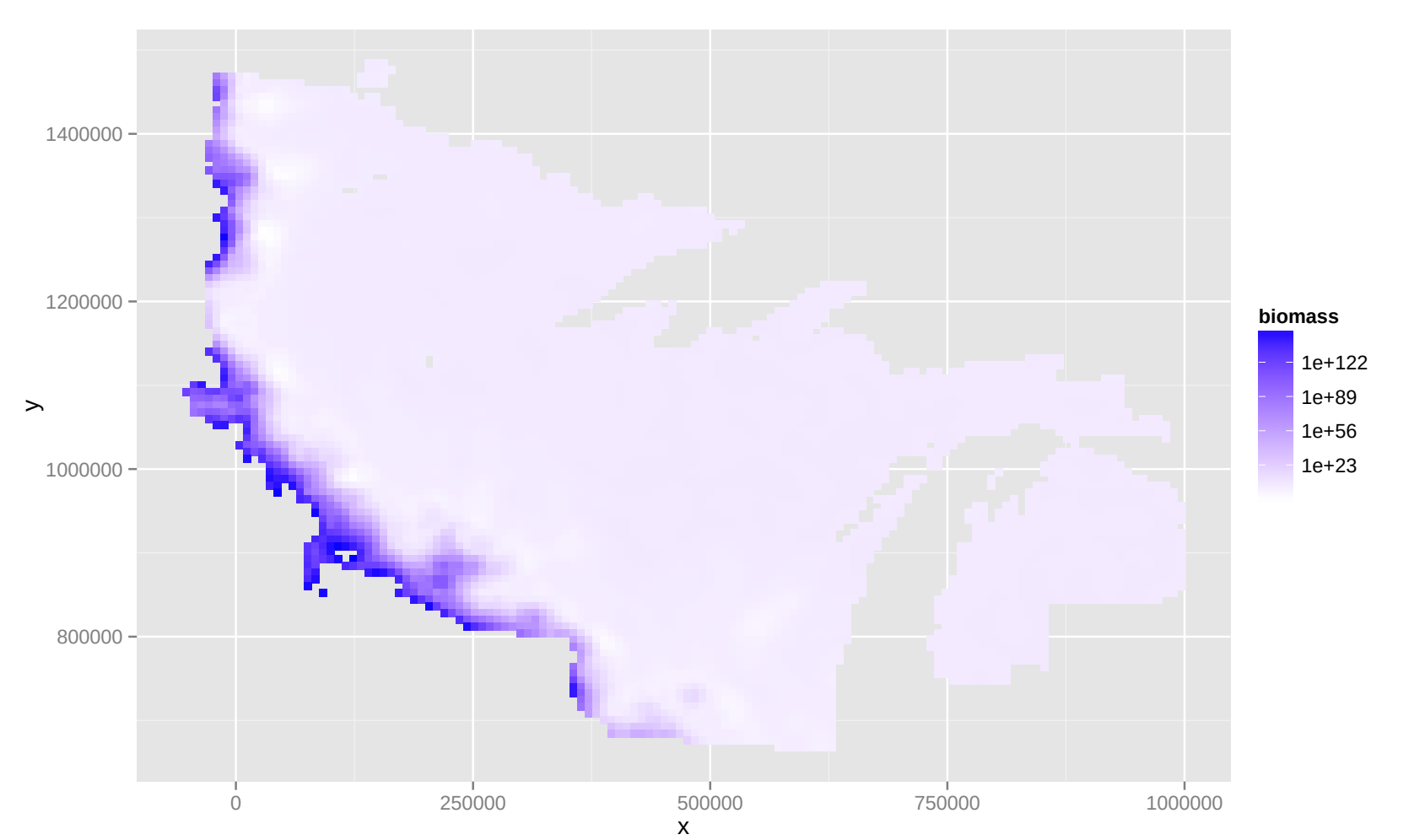


Figure: Cell means of draws for softwood biomass for the upper midwest.

Results

Tri-state hardwoods

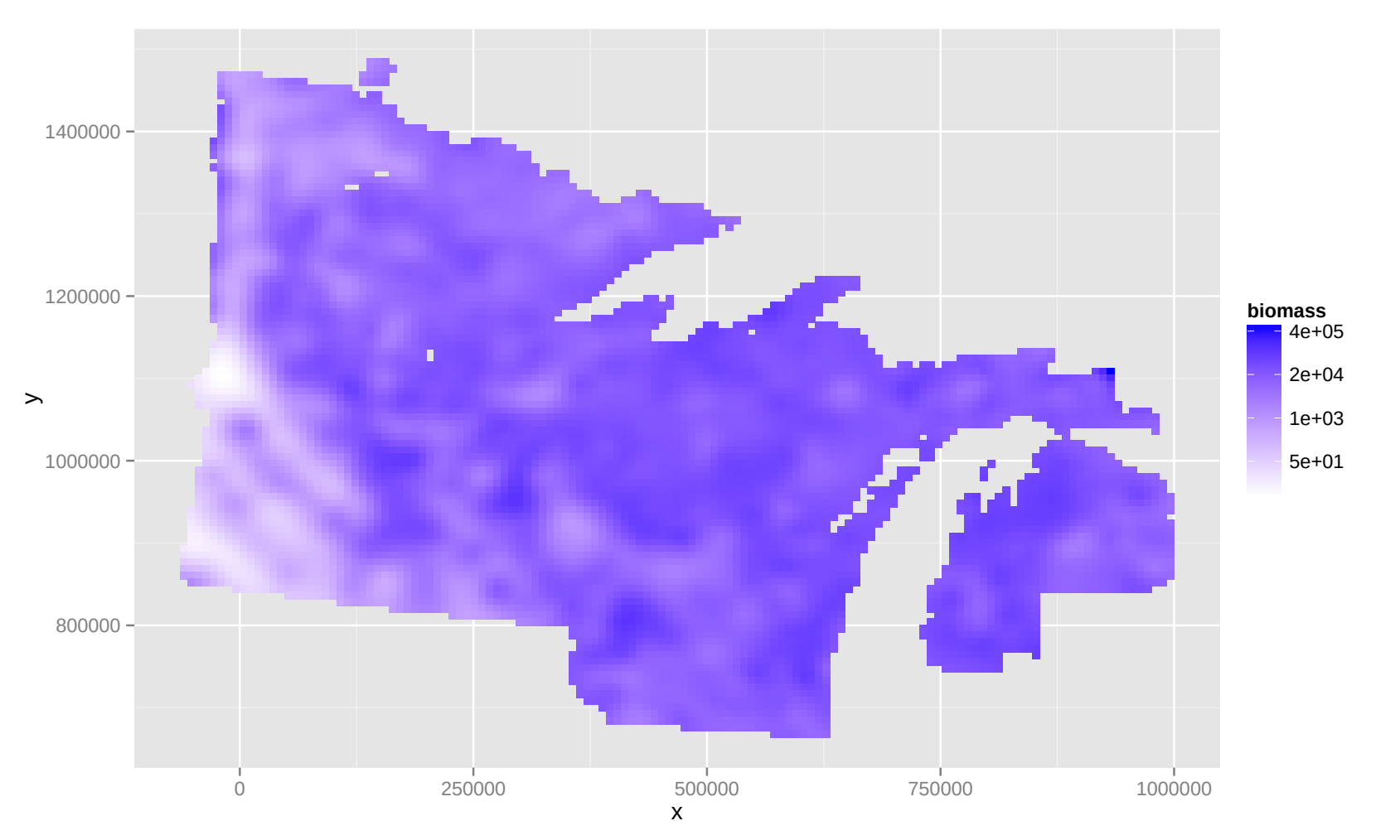
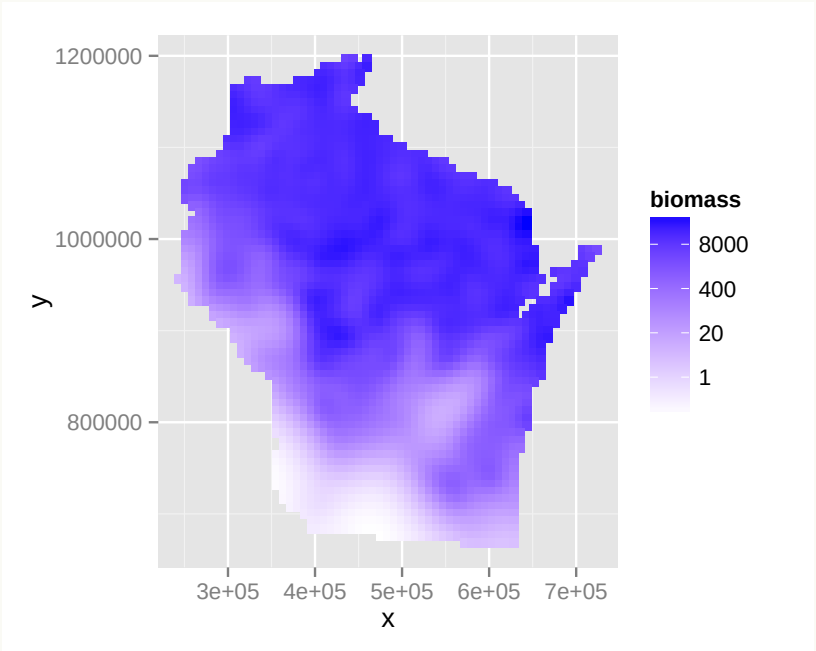


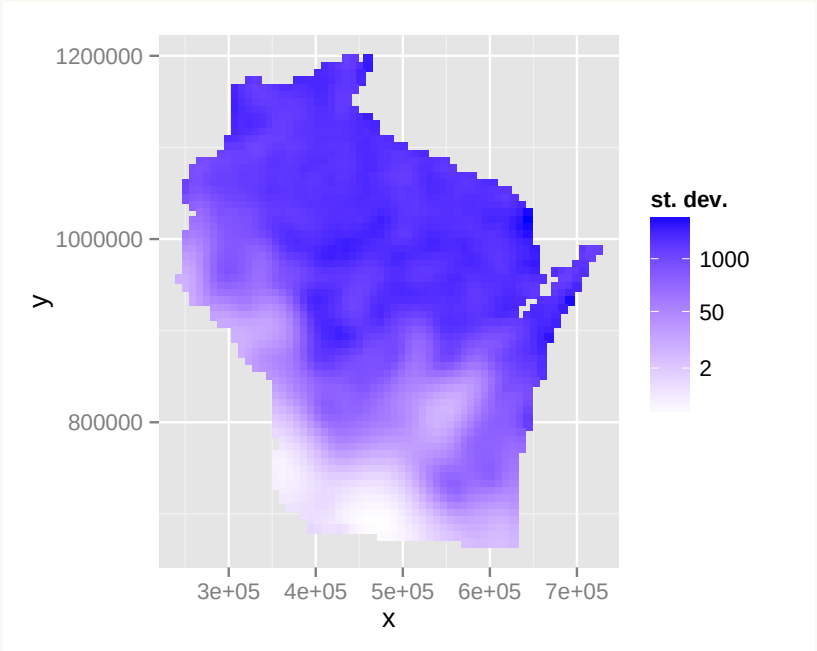
Figure: Cell means of draws for hardwood biomass for the upper midwest.

Results

Wisconsin softwood



(a) Mean

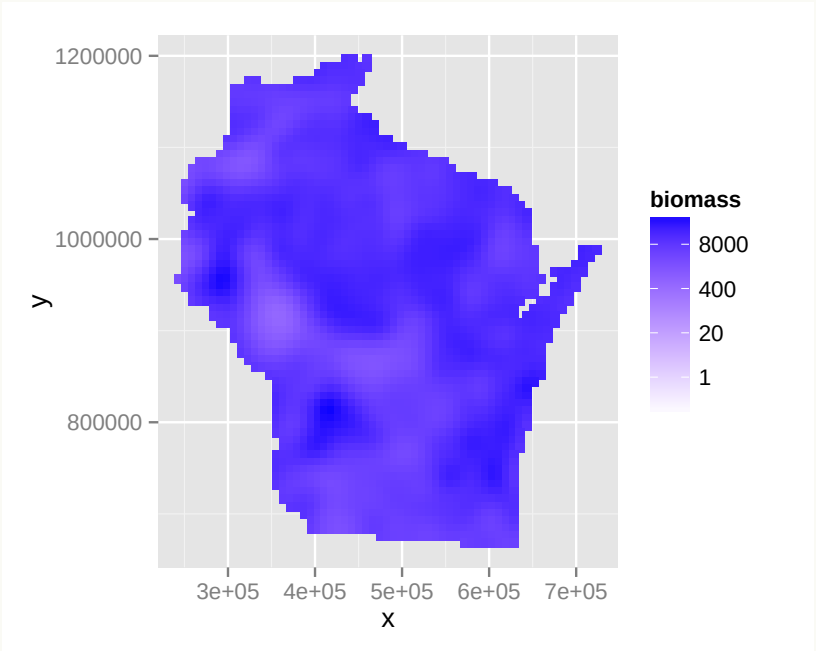


(b) Standard deviation

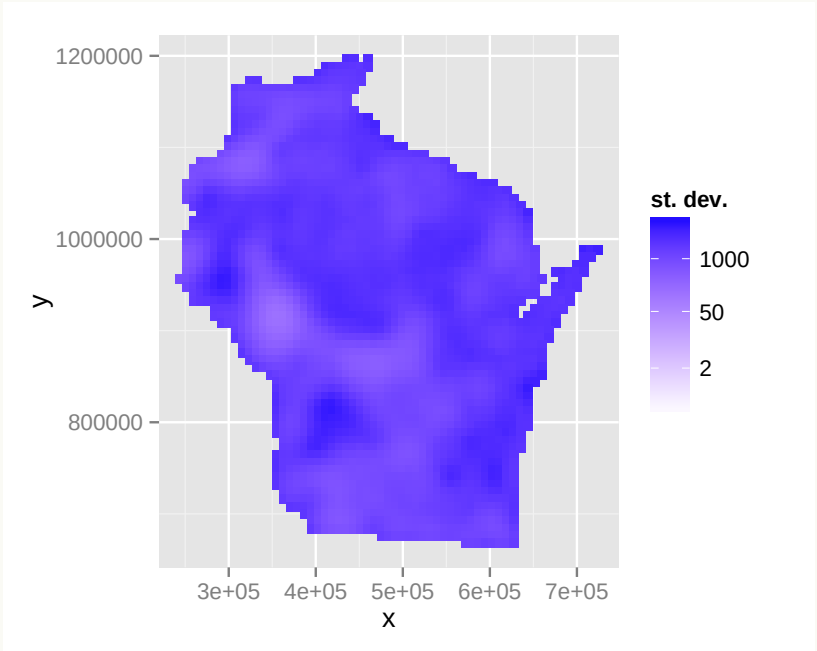
Figure: Mean and standard deviation of draws of per-cell biomass from the model for Wisconsin softwoods

Results

Wisconsin hardwood



(a) Mean

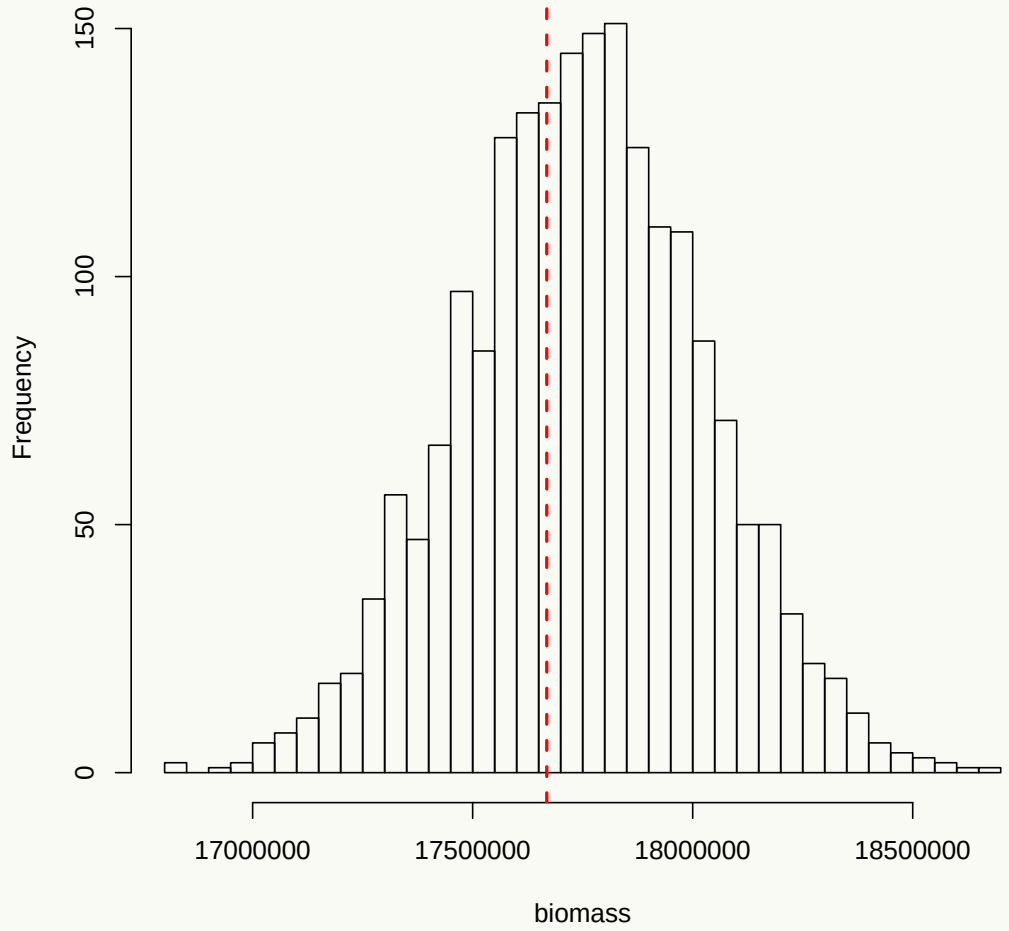


(b) Standard deviation

Figure: Mean and standard deviation of draws of per-cell biomass from the model for Wisconsin hardwoods

Results

Draws of total softwood biomass for Wisconsin - the vertical line represents the observed biomass.



Results

Draws of total hardwood biomass for Wisconsin - the vertical line represents the observed biomass.