

Modeling change in forest biomass across the eastern US

Erin M. Schliep¹ · Alan E. Gelfand¹ · James S. Clark^{1,2} · Kai Zhu^{2,3,4}

Received: 2 February 2015 / Revised: 7 June 2015 / Published online: 27 June 2015 © Springer Science+Business Media New York 2015

Abstract Predictions of above-ground biomass and the change in above-ground biomass require attachment of uncertainty due the range of reported predictions for forests. Because above-ground biomass is seldom measured, there have been no opportunities to obtain such uncertainty estimates. Standard methods involve applying an allometric equation to each individual tree on sample plots and summing the individual values. There is uncertainty in the allometry which leads to uncertainty in biomass at the tree level. Due to interdependence between competing trees, the uncertainty at the plot level that results from aggregating individual tree biomass in this way is expected to overestimate variability. That is, the variance at the plot level should be less than the sum of the individual variances. We offer a modeling strategy to learn about change in biomass at the plot level and model cumulative uncertainty to accommodate this dependence among neighboring trees. The plot-level variance is modeled using a parametric density-dependent asymptotic function. Plot-by-time covariate information is introduced to explain the change in biomass. These features are incorporated into a hierarchical model and inference is obtain within a Bayesian framework. We analyze data for the eastern United States from the Forest Inventory and Analysis (FIA) Program of the US Forest Service. This region contains roughly 25,000 FIA monitored

Handling Editor: Pierre Dutilleul.

- ¹ Department of Statistical Sciences, Duke University, Durham, NC, USA
- ² Nicholas School of the Environment, Duke University, Durham, NC, USA
- ³ Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
- ⁴ Department of Biology, Stanford University, Stanford, CA, USA

Erin M. Schliep erin.schliep@duke.edu

plots from which there are measurements of approximately 1 million trees spanning more than 200 tree species. Due to the high species richness in the FIA data, we combine species into plant functional types. We present predictions of biomass and change in biomass for two plant functional types.

Keywords Allometric equations · Bayesian hierarchical model · Cumulative uncertainty · Forest biomass

1 Introduction

Forests play an important role in the global carbon cycle (Pan et al. 2011). Given the range of predictions reported for carbon sinks in forests, estimates of associated uncertainty are critical. For eastern North America, predictions of the annual carbon sink range from 0.21 to 0.25 petagrams of carbon per year (Pg C/yr) (Pan et al. 2011). Unfortunately, there has been no way to assign statistical model-based uncertainty to these predictions because biomass per unit area is not directly measured. Instead, allometric equations are applied to the diameter (and sometimes height) of each tree on a plot to obtain predictions of above-ground biomass at the tree level, resulting in uncertainty in tree-level biomass. At the plot level, predictions of above-ground biomass are obtained by summing the independent tree-level predictions of biomass. Summing the tree-level variances to obtain variance in total biomass ignores the fact that trees interact. That is, due to the interdependence and crowding between competing trees, the sum of the variances will not equal the variance of the sum. One approach to this problem is to model the cumulative variance as density-dependent. The variance model should allow for dependence as a function of density in the form of diminishing returns; the plot-level variance is something less than the sum of the individual variances.

Understanding forest biomass change is essential for human society to cope with global climate change (Barford et al. 2001; Schimel et al. 2001; Wright 2005; Susan 2007). This has led to substantial literature providing modeling efforts for both biomass and the change in biomass. For example, McMahon et al. (2010) model biomass using the Monod function which describes the increase in biomass of forests during recovery. Their approach focuses on patterns of resource use and limitation and is a function of stand age and the age at half-saturation. Vayreda et al. (2012) use principal component analysis to model change in carbon, as well its components, growth rate and mortality rate. This approach has the disadvantage of making interpretation of model parameters difficult.

We offer a species-level modeling strategy to quantify forest biomass change that accounts for density and species differences as they vary geographically. Forests are made up of a mix of tree species with varying functional traits and growth patterns as a response to light, moisture, and nutrients. In the eastern United States, forests have experienced dramatic change due to human disturbance for centuries (MacCleery 1993). In the process of recovery, forests are changing their species compositions over time, a process referred to in forest ecology as *succession*. Thus, biomass change of eastern US forests depends heavily on successional status of species and management

practices across geography. To accurately quantify forest biomass change, we take into account species, environment, geography, shade tolerance, and seral stage.

Allometric equations are available for common tree species regionally and globally [e.g., Jenkins et al. (2003)]. Uncertainty at the tree level has been investigated in terms of measurement error, the specification in the allometric equation, the sampling protocol of stems in a plot, and the representativeness of small plots for a forest landscape (Chave et al. 2004, 2014). However, we are interested in above-ground biomass, referred to as biomass hereafter, at the plot level. Therefore, our estimates of uncertainty should also be at the plot level.

Our data come from the Forest Inventory and Analysis (FIA) Program of the US Forest Service and include species, size, and health of trees, as well as tree growth, mortality, and removal by harvest. We analyzed data obtained from the eastern United States. This region contains roughly 25,000 FIA monitored plots from which there are measurements of roughly 1 million trees spanning more than 200 tree species. Between 1997 and 2011, each plot in the region is surveyed twice. Due to the high species richness in the FIA data, we combine species into plant functional types (as described below) and model at that level.

To illustrate the nature of diminishing returns in the uncertainty of biomass at the plot level, we predict tree-level biomass for late successional hardwoods in the FIA data across the eastern US using species-specific allometric equations and associated errors. That is, for each late successional hardwood on the plot we predict biomass using species-specific allometric equations and the tree's diameter at breastheight. Using the estimates of uncertainty in the allometric equations, we repeat this for 10,000 iterations to obtain a distribution of predicted biomass for each tree. For each plot, we sum the tree-level predictions to obtain a distribution of predicted plotlevel biomass. We compute standard deviations of these samples for each plot to obtain estimates of uncertainty in plot-level biomass. Working with plots of size 0.067 ha, Fig. 1 (left) shows boxplots of the standard deviation of plot-level biomass is increasing as a function of the number of trees. When the plot has more than 20 trees, however, uncertainty saturates; additional trees do not add to the uncertainty



Fig. 1 Standard deviation (*left*) and coefficient of variation (*right*) of the distribution of predicted biomass at the plot level as a function of the number of trees

in plot-level biomass. Also shown in Fig. 1 (right) are boxplots of the coefficient of variation (CV) by plot density, where CV is computed as the ratio of the standard deviation to the mean. CV decreases as a function of the number of trees indicating, again, that the variance of biomass at the plot level reaches saturation. Chave et al. (2014) report a similar decrease in CV with increasing plot density for tropical and sub-tropical forests and woodland savannas. We argue that the CV decreases with the number of trees and it results from dependence among individuals.

We are interested in predicting biomass with uncertainty at the plot level and, therefore, we model at the plot level. Additionally, we want predictions of the change in biomass at the plot level. As a result of the FIA data collection protocol, small trees are measured only on a subset of a plot while larger trees are measured on the entire plot. This means that some trees will not have been included in a previous sample. For this reason, we model as separate responses plot-level biomass of *saplings* and *trees*. Due to the sparse surveying of plots, we model plot-level biomass statically at each survey time with change in biomass induced by differencing. Plot-by-time covariate information is introduced, as well as plot-level covariates, to explain change. A benefit of our model is that we have an explicit conditional distribution for the rate of change of biomass given current biomass for each plot and plant functional type. We propose a parametric functional specification for cumulative uncertainty at the plot level that results from aggregating individual-level biomass to plot-level biomass. All of these features are incorporated into a hierarchical model and we implement inference within a Bayesian framework.

The plan of the paper is as follows. In Sect. 2 we describe the FIA data, how and where they are collected and the species and plant functional types observed. Allometric equations for computing individual-level biomass are described in Sect. 3. We show how biomass is aggregated to the plot level where total biomass is defined as the summation of sapling biomass and tree biomass. We define Δ -biomass as the annual rate of change in total biomass. In Sect. 4, we outline the models for sapling and tree biomass. In Sect. 5, we apply the model to two plant functional types, late successional hardwoods and southern pines. The paper concludes with a summary and suggestions for future work in Sect. 6.

2 The FIA data

FIA applies a nationally consistent sampling protocol using a quasi-systematic design covering all ownerships across the United States resulting in national sample intensity of one plot per 2428 ha (Bechtold and Patterson 2005). Within the eastern US, the FIA surveys roughly 25,000 plots (Fig. 2, left). Data obtained for each plot include forest type, site attributes, tree species, tree size, and overall tree condition. We included only non-disturbed plots in this analysis. Between the years 1997 and 2011, each plot is surveyed twice and the time between surveys ranges from 1 to 12 years (Fig. 2, right). The inventory includes 218 species of trees. Due to the large number of different species and the rarity of some species across the region, we group species into 11 plant functional types (PFTs) (Dietze and Moorcroft



Fig. 2 Number of plots surveyed by the FIA across the eastern US (*left*) and years between FIA surveys across plots (*right*)

No.	PFT	Acronym	No. of plots
1.	Early successional hardwoods	ESH	16,364
2.	Evergreen hardwoods	Evergreen	1502
3.	Hydrics	Hydric	1213
4.	Late successional conifers	LSC	7627
5.	Late successional hardwoods	LSH	17,306
6.	Midsuccessional conifers	MC	3951
7.	Northern midsuccessional hardwoods	NMH	16,055
8.	Northern pines	NP	3531
9.	Southern midsuccessional hardwoods	SMH	10,810
10.	Southern pines	SP	4664
11.	Other		693

 Table 1
 Summary of the 23,259 FIA plots by plant functional type

2011). The PFTs are listed in Table 1 along with the number of plots each PFT was observed.

An FIA plot consists of four circular subplots arranged in the pattern shown in Fig. 3. The subplots each have a radius of 7.32 m and the distance between the subplot centroids is 35.58 m. Measurements are taken of all trees within subplots where a tree is classified as an individual with diameter greater than 12.7 cm. All saplings, referring to individuals with diameters less than or equal to 12.7 cm, are measured on the four microplots, each of which is a subset of a subplot. The radius of each micro plot is 2.07 m. The total microplot area is 53.85 m^2 and the total subplot area is 673.34 m^2 . Ecologically, the distance between plots in the FIA data is too large to adopt spatial dependence between plots. Instead, we capture heterogeneity across plots using random plot effects.

Fig. 3 FIA sampling scheme for subplots (*grey*) and microplots (*black*) on an FIA plot



3 Individual and plot-level biomass

3.1 Allometric equations for individual biomass

Biomass for each individual is computed using an allometric regression equation that converts diameter to above-ground biomass. The allometric equation proposed by Jenkins et al. (2003) is

$$\log(biomass) = \beta_0 + \beta_1 \log(dbh) \tag{1}$$

where *dbh* is diameter at breast-height (cm) and biomass is measured in kilograms (kg). The parameter values β_0 and β_1 are species-specific. There are numerous studies on biomass equations for different species and regions [e.g., Jenkins et al. 2003; Brown et al. 1999; Marklund 1988; Zianis et al. 2005]. With the exception of Chave et al. (2004), Wutzler et al. (2008), and Stephenson et al. (2014), however, there is much less in the literature about their uncertainty due the destructive sampling of trees that is required. Chave et al. (2004) investigated the error associated with the allometric equations in predicting biomass for tropical forests and reported the choice of allometric equation contributed to error greater than 20% of the above-ground biomass Stephenson et al. (2014) found that Eq. (1) has a tendency to overpredict biomass for larger trees for nine species from the temperate western USA. Using a similar biometric equation, Wutzler et al. (2008) found confidence intervals for biomass to be narrow where the coefficient of variation (CV) was 0.12 for an individual at an average stand.

3.2 Sapling and tree plot-level biomass

Using (1) we obtain predictions of biomass (kg) at the individual level for each plot at two survey times. Let *s* be an indicator for survey where s = 1 for the first survey and s = 2 for the second survey. Let Y_{ijs}^k denote the biomass (kg) of the *j*th individual at plot *i* of PFT *k* at survey *s*. Due to the different sampling intensities within the plot (i.e. subplot, microplot), we model total biomass in two components, (1) biomass of saplings and (2) biomass of trees. Let d_{ijs}^k denote the diameter of the individual.

Individuals with $d_{ijs}^k \le 12.7 \text{ cm}$ are classified as saplings at survey *s* and those with $d_{ijs}^k > 12.7 \text{ cm}$ are classified as trees at survey *s*. For plot *i* and PFT *k* we define sapling biomass for survey 1 as

$$b_{i1}^{k} = \sum_{j} Y_{ij1}^{k} \operatorname{I}_{[d_{ij1}^{k} \le 12.7]}$$

and at survey 2 as

$$b_{i2}^{k} = \sum_{j} Y_{ij2}^{k} \operatorname{I}_{[d_{ij2}^{k} \le 12.7]}$$

where b_{is}^k is in terms of kg per four microplots within the FIA plot for all *i*, *s*, and *k*. We denote total biomass of trees at plot *i* of PFT *k* for survey 1 as

$$B_{i1}^{k} = \sum_{j} Y_{ij1}^{k} \operatorname{I}_{[d_{ij1}^{k} > 12.7]}$$

and at survey 2 as

$$B_{i2}^{k} = \sum_{j} Y_{ij2}^{k} \operatorname{I}_{[d_{ij2}^{k} > 12.7]}$$

where B_{is}^k is in terms of kg per four subplots within the FIA plot for all *i*, *s*, and *k*. We consider b_{is}^k and B_{is}^k as "noisy" total sapling and tree biomass, respectively, due to the error in the predictions of each Y_{ijs}^k resulting from the allometry. That is, total sapling and tree biomass is *derived*; it is never observed. However, with interest in understanding the behavior of total biomass, below, we model the *b*'s and *B*'s.

Figure 4 gives histograms of the number of late successional hardwoods (LSH) classified as saplings (left) and trees (right) observed at each plot during the first survey. Also in this figure are histograms of total sapling and total tree biomass across plots for LSH.

Let TB_{is}^k denote total biomass at plot *i* and survey *s* of PFT *k*. Total biomass is the summation of sapling and tree biomass computed as

$$TB_{is}^{k} = \frac{b_{is}^{k}}{A_{s}} + \frac{B_{is}^{k}}{A_{t}}$$
⁽²⁾

where A_s and A_t are the total area (ha) of the four microplots where all saplings are observed and four subplots where all trees are observed, respectively. Therefore, TB_{is}^k is in terms of kg per hectare (kg/ha). The rate of change in biomass, given in kg per hectare per year (kg/ha/yr), referred to as Δ -biomass, for plot *i* and PFT *k* is

$$\Delta_i^k = \frac{T B_{i2}^k - T B_{i1}^k}{t_{i2} - t_{i1}} \tag{3}$$

where $t_{i2} - t_{i1}$ is the time between the two surveys for plot *i*. This assumes a constant rate of change in biomass between t_{i1} and t_{i2} as opposed to what would result from

29



Fig. 4 Histograms of the number of individuals (*top*) and total biomass (*bottom*) of saplings (*left*) and trees (*right*) for LSH for survey 1. Total biomass is given in kg per four microplots (saplings) and subplots (trees) on the FIA plots

rapid or slow onset disturbance. However, we are limited by having only two times points; more complex modeling for the rate of change would require additional surveys.

4 The model

We define models for noisy total sapling and tree biomass in a hierarchical framework. Noisy total sapling biomass, henceforth referred to as sapling biomass, is modeled as

$$b_{is}^{k} = n_{is}^{k} \mu_{is}^{k} + g(n_{is}^{k}; \phi_{b^{k}}) \epsilon_{is}^{k}$$
(4)

where n_{is}^k is the number of saplings, μ_{is}^k is the average sapling biomass, and $g(n_{is}^k; \phi_{b^k}) \epsilon_{is}^k$ is first-stage measurement error. We assume ϵ_{is}^k is independent error with variance $\sigma_{b^k}^2$ for all *i*, *s*, and *k*, and $g(n_{is}^k; \phi_{b^k})$ is a function of n_{is}^k and a para-

meter ϕ_{b^k} . Similarly, we model noisy total tree biomass, henceforth referred to as tree biomass, as

$$B_{is}^{k} = N_{is}^{k} \theta_{is}^{k} + g(N_{is}^{k}; \phi_{B^{k}}) \eta_{is}^{k}$$
(5)

where N_{is}^k is the number of trees, θ_{is}^k is the average tree biomass, and $g(N_{is}^k; \phi_{B^k})\eta_{is}^k$ is first-stage measurement error where η_{is}^k is independent error with variance $\sigma_{B^k}^2$. Again, $g(N_{is}^k; \phi_{B^k})$ is a function of N_{is}^k and a parameter ϕ_{B^k} .

We assume the function $g(m; \phi)$ is an exponential asymptote function with parameter ϕ that is bounded above where

$$g(m;\phi) = \sqrt{1 - \exp^{-m/\phi}}.$$
(6)

This parametric functional form was motivated by Fig. 1 to govern cumulative uncertainty at the plot level that results from aggregating from individual-level biomass to plot-level biomass. It assumes that measurement error increases as the number of individuals on the plot increases while restricting total plot-level measurement error. Here, ϕ_{b^k} denotes the range parameter controlling the asymptote of sapling biomass measurement error and ϕ_{B^k} denotes the range parameter controlling the asymptote of tree biomass measurement error.

Next, we model average sapling and tree biomass as incorporated in (4) and (5). To ensure that average biomass is non-negative, we model both μ_{is}^k and θ_{is}^k using a tobit model with latent random variables $\tilde{\mu}_{is}^k$ and $\tilde{\theta}_{is}^k$, respectively. That is,

$$\mu_{is}^{k} = \begin{cases} \widetilde{\mu}_{is}^{k} & \widetilde{\mu}_{is}^{k} > 0\\ 0 & \widetilde{\mu}_{is}^{k} \le 0 \end{cases}$$
(7)

and

$$\theta_{is}^{k} = \begin{cases} \widetilde{\theta}_{is}^{k} & \widetilde{\theta}_{is}^{k} > 0\\ 0 & \widetilde{\theta}_{is}^{k} \le 0. \end{cases}$$

$$\tag{8}$$

In general, the tobit model is a more natural model than say, a log-normal, if we expect many average sapling or tree biomasses to be small, and thus, don't think the left tail of the density should go to 0 at 0.

Both latent average biomass variables are specified through a linear mixed model with normally distributed error. For PFT k, latent average sapling biomass is

$$\widetilde{\mu}_{is}^{k} = W_{is}^{\prime} \boldsymbol{\alpha}_{1}^{k} + \lambda_{i}^{k}$$

$$\lambda_{i}^{k} = X_{i}^{\prime} \boldsymbol{\alpha}_{2}^{k} + \nu_{i}^{k}$$
(9)

where W_{is} is a vector of survey year-specific covariates for plot *i*, α_1^k is a vector of coefficients, and λ_i^k is a plot random effect. The plot random effect λ_i^k is centered with mean $X_i'\alpha_2^k$ where X_i is a vector of covariates for plot *i* and α_2^k is a vector of coefficients. Lastly, v_i^k is normally distributed random noise with mean 0 and variance $\tau_{\lambda_k}^2$. The multilevel structure of the model is specified in the form of hierarchical

centering¹ which better identifies the model parameters and leads to better behaved Markov chain Monte Carlo (MCMC) model fitting (Gelfand et al. 1995). Similarly, latent average tree biomass is modeled as

$$\widetilde{\theta}_{is}^{k} = \boldsymbol{W}_{is}^{\prime} \boldsymbol{\beta}_{1}^{k} + \gamma_{i}^{k}
\gamma_{i}^{k} = \boldsymbol{X}_{i}^{\prime} \boldsymbol{\beta}_{2}^{k} + \zeta_{i}^{k}.$$
(10)

Here, $\boldsymbol{\beta}_1^k$ and $\boldsymbol{\beta}_2^k$ are vectors of coefficients, γ_i^k is a plot random effect, and ζ_i^k is normally distributed random noise with mean zero and variance $\tau_{\gamma_k}^2$.

Given the model parameters, we have explicit distributions of both TB_{i1}^k and TB_{i2}^k , as well as $\Delta_i^k | TB_{i1}^k$ using (2) and (3) (See "Appendix"). From (4) and (5), total biomass might possibly be less than 0, in which case we set it to 0. Using these distributions and composition sampling we are able to obtain draws from the posterior predictive distributions at the plot level for total biomass and Δ -biomass.

We assign prior distributions to the model parameters defined in (4), (5), (9), and (10) above as follows. At the data level, we assign diffuse, conjugate inverse-Gamma distributions to the variance parameters, $\sigma_{b^k}^2$ and $\sigma_{B^k}^2$. The range parameters, ϕ_{b^k} and ϕ_{B^k} , controlling the exponential asymptote functions are assigned truncated-normal distributions constrained to be positive. We assign mean zero multivariate normal distributions to the coefficient vectors α_1^k and β_1^k . These priors are not conjugate because both true average sapling and tree biomass parameters, μ_{is}^k and θ_{il}^k , are nonnegative as defined by the tobit model in (7) and (8). Thus, sampling of these parameters requires Metropolis-Hastings steps and the details of these algorithms are given in "Appendix".

The coefficients α_2 and β_2 are assigned noninformative conjugate multivariate normal distributions. We include plot-level random effects, v_i^k and ζ_i^k , to capture any remaining heterogeneity in sapling and tree biomass across plots beyond that being explained by the covariates. The random noise parameters v_i^k and ζ_i^k are assumed i.i.d. normal random variables with mean 0 and variances τ_{λ}^k and τ_{γ}^k , respectively, where τ_{λ}^k and τ_{γ}^k have conjugate inverse-Gamma distributions.

5 Application: Modeling biomass in the eastern US

We model biomass for two plant functional types: late successional hardwood (LSH) and southern pine (SP). There are 17,306 plots that contain at least one LSH and 4664 plots that contain at least one SP. The plots are mixed stands such that other PFTs may also be present. LSH are found throughout the majority of the eastern US while SP are concentrated in the southeast. Again, each plot is surveyed twice according to the

¹ Hierarchical centering is a reparameterization technique for models with multiple levels of random effects. It can be applied in its simplest form in the context of a standard ANOVA model where population means are often expressed as a global mean and a population level deviation, e.g., $\mu + \alpha_i$. Here, the data will well-identify the sum but not as well the components. Hierarchical centering entails reparameterizing from μ and α_i [with prior $\pi(\mu)\pi(\alpha_i)$] to $\eta_i = \mu + \alpha_i$ and μ and specifying the prior as $\pi(\eta_i|\mu)\pi(\mu)$, i.e., *centering* the η_i hierarchically.

sampling scheme outlined in Sect. 2. We use 80% of the plots of each plant functional type to fit the model and hold out the remaining 20% to do out-of-sample prediction for model evaluation.

As shown in Fig. 4, the average total biomass of LSH saplings per four microplots is 36.6 kg and average total biomass of trees per four subplots is 2340 kg. The number of LSH saplings observed on the four microplots ranges from 1 to 45 with a median of 3. The number of LSH trees observed on the four subplots ranges from 1 to 64 with a median of 6. For SP, the average total biomass of saplings is 27.5 kg and average total biomass of trees is 2318 kg. The total number of SP saplings ranges from 1 to 105 with a median of 7 (Figures not included).

The survey year-specific covariates in each model include tree density and stand age of the plot. Tree density is the total number of trees of all species observed in the plot at the time of the survey. Other covariates in the model include temperature and precipitation, both of which are centered and scaled, and indicator variables for physiographic class code of available moisture in the soil. The three classes of moisture availability are xeric, mesic, and hydric where xeric is low or deficient, mesic is moderate and used as the base level, and hydric is abundant. These covariates are not survey year-specific.

The prior distributions assigned to the parameters $\sigma_{b^k}^2$ and ϕ_{b^k} of the measurement error variance of sapling biomass are inverse-gamma and truncated normal, respectively, where $\sigma_{b^k}^2 \sim IG(3, 10^4)$ and $\phi_{b^k} \sim TN(0, 40, 0, \infty)$. The tree biomass measurement error parameters are assigned $\sigma_{B^k}^2 \sim IG(3, 10^5)$ and $\phi_{B^k} \sim$ $TN(0, 40, 0, \infty)$. The priors for $\sigma_{b_k}^2$ and $\sigma_{B_k}^2$ were both chosen to be diffuse. Since tree biomass is larger than sapling biomass, we assume measurement error may also be larger. Thus, the median of the distribution of $\sigma_{B^k}^2$ is greater than the median of the distribution of $\sigma_{b_k}^2$. The priors for ϕ_{b_k} and ϕ_{B_k} were chosen such that the distributions spanned the observed $n_{i_s}^k$ and $N_{i_s}^k$.

The coefficient vectors $\boldsymbol{\alpha}_1^k$ and $\boldsymbol{\beta}_1^k$ have mean zero multivariate normal prior distributions with variance $10^6 \boldsymbol{I}_{p_1 \times p_1}$ where $p_1 = 2$. The coefficient vectors $\boldsymbol{\alpha}_2^k$ and $\boldsymbol{\beta}_2^k$ each contain an intercept term and are assigned mean zero multivariate normal prior distributions with variance $10^6 \boldsymbol{I}_{p_2 \times p_2}$ where $p_2 = 5$. Lastly, the variances $\tau_{\lambda^k}^2$ and $\tau_{\gamma^k}^2$ are assigned IG(4, 4) prior distributions.

The model is fitted using R software (R Development Core Team 2007) running on an Intel Core i7 processor with Scientific Linux 6.4. We use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution of the parameters given the data. The MCMC algorithm was contrived specifically for this application and some explicit details are included in 1. We run MCMC for 100,000 iterations. Convergence was assessed by computing the Gelman and Rubin R statistic for each of the model parameters using three chains with varying starting values. The upper 97.5% bound on the statistic was less than 1.10 for each of the parameters, indicating no issues with convergence. We disregard the first 50,000 samples as burn-in and retain every 10th iteration for inference. Posterior median estimates and 95% credible intervals are given in Table 2.

34

Parameter	Late successional hardwoods	Southern pines
α_{11} (Tree density)	0.14 (0.12, 0.17)	0.23 (0.21, 0.24)
α_{12} (Stand age)	0.07 (0.06, 0.09)	-0.05 (-0.07, -0.03)
α_{20} (Intercept)	6.28 (5.53, 7.00)	11.56 (10.78, 12.36)
α_{21} (Temperature)	-1.58 (-1.99, -1.19)	0.56 (0.05, 1.11)
α_{22} (Precipitation)	0.31 (-0.08, 0.69)	0.05 (-0.40, 0.49)
α_{23} (Xeric)	-1.12 (-2.03, -0.16)	2.13 (0.83, 3.41)
α ₂₄ (Hydric)	1.13 (-0.17, 2.48)	3.09 (0.39, 6.02)
β_{11} (Tree density)	-3.30 (-3.51, -3.10)	-0.29 (-0.38, -0.21)
β_{12} (Stand age)	3.36 (3.26, 3.46)	6.41 (6.19, 6.61)
β_{20} (Intercept)	158.6 (150.9, 168.5)	-6.42 (-17.62, 5.67)
β_{21} (Temperature)	-14.46(-21.52, -7.83)	56.77 (47.96, 65.85)
β_{22} (Precipitation)	-4.81 (-11.40, 1.71)	-12.93 (-21.15, -4.75)
β_{23} (Xeric)	-131.2 (-149.3, -112.0)	-114.1 (-137.0, -91.62)
β_{24} (Hydric)	-12.72 (-34.03, 8.81)	-2.85 (-40.74, 36.02)

Table 2 The posterior medians and 95% credible intervals for parameters of average sapling and tree biomass

Bold indicates credible intervals not containing 0

The coefficient for tree density in average tree biomass is negative for both PFTs indicating that plots with high tree density have lower average tree biomass. Interestingly, however, the same coefficient for average sapling biomass is positive. This is because plots with high tree density tend to have fewer saplings, resulting in larger values of average sapling biomass. Both average sapling and tree biomass for LSH are increasing between the first and second survey indicated by positive coefficients for stand age. Average SP tree biomass is increasing between the first and second survey but average sapling biomass for SP is decreasing. Average LSH sapling and tree biomass increases with temperature while SP sapling and tree biomass increases with temperature. Additionally, high soil moisture tends to increase average sapling biomass.

We compute estimates of the measurement error variance of sapling biomass and tree biomass using posterior estimates of σ_{bk}^2 , ϕ_{bk} , σ_{Bk}^2 , and ϕ_{Bk} for LSH and SP. We plot posterior median estimates and 95% credible intervals of the variance as a function of the number of saplings and trees in Fig. 5. The histograms in each figure are of the the number of saplings (top) and trees (bottom) for each of the PFTs. The measurement error of sapling biomass is $\sigma_{bk}^2(1 - \exp^{-n_{is}^k/\phi_{bk}})$ and the measurement error variance for both sapling and tree biomass for LSH asymptotes within the range of the number of individuals per plot that we observed. Tree biomass reaches an upper bound for SP but sapling biomass does not. This is likely due to the majority of the plots having fewer than 30 saplings and two plots in the region being outliers with more than 60. Changes in the range parameter ϕ_{bk} can inflate the variance which explains the large values of posterior variability seen in the measurement error for sapling biomass of SP.



Fig. 5 Posterior median estimates and 95% credible intervals (kg^2) of measurement error variances for sapling (top) and tree (*bottom*) biomass for late successional hardwoods (*left*) and southern pines (*right*) as a function of the number of individuals. The histograms in each figure are of the number of individuals on the plot

 Table 3
 In-sample RMSE for total sapling and tree biomass (kg) at surveys 1 and 2 for late successional hardwoods and southern pines

Model	b_1	<i>b</i> ₂	<i>B</i> ₁	<i>B</i> ₂
Late successional hardwoods	15.66	15.07	266.02	266.73
Southern pines	24.72	23.46	240.02	236.19

In-sample root mean square error (RMSE) is reported in Table 3 for sapling and tree biomass at each survey time for both PFTs. Samples from the posterior distribution of b_{is}^k and B_{is}^k are obtained using posterior draws of the model parameters. The values are similar for both survey times and functional type. We obtain posterior samples of Δ_i^k using (2) and (3) and posterior draws of b_{is}^k and B_{is}^k .

Posterior median predictions of change in sapling and tree biomass per year are also shown spatially for both LSH and SP in Fig. 6. We see that sapling biomass for LSH is increasing in North Carolina, while changes in the rest of the eastern US are small. The



Fig. 6 Posterior median predictions of the change in biomass (kg/ha/yr) of saplings (*top*) and trees (*middle*) for late successional hardwoods (*left*) and southern pines (*right*). The *bottom panel* give posterior median predictions of Δ -biomass (kg/ha/yr) for the two PFTs





Fig. 7 Standard deviations of the posterior distributions of Δ -biomass (kg/h/yr) for late successional hardwoods (*left*) and southern pines (*right*)

mild increases in tree biomass across the eastern US result from growth and is driving the positive predictions of Δ -biomass in the bottom panel of the figure. Decreases in sapling biomass through time for SP is predicted in much of the southern states except for regions in North Carolina and Alabama. Nearly all plots are seeing increases in tree biomass of SP, many of which are large in comparison to the growth of LSH. This is, in part, due to the rapid growth rate of SP and their high tree density in the region.

Posterior median predictions of Δ -biomass (kg/ha/yr) are shown in the bottom panel of Fig. 6 for both functional types. Biomass is increasing in the upper midwest and northeast, as well as North Carolina for LSH. Biomass is increasing throughout the entire southeast for SP. The predominant contributor to positive predictions of Δ -biomass for SP is the increase in tree biomass, or growth. Decreases in biomass appear to be very localized events as there are no regions reporting clusters of negative Δ -biomass. Large decreases in biomass are often the result of the mortality of a large tree. One of the benefits of our model is that we also obtain estimates of uncertainty for Δ -biomass. Show in Fig. 7 is the standard deviation of our posterior distribution of Δ -biomass across space for both PFTs.

We predict biomass at hold-out plots for both PFTs. Table 4 reports 90% empirical coverage probabilities for total sapling and tree biomass for both survey times. Each of the coverage probabilities is slightly greater than the nominal level indicating that our prediction intervals for total sapling and tree biomass are conservative. We also compute out-of-sample root mean square prediction errors (RMSPE) for the hold-out plots (Table 5). RMSPE values are similar between the two survey times and PFTs.

6 Discussion

The interdependence between trees introduces challenges in estimating the uncertainty of biomass at the plot level. Due to this dependence, the variance of plot-level biomass

Model	<i>b</i> ₁	<i>b</i> ₂	<i>B</i> ₁	<i>B</i> ₂
Late successional hardwoods	90.49	93.09	96.19	95.90
Southern pines	92.88	94.84	94.37	93.64

Table 4 Out-of-sample 90% empirical coverage probabilities

 Table 5
 RMSPE for total sapling and tree biomass (kg) at surveys 1 and 2 for late successional hardwoods and southern pines

Model	b_1	b_2	<i>B</i> ₁	<i>B</i> ₂
Late successional hardwoods	46.55	45.65	1753.36	1885.80
Southern pines	52.15	47.15	1877.32	2381.18

should be less than the sum of the variances of the individual trees. Therefore, we propose a parametric density-dependent asymptotic functional form for the plot-level variance of biomass as motivated by our illustration.

We model biomass for each plant functional type by first defining total biomass as the summation of sapling and tree biomass. Sapling and tree biomass are modeled in terms of the average biomass of an individual in each size class. Modeling in terms of average biomass is advantageous due to the challenges of directly measuring change in biomass. We model sapling and tree biomass separately due to the unequal sampling intensities between saplings and trees. The model is defined specifically for the sampling scheme of the FIA data in the eastern US and would therefore need to be modified to accommodate other sampling schemes. Predictions of the rate of change in biomass, Δ -biomass, are computed as the difference between total biomass at the two surveys divided by the time between surveys. We applied the model to two plant functional types, late successional hardwoods and southern pines.

Ongoing work includes scaling biomass predictions at the plot level to larger spatial regions. Both regional and global predictions of biomass are of interest to management as they assess the sustainability of biomass. Additionally, we plan to model more plant functional types. In fact, if we model all the functional types we will be able to consider the behavior of and change in total biomass across the region. The modeling challenge then becomes to incorporate suitable dependence between PFT's as we sum across them in order to obtain appropriate estimates of uncertainty.

Further challenges include projecting biomass change under varying covariate scenarios to address the carbon cycling issue raised at the outset of the paper. Also, we may be able to avail ourselves of additional data sources (e.g., National Ecological Observatory Network (NEON) products) presenting the opportunity to implement data fusion to enhance our understanding of the process.

Acknowledgments This research was supported by the National Science Foundation under grant numbers EF-1137364 and CDI-0940671 and the Coweeta LTER. The authors would also like to thank Bradley Tomasek for providing useful discussion on biomass allometry.

Appendix

Implicit distributions of total biomass and Δ -biomass

The model implies distributions relating to TB_{i1}^k , TB_{i2}^k and Δ_i^k . Let $\Theta_s^k = \{\mu_{is}^k, \theta_{is}^k, \sigma_{b^k}^2, \phi_{b^k}, \sigma_{B^k}^2, \phi_{B^k}\}$ for s = 1, 2. Then

(a)
$$TB_{i1}^{k}|\Theta_{1}^{k} \sim N\left(\frac{1}{A_{s}}n_{i1}^{k}\mu_{i1}^{k} + \frac{1}{A_{t}}N_{i1}^{k}\theta_{i1}^{k}, \frac{1}{A_{s}^{2}}\sigma_{b^{k}}^{2}g(n_{i1}^{k};\phi_{b^{k}})^{2} + \frac{1}{A_{t}^{2}}\sigma_{B^{k}}^{2}g(N_{i1}^{k},\phi_{B^{k}})^{2}\right)$$

(b) $TB_{i2}^{k}|\Theta_{2}^{k} \sim N\left(\frac{1}{A_{s}}n_{i2}^{k}\mu_{i2}^{k} + \frac{1}{A_{t}}N_{i2}^{k}\theta_{i2}^{k}, \frac{1}{A_{s}^{2}}\sigma_{b^{k}}^{2}g(n_{i2}^{k};\phi_{b^{k}})^{2} + \frac{1}{A_{t}^{2}}\sigma_{B^{k}}^{2}g(N_{i2}^{k},\phi_{B^{k}})^{2}\right).$

Additionally, TB_{i2}^k and TB_{i2}^k are conditionally independent given Θ_1^k and Θ_2^k . Furthermore, since $\Delta_i^k = \frac{TB_{i2}^k - TB_{i1}^k}{t_{i2} - t_{i1}}$,

$$\begin{array}{l} \text{(c)} \ \ \Delta_{i}^{k} | \Theta_{1}^{k}, \ \Theta_{2}^{k} \sim N\left(\frac{1}{t_{i2}-t_{i1}}\left(\frac{1}{A_{s}}n_{i2}^{k}\mu_{i2}^{k}+\frac{1}{A_{t}}N_{i2}^{k}\theta_{i2}^{k}-\frac{1}{A_{s}}n_{i1}^{k}\mu_{i1}^{k}-\frac{1}{A_{t}}N_{i1}^{k}\theta_{i1}^{k}\right), \\ & \quad \frac{1}{(t_{i2}-t_{i1})^{2}}\left(\frac{1}{A_{s}^{2}}\sigma_{b^{k}}^{2}(g(n_{i1}^{k};\phi_{b^{k}})^{2}+g(n_{i2}^{k};\phi_{b^{k}})^{2}\right) \\ & \quad +\frac{1}{A_{t}^{2}}\sigma_{B^{k}}^{2}\left(g(N_{i1}^{k},\phi_{B^{k}})^{2}+g(N_{i1}^{k},\phi_{B^{k}})^{2}\right)\right) \\ \text{(d)} \ \ \Delta_{i}^{k} | TB_{i1}^{k}, \ \Theta_{1}^{k}, \ \Theta_{2}^{k} \sim N\left(\frac{1}{t_{i2}-t_{i1}}\left(\frac{1}{A_{s}}n_{i2}^{k}\mu_{i2}^{k}+\frac{1}{A_{t}}N_{i2}^{k}\theta_{i2}^{k}-TB_{i1}^{k}\right), \\ & \quad \frac{1}{(t_{i2}-t_{i1})^{2}}\left(\frac{1}{A_{s}^{2}}\sigma_{b^{k}}^{2}g(n_{i2}^{k};\phi_{b^{k}})^{2}+\frac{1}{A_{t}^{2}}\sigma_{B^{k}}^{2}g(N_{i2}^{k},\phi_{B^{k}})^{2}\right)\right). \end{array}$$

Therefore, given the model parameters, we have an explicit distribution for $\Delta_i^k | T B_{i1}^k$. These distributions can be used in conjunction with prior samples through composition sampling to obtain samples from the posterior predictive distributions at the plot level for total biomass and Δ -biomass.

MCMC algorithm

The tobit latent variable approach to modeling average sapling and tree biomass requires Metropolis-Hastings algorithms for iterative sampling of α_1^k , β_1^k , λ^k , and γ^k . As a proposal distribution for each of these parameters, we use the full conditional distribution of the parameter under the assumption that $\mu_{is}^k = \tilde{\mu}_{is}^k$ and $\theta_{is}^k = \tilde{\theta}_{is}^k$ (i.e., dropping the tobit models (7) and (8) for average sapling and tree biomass, respectively). For simplicity, let $p_1 = 1$. Dropping k for ease of notation, the full conditional distribution of α_1 is

$$p(\alpha_1|\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{\lambda}, \boldsymbol{\beta}_2, \tau_{\boldsymbol{\lambda}}^2, \sigma_b^2, \phi_b) \propto p(\boldsymbol{b}_1, \boldsymbol{b}_2|\alpha_1, \boldsymbol{\lambda}, \sigma_b^2, \phi_b) p(\alpha_1)$$
$$\propto p(\boldsymbol{b}_1|\alpha_1, \boldsymbol{\lambda}\sigma_b^2, \phi_b) p(\boldsymbol{b}_2|\alpha_1, \boldsymbol{\lambda}, \sigma_b^2, \phi_b) p(\alpha_1)$$

Letting $\mu_{is}^k = \tilde{\mu}_{is}^k$, this conditional distribution is normal with known mean and variance. We propose a candidate value from this distribution, denoted α_1^* . Letting α_1^c denote the current value of α_1 , we compute $\tilde{\mu}_{is}^* = W_{is}' \alpha_1^* + \lambda_i$, $\tilde{\mu}_{is}^c = W_{is}' \alpha_1^c + \lambda_i$,

Deringer

 $(\mu_{is})^* = \max(0, \tilde{\mu}_{is}^*)$, and $\mu_{is}^c = \max(0, \tilde{\mu}_{is}^c)$ for each plot *i* and survey *s*. The candidate value is accepted with probability

$$\min\left\{\frac{p(\boldsymbol{b}_{1}|\boldsymbol{\mu}_{1}^{*},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{b}_{2}|\boldsymbol{\mu}_{2}^{*},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{\beta}^{*}) p(\boldsymbol{b}_{1}|\boldsymbol{\tilde{\mu}}_{1}^{c},\sigma_{bk}^{2},\phi_{b}) p(\boldsymbol{b}_{2}|\boldsymbol{\tilde{\mu}}_{2}^{c},\sigma_{bk}^{2},\phi_{b}) p(\boldsymbol{\beta}^{c})}{p(\boldsymbol{b}_{1}|\boldsymbol{\mu}_{1}^{c},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{b}_{2}|\boldsymbol{\mu}_{2}^{c},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{\beta}^{c}) p(\boldsymbol{b}_{1}|\boldsymbol{\tilde{\mu}}_{1}^{*},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{b}_{2}|\boldsymbol{\tilde{\mu}}_{2}^{*},\sigma_{b}^{2},\phi_{b}) p(\boldsymbol{\beta}^{c})},1\right\}$$

which reduces to

$$\min\left\{\frac{p(\boldsymbol{b}_{1}|\boldsymbol{\mu}_{1}^{*},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{2}|\boldsymbol{\mu}_{2}^{*},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{1}|\boldsymbol{\tilde{\mu}}_{1}^{c},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{2}|\boldsymbol{\tilde{\mu}}_{2}^{c},\sigma_{b^{k}}^{2},\phi_{b^{k}})}{p(\boldsymbol{b}_{1}|\boldsymbol{\mu}_{1}^{c},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{2}|\boldsymbol{\mu}_{2}^{c},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{1}|\boldsymbol{\tilde{\mu}}_{1}^{*},\sigma_{b^{k}}^{2},\phi_{b^{k}}) p(\boldsymbol{b}_{2}|\boldsymbol{\tilde{\mu}}_{2}^{*},\sigma_{b^{k}}^{2},\phi_{b^{k}})},1\right\}.$$
(11)

Note that when all values of the latent variable μ_{is}^* are greater than or equal to zero, the candidate value will be accepted with probability 1. Similar proposal distributions and algorithms are employed for β_1 , λ , and γ .

References

- Barford CC, Wofsy SC, Goulden ML, Munger JW, Pyle EH, Urbanski SP, Hutyra L, Saleska SR, Fitzjarrald D, Moore K (2001) Factors controlling long-and short-term sequestration of atmospheric CO2 in a mid-latitude forest. Science 294(5547):1688–1691
- Bechtold WA and Patterson PL (2005) The enhanced forest inventory and analysis program: national sampling design and estimation procedures. Technical report, US Department of Agriculture Forest Service, Southern Research Station Asheville, North Carolina
- Brown SL, Schroeder P, Kern JS (1999) Spatial distribution of biomass in forests of the eastern USA. For Ecol Manag 123(1):81–90
- Chave J, Condit R, Aguilar S, Hernandez A, Lao S, Perez R (2004) Error propagation and scaling for tropical forest biomass estimates. Philos Trans R Soc Lond Ser B Biol Sci 359(1443):409–420
- Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan MS, Delitti WBC, Duque A, Eid T, Fearnside PM, Goodman RC et al (2014) Improved allometric models to estimate the aboveground biomass of tropical trees. Glob Change Biol 20(10):3177–3190
- Dietze MC, Moorcroft PR (2011) Tree mortality in the eastern and central United States: patterns and drivers. Glob Change Biol 17(11):3312–3326
- Gelfand AE, Sahu SK, Carlin BP (1995) Efficient parameterizations for normal linear mixed models. Biometrika 82(3):479–488
- Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA (2003) National-scale biomass estimators for United States tree species. For Sci 49(1):12–35
- MacCleery DW (1993) American forests: a history of resiliency and recovery, vol 540. Forest History Society, Durham, North Carolina
- Marklund LG (1988) Biomass functions for pine, spruce and birch in Sweden. Swedish University of Agricultural Sciences, Uppsala
- McMahon SM, Parker GG, Miller DR (2010) Evidence for a recent increase in forest growth. Proc Natl Acad Sci 107(8):3611–3615
- Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, Phillips OL, Shvidenko A, Lewis SL, Canadell JG et al (2011) A large and persistent carbon sink in the worlds forests. Science 333(6045):988–993
- R Development Core Team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org
- Schimel DS, House J, Hibbard K, Bousquet P, Ciais P, Peylin P, Braswell BH, Apps MJ, Baker D, Bondeau A et al (2001) Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. Nature 414(6860):169–172

- Stephenson NL, Das AJ, Condit R, Russo SE, Baker PJ, Beckman NG, Coomes DA, Lines ER, Morris WK, Rüger N et al (2014) Rate of tree carbon accumulation increases continuously with tree size. Nat 507(7490):90–93
- Susan S (2007). Climate change 2007—the physical science basis: Working group I contribution to the fourth assessment report of the IPCC, vol 4. Cambridge University Press
- Vayreda J, Martinez-Vilalta J, Gracia M, Retana J (2012) Recent climate changes interact with stand structure and management to determine changes in tree carbon stocks in Spanish forests. Glob Change Biol 18(3):1028–1041
- Wright SJ (2005) Tropical forests in a changing environment. Trends Ecol Evol 20(10):553-560
- Wutzler T, Wirth C, Schumacher J (2008) Generic biomass functions for common beech (Fagus sylvatica) in Central Europe: predictions and components of uncertainty. Can J For Res 38(6):1661–1675
- Zianis D, Muukkonen P, Mäkipää R, Mencuccini M (2005) Biomass and stem volume equations for tree species in Europe, vol 4. Finnish Society of Forest Science, Finnish Forest Research Institute

Erin M. Schliep is a postdoctoral fellow at Duke University in the Department of Statistical Science.

Alan E. Gelfand is James B Duke Professor of Statistical Science and Professor of Environmental Sciences and Policy at Duke University.

James S. Clark is H.L. Blomquist Professor of the Nicholas School of the Environment, Professor of Biology, and Professor of Statistical Science at Duke University.

Kai Zhu is a postdoctoral fellow at Carnegie Institution for Science in the Department of Global Ecology and Stanford University in the Department of Biology.