Remote Sensing of Environment 182 (2016) 1–12

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Modeling forest biomass and growth: Coupling long-term inventory and LiDAR data

Chad Babcock [a,*], Andrew O. Finley [b], Bruce D. Cook [c], Aaron Weiskittel [d], Christopher W. Woodall [e]

[a] School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, USA
[b] Forestry Department, Michigan State University, East Lansing, MI 48824, USA
[c] Code 618, Biospheric Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD 20742, USA
[d] School of Forest Resources, University of Maine, Orono, ME 04469, USA
[e] USDA Forest Service, Northern Research Station, Forest Inventory and Analysis Program, 1992 Folwell Avenue, Saint Paul, MN 55114, USA

ARTICLE INFO

Article history:
Received 2 July 2014
Received in revised form 22 March 2016
Accepted 16 April 2016
Available online xxxx

Keywords:
LiDAR
Forest biomass
Biomass growth
Temporal misalignment
Long-term forest inventory
Bayesian hierarchical models
Markov Chain Monte Carlo
Gaussian process
Geospatial

ABSTRACT

Combining spatially-explicit long-term forest inventory and remotely sensed information from Light Detection and Ranging (LiDAR) datasets through statistical models can be a powerful tool for predicting and mapping above-ground biomass (AGB) at a range of geographic scales. We present and examine a novel modeling approach to improve prediction of AGB and estimate AGB growth using LiDAR data. The proposed model accommodates temporal misalignment between field measurements and remotely sensed data—a problem pervasive in such settings—by including multiple time-indexed measurements at plot locations to estimate AGB growth. We pursue a Bayesian modeling framework that allows for appropriately complex parameter associations and uncertainty propagation through to prediction. Specifically, we identify a space-varying coefficients model to predict and map AGB and its associated growth simultaneously. The proposed model is assessed using LiDAR data acquired from NASA Goddard's LiDAR, Hyper-spectral & Thermal imager and field inventory data from the Penobscot Experimental Forest in Bradley, Maine. The proposed model outperformed the time-invariant counterpart models in predictive performance as indicated by a substantial reduction in root mean squared error. The proposed model adequately accounts for temporal misalignment through the estimation of forest AGB growth and accommodates residual spatial dependence. Results from this analysis suggest that future AGB models informed using remotely sensed data, such as LiDAR, may be improved by adapting traditional modeling frameworks to account for temporal misalignment and spatial dependence using random effects.

1. Introduction

Linking long-term forest inventory with air- and space-borne Light Detection and Ranging (LiDAR) datasets via regression models offers an attractive approach to mapping forest above-ground biomass (AGB) at stand, regional, continental, and global scales. LiDAR data have shown great potential for use in estimating spatially explicit forest variables, including AGB, over a range of geographic scales (Asner, Hughes, Varga, Knapp, & Kennedy-Bowdoin, 2009; Babcock, Matney, Finley, Weiskittel, & Cook, 2013; Babcock et al., 2015; Finley, Banerjee, & MacFarlane, 2011; Iqbal, Dash, Ullah, & Ahmad, 2013; Muss, Mladenoff, & Townsend, 2011; Næsset, 2011; Neigh et al., 2013). Encouraging results from these and many other studies have spurred massive investment in new LiDAR sensors, sensor platforms, as well as extensive campaigns to collect field-based calibration data. For example, ICESat-2—planned for launch in 2017—will be equipped with a LiDAR sensor able to gather data from space at unprecedented spatial resolutions (Abdalati et al., 2010). As currently proposed, ICESat-2 will be a photon-counting sensor capable of recording photon-returns on a ≈70 cm footprint (ICESat-2, 2015). The Global Ecosystem Dynamics Investigation LiDAR (GEDI) will be an International Space Station mounted system capable of producing 25 m diameter footprint waveforms and is scheduled to be operational in 2018 (GEDI, 2014). One of GEDI's core objectives is to quantify the distribution of AGB at a fine spatial resolution. NASA Goddard's LiDAR, Hyper-spectral, & Thermal (G-LiHT) imager is an air-borne platform developed, in part, to examine how future space-originating LiDAR, e.g., ICESat-2, GEDI, or other platforms, may be combined with field-based validation measurements to build predictive models for AGB and other forest variables (Awadallah, Abbott, Wynne, & Nelson, 2013; Cook et al., 2013).

Along with increasing attention given to development of LiDAR data acquisition systems, there is continued and expanding interest in acquiring long-term forest inventory datasets for ecosystem monitoring initiatives (Albercht, Hanewinkel, & Bauhus, 2012; Phillips et al., 1998; Talbot et al., 2014). Long-term forest inventories including the United States Forest Service's Forest Inventory and Analysis (FIA) Program, Long-Term Ecological Research Network (LTER), and National

* Corresponding author.

http://dx.doi.org/10.1016/j.rse.2016.04.014
0034-4257/

Ecological Observatory Network (NEON) are collecting massive amounts of space and time indexed ecological data (Kampe, Johnson, Kuester, & Keller, 2010; Michener, Porter, Servilla, & Vanderbilt, 2011; Smith, 2002). These datasets are able to provide the calibration measurements necessary for large-scale AGB model development and mapping using LiDAR systems (Healey et al., 2012).

Much of the interest in LiDAR based AGB mapping is to support carbon monitoring, reporting, and verification (MRV) systems, such as defined by the United Nations Programme on Reducing Emissions from Deforestation and Forest Degradation (UN-REDD) and NASA's Carbon Monitoring System (CMS) (CMS, 2010; Le Toan et al., 2011; Ometto et al., 2014; UN-REDD, 2009). There are several challenges to specifying models that yield prediction useful for MRV systems. Here, we consider two common data characteristics that, if ignored in the modeling process, can result in misleading model-based inference about AGB or similar forest variables of interest. The first, and central focus of this study, is temporal misalignment between LiDAR data acquisition and AGB calibration data measurements. The second is spatial independence among model residuals—a key assumption of regression models. These characteristics are not simple nuisances that need accommodation to ensure valid inference. Explicit development of models that incorporate misalignment and residual spatial dependence can yield improved prediction of forest AGB and growth superior to that of commonly applied models.

In regression models, temporal misalignment occurs when the response variable and covariates (e.g., LiDAR metrics) are measured at different times. Temporal misalignment is common in settings where permanent sample plots (PSPs)—remeasured on a periodic basis as part of a long-term inventory or monitoring system—are used to calibrate remotely sensed data. For example, Saatchi et al. (2011) use field data from inventories conducted between 1995 and 2005 when building regression models for AGB with LiDAR data. Zhang et al. (2014) use FIA inventories collected between 2001 and 2007 to predict AGB using LiDAR data for California. Blackard et al. (2008), trained MODIS based regression models using national inventory data ranging from 1990 to 2003. Gregoire et al. (2011), García, Riaño, Chuvieco, and Danson (2010), Chen, Vaglio Laurin, Battles, and Saah (2012) and Anderson et al. (2008) use data from long-term inventories spanning a three year period to calibrate regression models. Some researchers elect to retain field data from long-term inventories spanning 10+ year intervals (Blackard et al., 2008; Saatchi et al., 2011) to obtain calibration datasets with large numbers of observations, whereas other researchers sacrifice sample size by only using inventories spanning a small time interval (Anderson et al., 2008; Chen et al., 2012; García et al., 2010; Gregoire et al., 2011). Because the FIA (the largest long-term forest inventory network both in spatial extent and number of PSPs) only measures between ten and twenty percent of their PSPs annually, we should expect to continue to see temporal misalignment to be an point of concern when attempting to combine long-term inventory and remote sensing data (Bechtold & Patterson, 2005). As another example, the recently established long-term inventory, NEON, only plans to acquire field inventories from their PSPs every five years, but collect airborne remote sensing data over these plots annually (Kampe et al., 2010; Kao et al., 2012). In order to develop yearly maps of forest structure and prediction uncertainty leveraging the annual LiDAR acquisitions, NEON will likely need sound statistical modeling strategies that accommodate temporal misalignment.

Regression model parameters (i.e., intercept, slope coefficients associated with LiDAR metrics, and residual variance) estimated using temporally misaligned data can lead to erroneous model-based inference and prediction especially as forest disturbances operate on temporal time scales rarely aligned with field and remotely sensed inventory programs. For AGB modeling, the worst cases will arise when there is a large time lag between LiDAR acquisition (covariates) and AGB measurements (response variable). The posited relationship between LiDAR metrics and response variables will be further weakened if the PSPs are visited across a large time interval (e.g., FIA sample cycle lengths) in concert with rapid changes in AGB (e.g., due to wildfire or wind throw). In both cases, *large* is relative to forest growth rates and disturbance intervals. Strategies to address misalignment include discarding all measurements outside of some *acceptable* time interval around the LiDAR acquisition date, using growth models to project PSP measurements to align with LiDAR acquisition date, or simply assume temporal misalignment effects are negligible (Blackard et al., 2008; Weiskittel, Hann, Kershaw, & Vanclay, 2011; Zhang et al., 2014). In many cases, such strategies are less than ideal—due to small resulting sample size, desire to avoid unrealistic assumptions about growth, or need to quantify components of uncertainty in AGB predictions.

Repeated time- and space-indexed measurements within long-term inventory datasets make possible an approach to the misalignment problem that does not discard data and facilitates uncertainty quantification on both AGB and growth. The framework we pursue builds on work by Banerjee and Johnson (2006) and Nathoo (2010) who model spatially varying growth curves using repeated measures. In both works, regression models are specified with intercept and slope coefficients arising from univariate or multivariate spatial Gaussian processes. Here, we also use a space-varying coefficients framework but incorporate LiDAR covariates to model and map AGB and growth simultaneously. In addition to accommodating temporal misalignment the proposed model captures residual spatial dependence (i.e., spatially structured variability not explained by LiDAR covariates) via a space-varying intercept (Banerjee, Carlin, & Gelfand, 2014; Gelfand, Kim, Sirmans, & Banerjee, 2003). Similar to Banerjee and Johnson (2006) and Nathoo (2010), we use a Bayesian model specification that allows for appropriately complex parameter associations and uncertainty propagation through to prediction.

To better understand how LiDAR and long-term inventory datasets can be used to map AGB, we look to Penobscot Experimental Forest (PEF) in Bradley, Maine. Using the PEF's long-term inventory database along with G-LiHT data acquired for the area in 2012, we present and examine the proposed modeling approach to improve AGB prediction and estimate growth. The proposed model is compared with two benchmark regressions to examine predictive performance, ability to account for temporal misalignment, and capacity to accommodate spatial dependence.

2. Methods

2.1. Field data description and preparation

The PEF is a 1600 ha tract of Acadian forest land located in Bradley, Maine (44° 52′ N, 68° 38′ W) (Fig. 1). The average annual temperature and precipitation near Bradley is 6 °C and 110 cm respectively. Composition of the PEF is a mix of coniferous and deciduous species including spruce (*Picea* spp.), balsam fir (*Abies balsamea*), red maple (*Acer rubrum*), birch (*Betula* spp.), and aspen (*Populus* spp.), among others. Since the 1950s, the PEF has been subjected to routine management and monitoring initiatives. With over 600 PSPs currently established and scheduled for remeasurement on an approximate ten-year cycle (and immediately pre- and post-harvest), the PEF has a wealth of calibration data available for fitting AGB models using LiDAR (Brissette & Kenefic, 2008). The PEF is divided into over 50 management units (MU). Within each MU, different silvicultural treatments are implemented, e.g., unregulated harvest, shelterwood, diameter limit cutting, or natural regeneration (Fig. 1). MU 32 is a mature, natural area that serves as a reference stand for the PEF with limited harvesting or other management actions conducted since 1954. MU 8 and MU 22 were commercially clearcut in 1984 and 1989, respectively (Fig. 1). MU 23 is undergoing a three-stage shelterwood management plan (Sendak, Brissette, & Frank, 2003). For further details regarding the

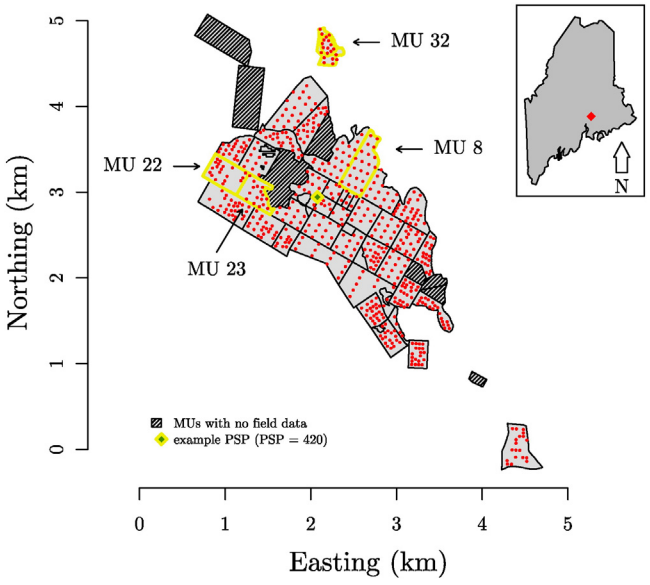

Fig. 1. Map of PEF. PSP's highlighted in red. Example PSP (site 420) colored in green. Black polygon boundaries outline different management units (MU) on the PEF. Hashed out polygons identify MUs with no inventory data. Select MUs have been labeled and highlighted in yellow. MU 8 and MU 22 were clear cut in 1984 and 1989 respectively. MU 32 is an old-growth stand that has experienced very limited management activity since 1954. MU 23 is a stand undergoing a three-stage shelterwood harvesting technique. The inset map in the upper right corner shows the location of the PEF with respect to Maine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

PEF's MU silvicultural treatments see Hayashi, Weiskittel, and Sader (2014).

For this study we use 2203 field measurements from 604 PSPs (Fig. 2). Only post-disturbance measurements from each PSP were included. Further, only PSPs with at least one post-disturbance measurement recorded between 1999 and 2011 were used. Data for the 33 measurements collected in 2012 were held out for model fit comparison (2012 was the most current data available for the PEF at the time of our study). Each PSP contains three nested sampling areas with radii of approximately 16 m (1/5 acre), 8 m (1/20 acre), and 5 m (1/50 acre), which are used to measure trees of progressively smaller diameter at breast height (dbh) classes (Brissette, Kenefic, Russel, & Puhlick, 2012). For this analysis, only stems with dbh measures greater than or equal to 12.7 cm (5 in) were used to calculate AGB, which precluded trees measured only in the 8 m and 5 m plots. AGB was calculated using dbh based, species specific, allometric equations provided by (Jenkins, Chojnacky, Heath, & Birdsey, 2003). The response variable used in subsequent regression models was calculated for each PSP measurement date by summing individual tree AGB values, scaling to megagrams per ha (Mg/ha), then applying a square-root transformation to better approximate a Gaussian distribution and ensure positive support following back-transformation of predicted values.

2.2. LiDAR acquisition and preparation

Large footprint waveforms characteristic of upcoming space-based LiDAR instruments were simulated using discrete, multistop returns from the G-LiHT airborne imager collected over the PEF in 2012. G-LiHT is a portable multi-sensor airborne system developed by NASA Goddard Space Flight Center that simultaneously maps the composition and structure of terrestrial ecosystems. The G-LiHT airborne laser scanner (VQ-480, Riegl Laser Measurement Systems, Horn, Austria) uses a 1550 nm laser that provides an effective measurement rate of up to 150 kHz along a 60° swath perpendicular to the flight direction. At a nominal flying altitude of 335 m, each laser pulse has a footprint approximately 10 cm in diameter and is capable of producing up to 8 returns.

Pseudo-waveforms were created over the PEF's 604 PSPs by aggregating G-LiHT LiDAR returns and weighting return heights using a Gaussian shaped 25 m diameter footprint (Blair & Hofton, 1999). The pseudo-waveforms were then used to calculate percentile heights at 5% intervals between 5% and 100% (Fig. 3). Principal Components Analysis (PCA) via eigen decomposition of the percentile height variables correlation matrix was used to reduce the dimension of the data set and ensure the variables used as covariates in the subsequent regression analysis were uncorrelated. An assessment of the principal component eigenvalues showed that >85% of the variation in the percentile height data was accounted for by the first two eigenvectors (principal components). Hence, scores generated from the first two eigenvectors were used as covariates for model fitting. Pseudo-waveforms were also generated for a grid of 8283 25 × 25 m cells over the PEF. PCA scores were calculated for each grid cell and used for subsequent prediction and mapping of AGB and AGB growth.

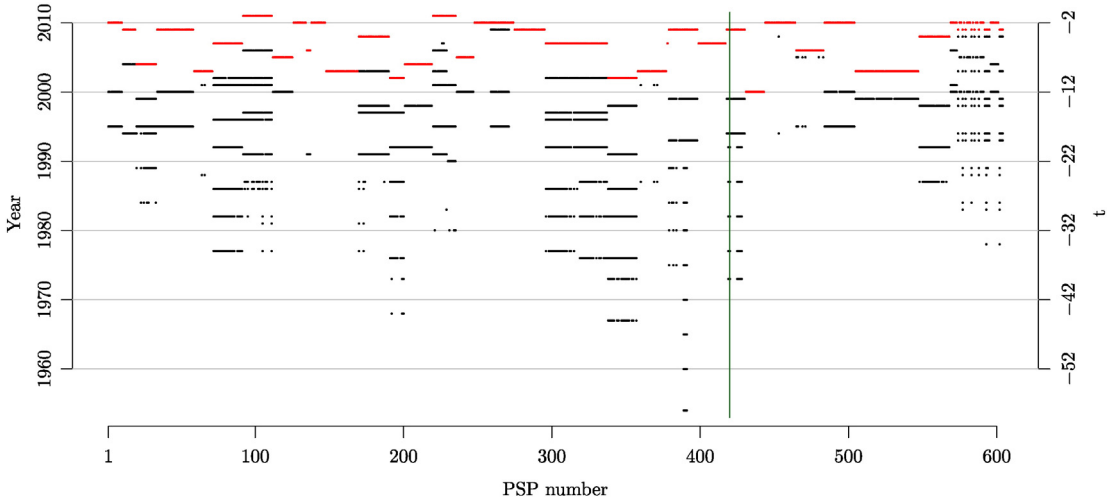

Fig. 2. Above-ground biomass remeasurements used in analysis. Most recent remeasurement for each permanent sample plot (PSP) is highlighted in red. Horizontal axis is the PSP identification number. Left vertical axis shows the remeasurement year and the right vertical axis shows the corresponding t value. t is the number of years after LiDAR collection. Since all remeasurements were conducted before the year of LiDAR acquisition, all t's are negative. Vertical green line identifies example PSP (site 420, highlighted in Fig. 4). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

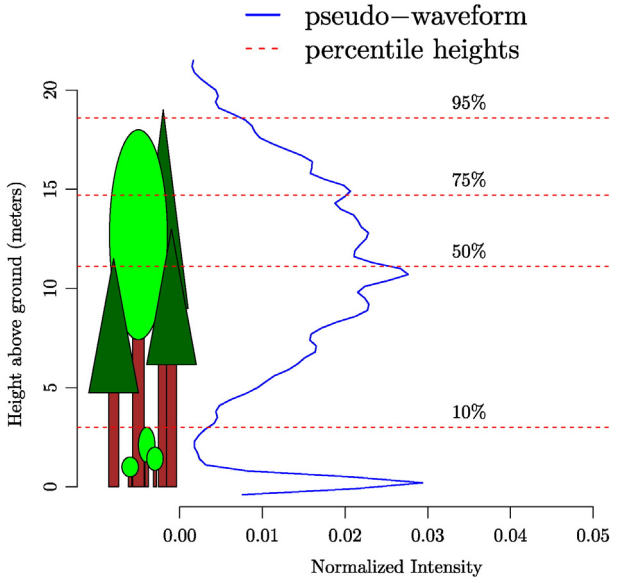

Fig. 3. Illustration showing the normalized G-LiHT pseudo-waveform intensity of return energy height profile (blue) along with example percentile height values (red). Return intensity is greater at heights where the forest canopy is encountered. The energy spike at 0 m signifies energy returning from ground. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. Model description

Let $Y_t(\mathbf{s})$ be the square-root transformed AGB value at location $\mathbf{s}\in\mathcal{D}$ and year t centered on the year of LiDAR acquisition, where $\mathbf{s}$ is a vector of location coordinates, e.g., easting and northing, and $\mathcal{D}$ is the forested domain. A model describing AGB growth at site $\mathbf{s}$ can be written as,

$$Y_t(\mathbf{s}) = \alpha_0(\mathbf{s}) + \alpha_1(\mathbf{s})t + \varepsilon_t(\mathbf{s}) \quad (1)$$

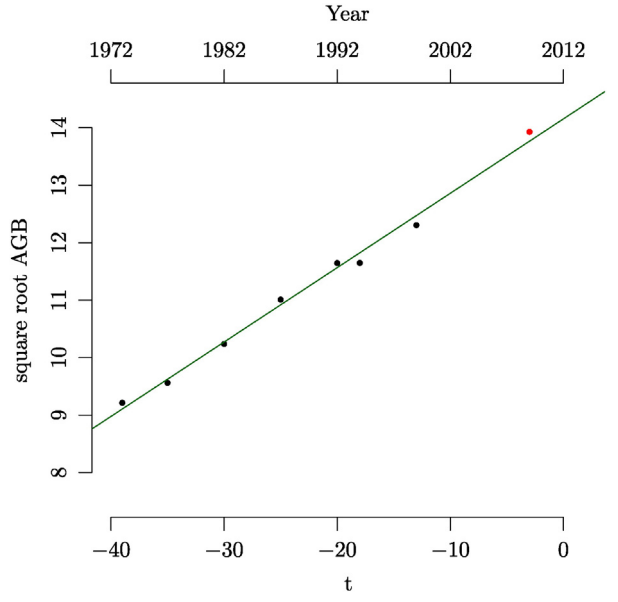

Fig. 4. Square root transformed AGB versus time graph for example PSP (site 420, identified with a vertical green line in Fig. 2). Upper horizontal axis shows the remeasurement year and the lower horizontal axis shows the corresponding t value. t is the number of years after LiDAR collection. Since all remeasurements were conducted before the year of LiDAR acquisition, all t's are negative. Most recent remeasurement highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\alpha_0(\mathbf{s})$ is the estimated value of $Y_t(\mathbf{s})$ when $t=0$ and $\alpha_1(\mathbf{s})$ equals the additive adjustment for every one year change in t. For the example PSP shown in Fig. 4, $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ correspond to the intercept and slope of the green line respectively. Since $t=0$ for the year the LiDAR data was collected, $\alpha_0(\mathbf{s})$ can be interpreted as an estimate of AGB at the time of LiDAR acquisition. Higher values of $\alpha_1(\mathbf{s})$ correspond to faster AGB accumulation and lower values of $\alpha_1(\mathbf{s})$ suggest slower AGB growth over time. Note the $\mathbf{s}$ on α_0 and α_1 identify these as location specific parameters. An independent white-noise process $\varepsilon_t(\mathbf{s})$ captures measurement error in $Y_t(\mathbf{s})$ and is modeled as $\mathcal{N}(0,\tau^2)$. This is an attractive choice for modeling long-term inventory and LiDAR data. Specifying the first stage of our hierarchical design as (1) allows the model to acknowledge and account for temporal misalignment between field and remote sensing data through the introduction of $\alpha_1(\mathbf{s})t$. Further, this framework allows for the inclusion of multiple time-indexed measurements at one location to inform AGB growth estimation.

Letting $\boldsymbol{\alpha}_0=(\alpha_0(\mathbf{s}_i))_{i=1}^N$ and $\boldsymbol{\alpha}_1=(\alpha_1(\mathbf{s}_i))_{i=1}^N$, where N equals the total number of observed locations ($N=604$ PSPs), we write the second level of our hierarchical design as,

$$\boldsymbol{\alpha}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \quad (2)$$

$$\boldsymbol{\alpha}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad (3)$$

where, $\boldsymbol{\mu}_0=(\mu_0(\mathbf{s}_i))_{i=1}^N$ and $\boldsymbol{\mu}_1=(\mu_1(\mathbf{s}_i))_{i=1}^N$ constitute the mean components of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$ *Gaussian processes* (*GP*), respectively. For our subsequent analysis, we set $\mu_0(\mathbf{s})=\mathbf{x}(\mathbf{s})\boldsymbol{\beta}$, where $\mathbf{x}(\mathbf{s})$ is a row vector comprised of a 1 and the two LiDAR PCA scores for site $\mathbf{s}$ and $\boldsymbol{\beta}$ is the corresponding column vector of regression parameters, i.e., $\mathbf{x}(\mathbf{s})=(1,x_1(\mathbf{s}),x_2(\mathbf{s}))$ and $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)^T$. We set $\mu_1(\mathbf{s})=\gamma$, where γ is a single regression parameter. Note that this framework can be easily generalized to include any number of covariates to help estimate $\boldsymbol{\alpha}_0$ (AGB) and $\boldsymbol{\alpha}_1$ (AGB growth) simply by redefining $\mu_0(\mathbf{s})$ and $\mu_1(\mathbf{s})$ accordingly.

The $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ describe the spatial covariance of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$. We can write (2) and (3) as

$$\boldsymbol{\alpha}_0 = \boldsymbol{\mu}_0 + \mathbf{w}_0, \text{ where } \mathbf{w}_0 \sim \mathcal{N}(0_-, \boldsymbol{\Sigma}_0) \quad (4)$$

$$\boldsymbol{\alpha}_1 = \boldsymbol{\mu}_1 + \mathbf{w}_1, \text{ where } \mathbf{w}_1 \sim \mathcal{N}(0_-, \boldsymbol{\Sigma}_1). \quad (5)$$

$\mathbf{w}_0=(w_0(\mathbf{s}_i))_{i=1}^N$ and $\mathbf{w}_1=(w_1(\mathbf{s}_i))_{i=1}^N$ are zero centered random effects and provide local adjustment with spatial dependence to $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$. Such flexibility is desired because we might expect AGB and growth, i.e., $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$, to have some spatial pattern over a forest domain. If the covariates used to inform the model (i.e., LiDAR metrics) fail to fully account for the spatial structure in the outcome variable (AGB), $\mathbf{w}_0$ and $\mathbf{w}_1$ will absorb any unaccounted for spatial structure, thereby reducing spatial dependence among the residuals. Aside from helping to ensure statistical validity, the spatial random effects yield other modeling benefits. For example, we can interpret $\mathbf{w}_0$ and $\mathbf{w}_1$ as capturing the contribution of unobserved (or unobservable) covariates with spatial pattern. Capturing these latent covariates via the spatial random effects should result in improved fit and predictive performance. The inclusion of $\mathbf{w}_1$ provides an additional benefit specific to modeling growth with long-term inventory data. Often at locations, only one post-disturbance inventory will be available—for this analysis 173 PSPs had one post-harvest measurement (Fig. 2). By introducing $\mathbf{w}_1$, estimation of growth at sites with one measurement becomes possible because the model borrows biomass accumulation information from neighboring sites where multiple inventories may be available.

The spatial random effects are specified as $w_0(\mathbf{s}) \sim GP(0, C(\cdot,\cdot,\boldsymbol{\theta}_0))$ and $w_1(\mathbf{s}) \sim GP(0, C(\cdot,\cdot,\boldsymbol{\theta}_1))$, where $C(\mathbf{s},\mathbf{s}',\boldsymbol{\theta}_0)=Cov(w_0(\mathbf{s}),w_0(\mathbf{s}'))$ and $C(\mathbf{s},\mathbf{s}',\boldsymbol{\theta}_1)=Cov(w_1(\mathbf{s}),w_1(\mathbf{s}'))$ are functions that model the covariance between any pair of locations $\mathbf{s}$ and $\mathbf{s}'$. The (i,j)-th elements of the $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ spatial covariance matrices are given by $C(\mathbf{s}_i,\mathbf{s}_j,\boldsymbol{\theta}_0)$ and

$C(\mathbf{s}_i, \mathbf{s}_j, \boldsymbol{\theta}_1)$, respectively. To ensure $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are symmetric and positive definite, the spatial covariance functions are defined as

$$C(\mathbf{s}, \mathbf{s}', \boldsymbol{\theta}_0) = \sigma_0^2 \rho(\mathbf{s}, \mathbf{s}'; \phi_0), \text{and } \boldsymbol{\theta}_0 = \{\sigma_0^2, \phi_0\} \quad (6)$$

$$C(\mathbf{s}, \mathbf{s}', \boldsymbol{\theta}_1) = \sigma_1^2 \rho(\mathbf{s}, \mathbf{s}'; \phi_1), \text{and } \boldsymbol{\theta}_1 = \{\sigma_1^2, \phi_1\}, \quad (7)$$

where $\rho(\cdot;\phi_0)$ and $\rho(\cdot;\phi_1)$ are *correlation functions* with ϕ_0 and ϕ_1 determining the rate of correlation decay. The spatial variance parameters σ_0^2 and σ_1^2 equal $Var(w_0(\mathbf{s}))$ and $Var(w_1(\mathbf{s}))$, respectively. For the subsequent analysis, exponential correlation functions were assumed for $\mathbf{w}_0$ and $\mathbf{w}_1$, where $\rho(||\mathbf{s}-\mathbf{s}'||;\phi_0) = \exp(-\phi_0||\mathbf{s}-\mathbf{s}'||)$ and $\rho(||\mathbf{s}-\mathbf{s}'||;\phi_1) = \exp(-\phi_1||\mathbf{s}-\mathbf{s}'||)$ with $||\mathbf{s}-\mathbf{s}'||$ being defined as the Euclidean distance separating sites $\mathbf{s}$ and $\mathbf{s}'$. To ease interpretation of the ϕ estimates, corresponding *effective spatial range* estimates are presented in Table 1 and labeled as esr_0 and esr_1. We define *effective spatial range* as the distance (m) where the spatial correlation between locations drops to 0.05.

The data model for (1) is constructed as follows. Let $n(\mathbf{s})$ be the number of remeasurements at location $\mathbf{s}$. Defining the $M\times 1$ column vectors $\mathbf{Y} = ((Y_{t_j}(\mathbf{s}_i))_{j=1}^{n(\mathbf{s}_i)})_{i=1}^{N}$, $\mathbf{t} = ((t_j(s))_{j=1}^{n(\mathbf{s}_i)})_{i=1}^{N}$, and $\boldsymbol{\varepsilon} = ((\varepsilon_{t_j}(\mathbf{s}_i))_{j=1}^{n(\mathbf{s}_i)})_{i=1}^{N}$, where $M = \sum_{i=1}^{N} n(\mathbf{s}_i)$, we write the model as,

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\alpha}_0 + \mathbf{Z}\boldsymbol{\alpha}_1 \mathbf{t} + \boldsymbol{\varepsilon}, \quad (8)$$

where $\mathbf{Z}$ is an $M\times N$ block-diagonal matrix with the i-th diagonal element being an $n(\mathbf{s}_i)\times 1$ vector of ones.

To complete the Bayesian specification of our hierarchical model, prior distributions are assigned to the parameters and inference proceeds by sampling from the posterior distribution. As customary, we assume $\boldsymbol{\beta}$ and γ follow $\mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$ and $\mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$ priors (Gelman, Carlin, Stern, & Rubin, 2013). The spatial variance components σ_0^2 and σ_1^2, along with the measurement error τ^2, are assigned inverse-Gamma, $IG(a,b)$, priors. The spatial decay parameters ϕ_0 and ϕ_1 follow a Uniform prior distribution, $Unif(a,b)$, with support over the geographic range of the study area.

Using notation similar to Gelman et al. (2013), we can write the posterior distribution of the parameters as $p(\boldsymbol{\Omega}|\mathbf{Y})$, where $\boldsymbol{\Omega} = \{\boldsymbol{\beta}, \gamma, \mathbf{w}_0, \mathbf{w}_1, \sigma_0^2, \sigma_1^2, \phi_0, \phi_1, \tau^2\}$, which is proportional to

$$\begin{aligned} &Unif(\phi_0|a_{\phi_0}, b_{\phi_0}) \times Unif(\phi_1|a_{\phi_1}, b_{\phi_1}) \\ &\times IG\left(\sigma_0^2|a_{\sigma^2_0}, b_{\sigma_0^2}\right) \times IG\left(\sigma_1^2|a_{\sigma_1^2}, b_{\sigma_1^2}\right) \\ &\times \mathcal{N}\left(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) \times \mathcal{N}\left(\gamma|\mu_\gamma, \sigma_\gamma^2\right) \times IG(\tau^2|a_\tau, b_\tau) \\ &\times \mathcal{N}(\mathbf{w}_0|0_-, \boldsymbol{\Sigma}_0) \times \mathcal{N}(\mathbf{w}_1|0_-, \boldsymbol{\Sigma}_1) \times \mathcal{N}\left(\mathbf{Y}|\mathbf{Z}\boldsymbol{\alpha}_0 + \mathbf{Z}\boldsymbol{\alpha}_1\mathbf{t}, \tau^2\mathbf{I}_M\right). \end{aligned} \quad (9)$$

Posterior samples of $\boldsymbol{\Omega}$ are collected via Markov Chain Monte Carlo (MCMC) algorithms detailed in (Banerjee et al., 2014).

The posterior predictive distribution of the spatial random effects $\mathbf{w}_0$ and $\mathbf{w}_1$ at new (unobserved) locations are given by

$$p(\mathbf{w}_{\sim 0}|\mathbf{Y}) \propto \int p(\mathbf{w}_{\sim 0}|\mathbf{w}_0, \boldsymbol{\Omega}_0, \mathbf{Y}) p(\mathbf{w}_0|\boldsymbol{\Omega}_0, \mathbf{Y}) p(\boldsymbol{\Omega}_0|\mathbf{Y}) d\boldsymbol{\Omega}_0 \mathbf{w}_0 \quad (10)$$

$$p(\mathbf{w}_{\sim 1}|\mathbf{Y}) \propto \int p(\mathbf{w}_{\sim 1}|\mathbf{w}_1, \boldsymbol{\Omega}_1, \mathbf{Y}) p(\mathbf{w}_1|\boldsymbol{\Omega}_1, \mathbf{Y}) p(\boldsymbol{\Omega}_1|\mathbf{Y}) d\boldsymbol{\Omega}_1 \mathbf{w}_1, \quad (11)$$

where $\mathbf{w}_{\sim 0} = (w_0(\mathbf{s}_{\sim i}))_{i=1}^{m}$ and $\mathbf{w}_{\sim 1} = (w_1(\mathbf{s}_{\sim i}))_{i=1}^{m}$ are the collections over prediction locations $\mathbf{s}_{\sim}$'s, and $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ include the appropriate subsets of parameters from $\boldsymbol{\Omega}$. Then given the set of LiDAR covariates at location $\mathbf{s}_{\sim}$, (4) and (5) can be evaluated to obtain $\alpha_0(\mathbf{s}_{\sim})$ and $\alpha_1(\mathbf{s}_{\sim})$. Finally, the posterior predictive distribution of the response variable, $Y_t(\mathbf{s}_{\sim})$, is obtained from $\mathcal{N}(\alpha_0(\mathbf{s}_{\sim}) + \alpha_1(\mathbf{s}_{\sim})t, \tau^2)$. Samples from the predictive distribution of each component of interest, i.e., $w_0(\mathbf{s}_{\sim}), w_1(\mathbf{s}_{\sim}), \alpha_0(\mathbf{s}_{\sim}), \alpha_1(\mathbf{s}_{\sim})$, and $Y_t(\mathbf{s}_{\sim})$, are obtained via composition sampling, see, e.g., Finley, Banerjee, and Gelfand (2015) and Banerjee et al. (2014) for details. Summary statistics of these posterior predictive distributions, e.g., mean, median, variance, interquartile range and credible intervals, can then be computed and mapped as illustrated in Section 3.

2.4. Model comparison and validation

The modeling framework described in Section 2.3 (test model) is compared with the benchmark regression models

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})\boldsymbol{\beta} + \varepsilon(\mathbf{s}) \text{ (benchmark 1)} \quad (12)$$

and

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})\boldsymbol{\beta} + w_0(\mathbf{s}) + \varepsilon(\mathbf{s}) \text{(benchmark 2)}, \quad (13)$$

where $\boldsymbol{\beta}$ and $w_0(\mathbf{s})$ were defined previously. Further, the benchmark models provide a way to evaluate the test model's ability to accommodate temporal misalignment and need to accommodate residual spatial dependence. Specification of the benchmark models (including parameter prior distributions) are defined analogous to the test model, aside from $Y(\mathbf{s})$ which is the most recent measurement of square-root transformed AGB for location $\mathbf{s}$ (red measurements in Figs. 2 and 4) and $\varepsilon(\mathbf{s}) \sim \mathcal{N}(0, \tau^2)$. Again, posterior samples of the benchmark models' parameters and predictions are collected via an MCMC algorithm and composition sampling in the same fashion as the test model. Neither benchmark model acknowledges temporal misalignment between the time LiDAR and field data were acquired—which is typical of most contemporary modeling exercises that couple LiDAR and long-term forest inventory data. Because we can only include the most recent measurements at each PSP, the benchmark models lack the ability to estimate AGB growth and hence only estimate AGB. Benchmark model 2 includes a

Table 1
Parameter estimates and associated 95% credible intervals for benchmark model 1, 2, and the test model. 10-fold RMSE is presented in Mg/ha units. esr_0 and esr_1 are the *effective spatial ranges*, or the distances (m) at which correlation drops to 0.05, for $\mathbf{w}_0$ and $\mathbf{w}_1$ respectively.

		Benchmark 1	Benchmark 2	Test model
Parameter estimates 50% (2.5%, 97.5%)	β_0	9.73 (9.56, 9.90)	9.53 (8.52, 10.51)	10.55 (10.12, 10.98)
	β_1	−0.38 (−0.42, −0.33)	−0.28 (−0.32, −0.23)	−0.27 (−0.31, −0.23)
	β_2	0.72 (0.62, 0.82)	0.38 (0.29, 0.47)	0.34 (0.25, 0.43)
	γ	–	–	0.173 (0.111, 0.23)
	τ^2	4.84 (4.34, 5.45)	1.16 (0.82, 1.53)	0.048 (0.045, 0.052)
	esr_0	–	810.85 (537.45, 1860.61)	283.62 (227.49, 365.23)
	σ_0^2	–	4.10 (2.89, 7.66)	4.19 (3.55, 5.12)
	esr_1	–	–	1112.12 (693.52, 2501.50)
	σ_1^2	–	–	0.009 (0.006, 0.018)
Prediction metrics	10-fold RMSE	42.88	31.27	17.52
	95% CP	0.94%	0.94%	0.92%
	95% CIW	169.25	117.26	54.21

spatial random effect that should absorb any spatially structured variation that may be present in the residuals of benchmark model 1.

To test predictive performance of the three candidate models, a ten-fold holdout set design was constructed by separating the PSPs ($n = 604$) into ten approximately equal-sized groups. Square-root transformed AGB for the holdout groups was sequentially predicted given model parameters estimated using data in the remaining nine groups. Each posterior predictive distribution sample was then back-transformed prior to calculating the AGB posterior predictive distribution mean at each holdout location. Note, because we are working within a Bayesian setting, inference proceeds from the entire back-transformed posterior distribution and hence we avoid issues with back-transformation bias encountered within a frequentist inferential paradigm, see, e.g., Stow, Reckhow, and Qian (2006) for further discussion. Root mean squared error (RMSE) was calculated using the back-transformed holdout posterior predicted means and observed AGB for each of the three models (labeled 10-fold RMSE in Table 1). The model with the lowest 10-fold RMSE was considered the "best" predicting model.

To graphically assess the impact of temporal misalignment on 2012 AGB prediction, scatterplots comparing candidate model fitted values from the 33 PSPs with measurements taken in 2012 and the actual 2012 remeasurements were examined (recall, as noted in Section 2.1, 2012 data was not used to estimate model parameters and reserved for model comparison).

2.5. Posterior inference

Model-based parameter and predictive inference for each of the three candidate models were based on three MCMC chains run for 25,000 iterations. Diffuse hyper-parameters for the $IG(a,b)$, $\mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, and $\mathcal{N}(\mu_\gamma, \sigma^2_\gamma)$ prior distributions were selected to minimize influence on the posterior distribution. The *IG a* hyper-parameter was set to 2 which results in a prior distribution mean equal to *b* and infinite variance. The $\mathcal{N}$ hyper-parameters were set to mean zero and variances equal to 10,000. Each MCMC was initiated with different starting values and random seeds. Chain convergence was diagnosed by monitoring mixing and the Gelman-Rubin statistic (Cowles & Carlin, 1996; Gelman & Rubin, 1992). Satisfactory convergence was achieved after 10,000 iterations for all models. Posterior inference was based on a post burn-in sub-sample of 15,000 iterations (5000 from each chain).

3. Results

3.1. Parameter estimates and prediction accuracy

Parameter estimates and prediction metrics for the three candidate models, detailed in Sections 2.3 and 2.4, are given in Table 1. Here, we can see the 95% credible intervals for β_1 and β_2 do not include zero—the LiDAR PCA covariates explain a substantial portion of the variability in AGB for all three models. Regarding prediction performance results in Table 1, benchmark model 1 yields the largest 10-fold RMSE (42.88 Mg/ha) among the candidate models. In comparison, the spatial random effects in benchmark model 2 and the test model capture substantial residual spatial structure and hence improve prediction at unobserved locations, i.e., by borrowing information from observed measurements. There is substantial residual spatial dependence and space-varying impact of *t*, as indicated by long *effective spatial range* estimates (esr_0 and esr_1, noting a maximum inter-location distance between PSP of ~5.6 km) and relatively large spatial variances (σ^2_0 and σ^2_1). This, along with increased prediction accuracies, is compelling evidence that spatial dependence is present in the residuals of benchmark model 1—suggesting inference about parameters and predictions could be compromised. Indeed, benchmark model 1 prediction is less accurate compared with the other models.

Between the random effects models, the test model provides a substantially lower 10-fold RMSE compared to the benchmark model 2 (17.52 versus 31.27 Mg/ha). Similarity between the models' β_1 and β_2 estimates suggests that once residual spatial dependence is accommodated, via $\mathbf{w}_0$, both models provide similar LiDAR calibration and any improvement in prediction is due to correction for temporal misalignment.

The test model's increased prediction accuracy, along with higher β_0, suggests benchmark model 2 yields a negative prediction bias. Because the most recent measurement at each PSP occurred before G-LiHT data collection, a negative prediction bias should be expected. Calibrating LiDAR data to past measures of AGB, without accounting for the temporal mismatch, will result in underestimates of AGB for the year of remote sensing data collection. The difference in 10-fold RMSE between the two models suggests the negative bias in benchmark model 2 is substantial and correcting for it, as we have in the test model, increases prediction accuracy. Fig. 5 graphically demonstrates the negative bias introduced by ignoring the effects temporal misalignment between PSP remeasurement and remote sensing acquisition. We see in Fig. 5a that benchmark model 1 fitted values at the 33 PSPs with 2012 remeasurements are substantially lower than actual 2012 AGB. Fig. 5b shows that this bias is reduced but still present when a spatial random effect is introduced. Only after temporal misalignment is explicitly accommodated via the introduction of a growth related random effect is the model fit bias effectively removed (Fig. 5c).

3.2. Predictive uncertainty and precision

In addition to assessing prediction accuracy, we are interested in candidate models' capacity to provide appropriate measures of prediction uncertainty. To assess the quality of uncertainty quantification, we count the number of holdout set observations that fall within their respective posterior predictive distribution credible intervals, e.g., we can use a credible set that gives a 95% credible interval (Carlin & Louis, 2008, pp. 48), then divide this count by the total number of observations in the holdout set and multiply by 100 to yield an empirical coverage percentage (CP). We expect the CP to be approximately equal to the chosen credible set, e.g., ~95% of the holdout observations should fall within their respective 95% credible interval bounds. These CPs are given in Table 1 for the 10-fold holdout measurements. The CP results suggest that all models provide appropriate estimates of prediction uncertainty—even benchmark 1 that violates the independent and identically distributed residuals assumption.

CP is useful for identifying misspecified models and issues arising from prediction in extrapolation settings; however, it does not provide information about the precision of prediction. For this we can look at the posterior predictive distribution credible interval width (CIW)—more precise predictions have shorter CIW. Table 1 provides the mean CIW calculated using the 95% posterior predictive credible interval width for the 10-fold holdout sets. Here, the 10-fold CIW for the test model is approximately half the width of the benchmark 2 and about 1/3 that of benchmark 1.

3.3. Prediction and influence of PSP measurements

The test model was used to generate posterior predictive distributions for 2012 AGB and AGB growth ($\boldsymbol{\alpha}_1$) for each of the 8283 25×25 m cells over the PEF. The mean of each cell's posterior predictive distribution is mapped in Figs. 6a and 7a and serves as our best predictions of 2012 AGB and growth (posterior predictive samples were back-transformed prior to calculating summary statistics and subsequent maps). In addition to these *point estimate* maps, the 95% posterior predictive distribution CIW was calculated and mapped for AGB and growth, Figs. 6b and 7b respectively.

Figs. 6b and 7b show that CIW for AGB and growth is generally largest in MUs with no inventory data (hashed out MUs in Fig. 1). This

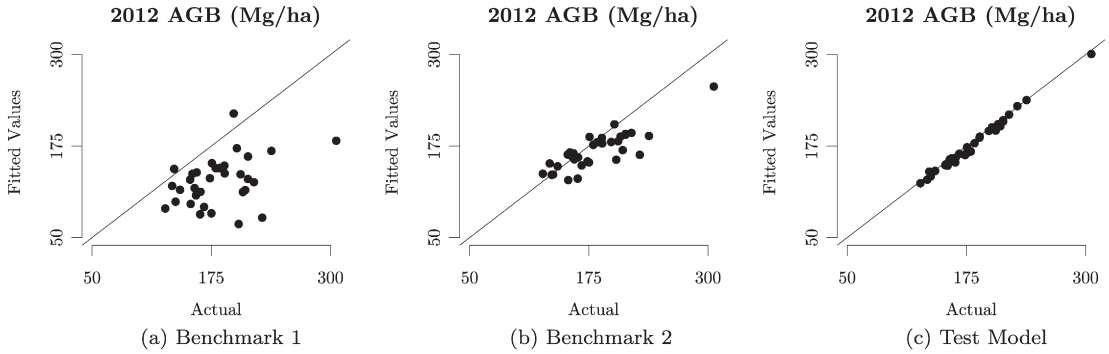

Fig. 5. Fitted values versus actual 2012 AGB scatter-plots for the three models. The line on each plot is the one-to-one line.

behavior is expected and clearly illustrates how precision of prediction decreases as we move away from the field data used to calibrate the model. In Fig. 6b, we can even identify PSP locations as points of high precision, i.e., narrow CIWs, within MUs. The plot of grid cell AGB CIW versus distance to nearest PSP (Fig. 8a) further illustrates the negative relationship between precision of prediction and distance to observed PSPs, i.e., high precision at and near observed PSPs. Additionally, Fig. 8a suggests that, on average, AGB CIW increases until distances of ~200–300 m from PSP locations, after which it levels off. This corresponds to the estimated value for esr_0 for the test model (Table 1). Similarly, Fig. 8b shows how the precision of predicted $\alpha_1(\mathbf{s})$ changes as a function of distance to the nearest PSP. Here, we see the effect PSP measurements have on prediction precision is farther reaching than for AGB (>1200 m)—also reflected in the esr_1 parameter estimate for the test model (Table 1).

To graphically demonstrate the test model's ability to account for uncertainty resulting from temporal misalignment and the benefit of including multiple measurements at PSPs, Fig. 9 summarizes annual AGB predictions from 1972 to 2012 for 4 example PSPs with associated estimates of uncertainty. We see in all 4 PSPs (especially PSP 120 and 150) that prediction uncertainty increases (wider credible intervals, lower precision) with increasing time since measurement. Fig. 9a and d shows that PSPs 420 and 302 have better overall predictions (narrower credible intervals) across the time gradient than PSPs 120 and 150–a direct result of the number of measurements and hence information to inform PSP specific intercept and growth. At PSP 150 (Fig. 9c), the single measurement does inform AGB prediction in proximity to the measurement year; however, as expected, the paucity of periodic measurements results in low precision at the time interval extremes. Also, these wide credible intervals suggest that PSP 150 is not close to other PSPs that offer measurements at different values of t to inform prediction. Overall, 2012 prediction of AGB at all 4 example PSPs is higher than at the time of the most recent measurement, and growth curve uncertainty is smallest at PSPs with more measurements.

Fig. 10a and c helps elucidate the relationship between 2012 AGB and growth prediction uncertainty and the number of measurements at influential PSPs (i.e., nearby PSPs that have the most influence on prediction at unobserved locations.) Fig. 10a shows little discernible trend in box-plots summarizing the 8283 grid cell prediction CIWs subset by number of inventories at nearest PSP, suggesting that 2012 AGB uncertainty is not substantially affected by the number of repeat measurements on proximate PSPs. In contrast, the narrowing of CIWs after ~2–3 remeasurements at the nearest PSPs seen in Fig. 10c underscores the need for such information to define the space-varying slope parameter on t, i.e., AGB growth.

Fig. 10b and d shows grid cell prediction CIW box-plots subset by the number of years since last inventory at nearest PSP for 2012 AGB and growth, respectively. Neither figure provides strong evidence that time since remeasurement affects precision of prediction on its own, although Fig. 10d might show a slight trend that predictions near PSPs measured in the last four years have higher precision, on average, than predictions near PSPs with older measurements.

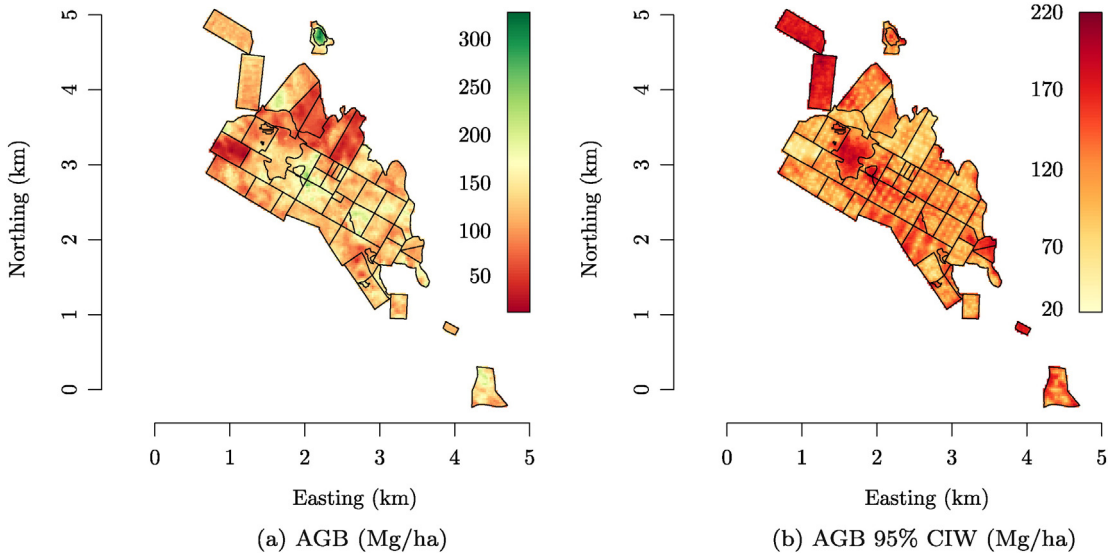

Fig. 6. Maps of predicted 2012 AGB (left) and associated 95% credible interval width (CIW) (right) using the test model. Black polygon boundaries outline different management units (MU) on the PEF.

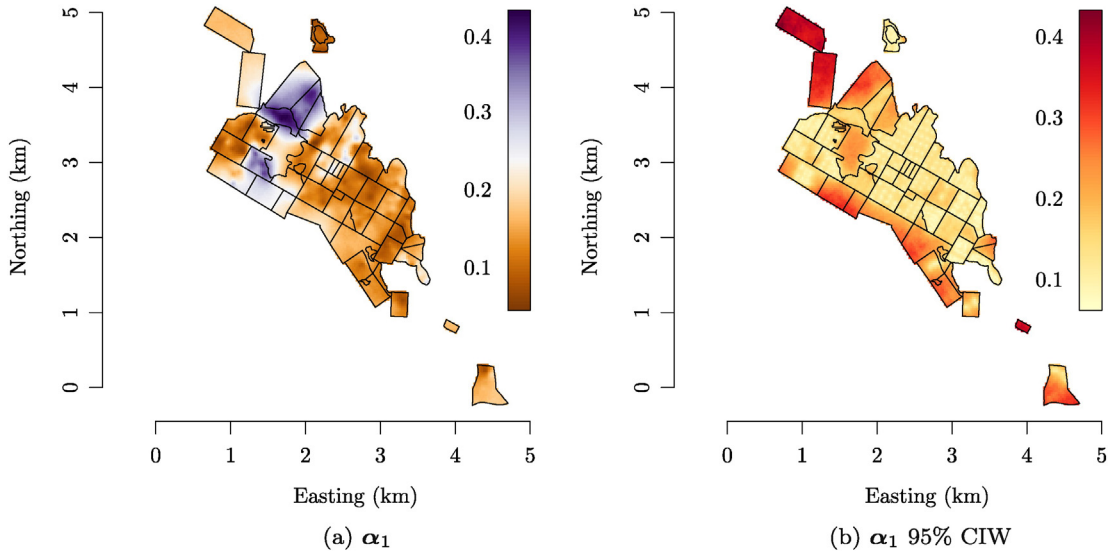

Fig. 7. Maps of predicted α_1 (growth) (left) and associated 95% credible interval width (CIW) (right) using the test model. Black polygon boundaries outline different management units (MU) on the PEF.

4. Discussion

4.1. Comparisons with other PEF studies

MU 8 and MU 22 were clearcut in 1984 and 1989, respectively (Fig. 1). We see the consequence of this in the map of predicted AGB (Fig. 6a) where both MU 8 and 22 are predominantly dark red, indicating comparatively low (<75 Mg/ha) AGB. Hayashi et al. (2014, Tab. 2) found MU 22 to have the lowest stem volume of the MUs they examined. We see slightly higher AGB growth rates in MU 8 compared to MU 22 in Fig. 7a.

These differences can be attributed to a time lag in stand *reorganization* (i.e., stand initiation) and *aggradation* (i.e., stem exclusion) phases (Oliver & Larson, 1996). Nyland (2007, pp. 204) points out the transition from the stand *reorganization* to *aggradation* phase, marked by the beginning of AGB accumulation, occurs approximately 20 years after clearcutting in this forest type and region. Since MU 22 and MU 8 were harvested 25 and 30 years ago, respectively, the differences in AGB growth might be attributed to the five-year difference in stand treatments and subsequent transition to the *aggradation* phase. Over the next decade, we expect a rise in AGB growth rates in both stands as they accrue biomass on ever larger trees until such a time that self-thinning processes start to reduce the growth of dominant trees (Oliver & Larson, 1996). MU 23 is undergoing a three-stage shelterwood management system where initial stand entries remove over-mature and undesirable trees to encourage the growth of advance regeneration present at the site. In an empirical study of the silvicultural treatments at the PEF Sendak et al. (2003, Fig. 4) found basal area growth to be fastest in this management system. In support of Sendak's finding, Fig. 7a shows high AGB growth rate in MU 23. MU 32 is a mature stand that serves as a reference treatment for the PEF. Only very limited management activity has occurred here since 1954. Fig. 6a shows that MU 32 has the highest accumulated AGB of any of the PEF's MUs and Fig. 7a indicates that AGB growth in MU 32 is currently very low—characteristics expected of mature, relatively undisturbed, old-growth stands in the *steady-state* phase of development (Nyland, 2007).

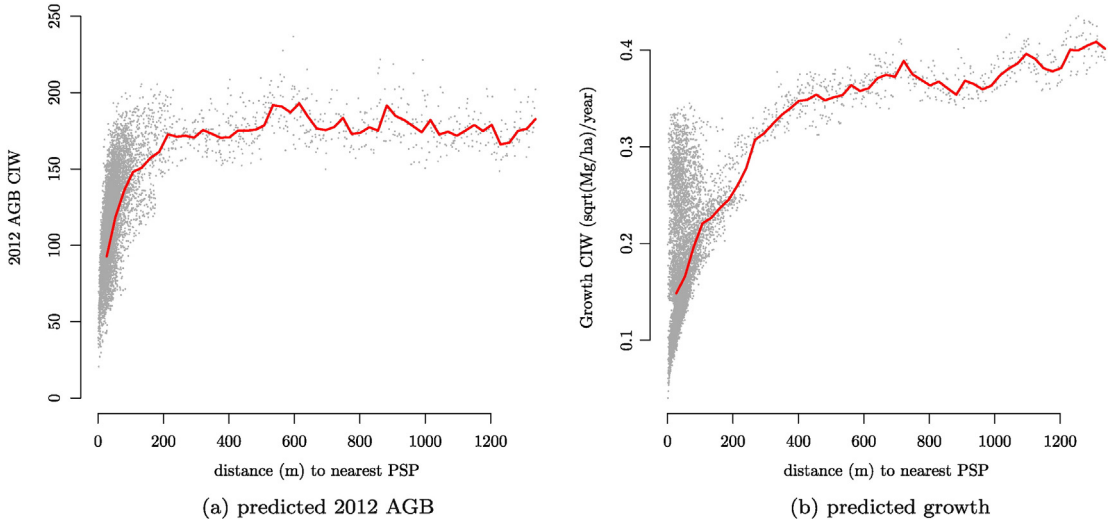

Fig. 8. 95% credible interval widths (CIW) for grid cell predictions of 2012 AGB and growth versus distance to nearest permanent sample plot (PSP). The overlayed red trend lines show mean AGB CIW and mean $\alpha_1(\mathbf{s})$ CIW values binned for 10 m intervals in uncertainty-agb and uncertainty-pro, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

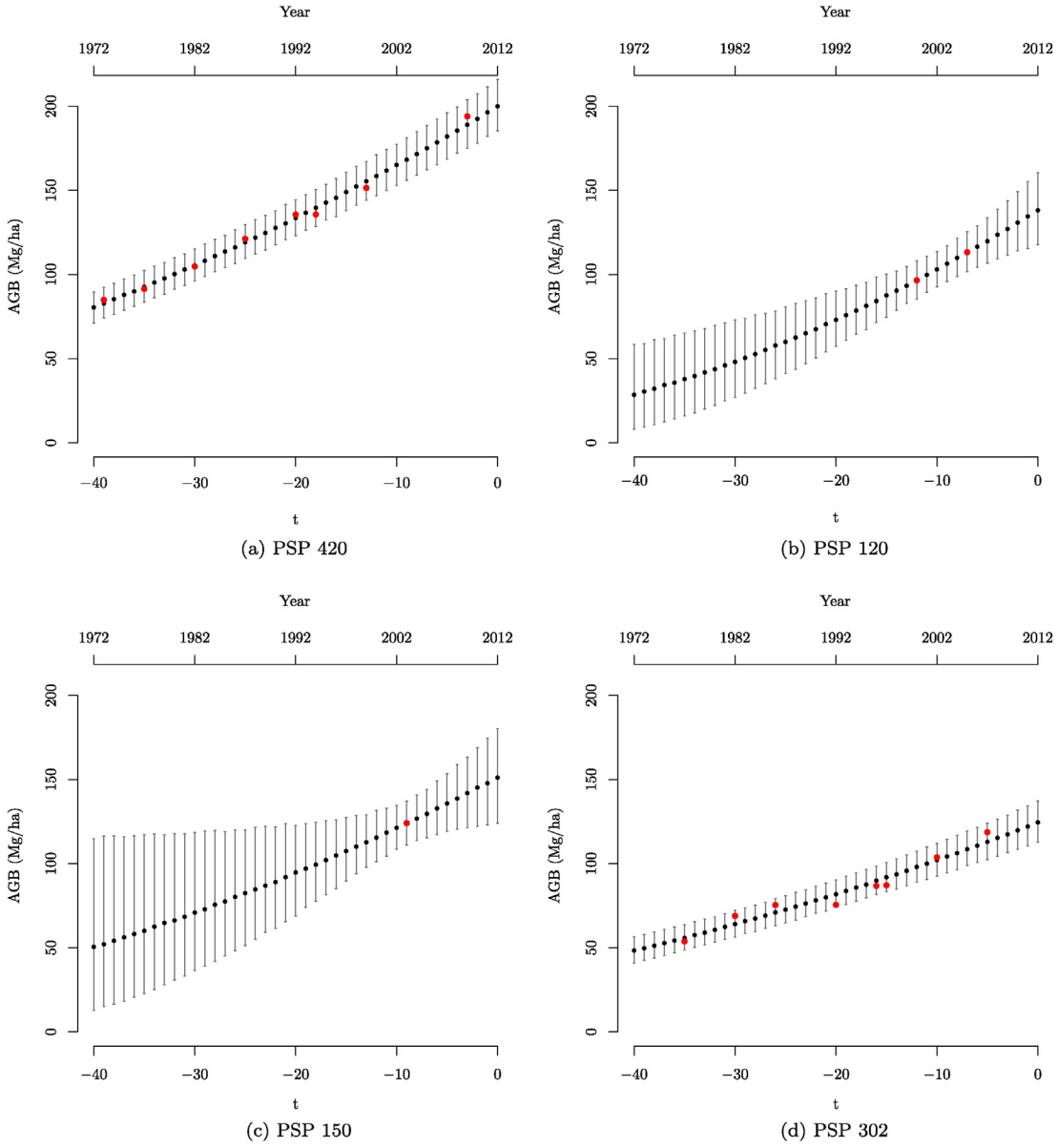

Fig. 9. Predictions of AGB from 1972 to 2012 for four example PSPs (black circles). Red circles indicate field inventory measures of AGB. Vertical gray lines highlight the 95% credible interval for the AGB predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Benefits of proposed framework

Results in Table 1 show the test model detailed in Section 2.3 provides greater prediction accuracy and precision compared to the benchmark models. These improvements are a direct result of appropriately modeling residual spatial dependence and capturing the space-varying nature of AGB change by leveraging repeated measurements. Other benefits of the proposed modeling framework are data driven estimates of AGB growth and full uncertainty quantification of parameters and predictions. The ability to estimate AGB and growth together and propagate their uncertainties through to prediction is clearly advantageous from an AGB monitoring and management perspective. As an example, scientists who are part of NASA's CMS initiative, in effort to advance the development of carbon MRV systems, need methods to spatially predict AGB with uncertainty. The modeling framework proposed here aims to help meet this goal.

The performance of the model introduced here provides evidence that temporal misalignment between LiDAR and field inventory data should be acknowledged during model fitting. The benchmark models' poor performance as compared to the test model, indicates the effects of temporally disjointed AGB and LiDAR measurements should not be treated as negligible even when the time discrepancies are ten years or less. This has implications for potential national scale mapping of AGB using air- or space-borne LiDAR and inventory systems, such as FIA. In most western U.S. states, only ten percent of FIA PSPs are measured annually, i.e., we would expect to encounter time disagreements of up to ten years when calibrating LiDAR with FIA inventory data. Results from this analysis suggest that future continental scale LiDAR informed AGB models using FIA inventory data may be improved by accounting for temporal misalignment.

One common solution to temporal disagreement between field inventory and remote sensing data is to apply growth models to “project” inventories forward to the time of LiDAR acquisition. A potential weakness of this approach is that uncertainty in growth model results are not easily propagated through statistical model components to prediction. Rather than attempting to correct for temporal misalignment in a

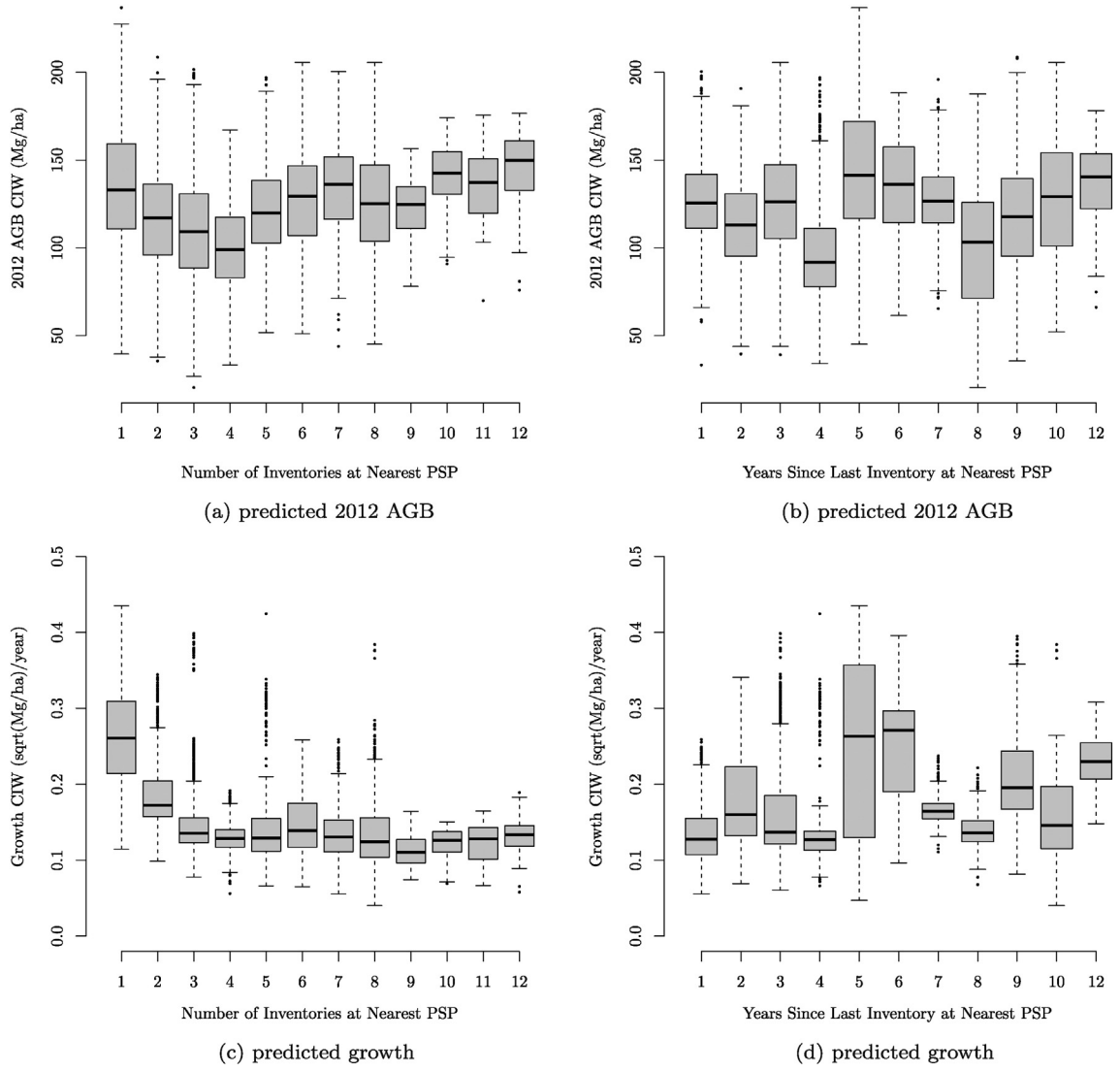

Fig. 10. Boxplots showing the distributions of 95% credible interval widths (CIW) for grid cell predictions of 2012 AGB (agb-box-num and agb-box-time) and productivity (pro-box-num and pro-box-time) versus number of remeasurements (agb-box-num and pro-box-num) and years since last remeasurement (agb-box-time and pro-box-time) at nearest permanent sample plot (PSP).

disjoint data preprocessing phase, it is more attractive to specify modeling frameworks capable of acknowledging misalignment explicitly and appropriately propagating resulting uncertainty.

Spatially and temporally explicit measures of uncertainty, a by-product of fitting Bayesian spatio-temporal hierarchical models, can help researchers determine when and where to collect additional inventory data given specific inferential goals. For example, if the goal is to increase growth prediction accuracy and precision at the PEF, Fig. 10c and d suggests remeasurement of PSPs with fewer than three measurements where the most recent measurement is at least four years old. Looking at Fig. 10c and d together suggests the number of remeasurements at a PSP may be more important than time since last inventory in determining growth uncertainty at the PEF. Fig. 10a and b indicates the number and timing of PSP remeasurements may not be a strong determining factor concerning AGB prediction uncertainty. Rather, Fig. 8a suggests that collecting new inventories further than 200 m from existing PSPs will improve AGB certainty more than any remeasurement effort. These types of questions cannot be addressed when RMSE alone is considered.

4.3. Limitations and indications for future research

The assumption of a linear increase of AGB over time as imposed by the first stage of the hierarchical model (1) is a potential drawback to this specific design's application to other forests. A linear trend was chosen to relate time and square-root transformed AGB as opposed to a sigmoidal growth curve that is commonly put forward for AGB growth rates. AGB growth on the PEF did not exhibit any strong curvilinear trends (Figs. 4 and 9) hence a linear trend was considered appropriate. Ongoing management at the PEF is likely keeping much of the forest in an active state of AGB growth. Areas are being harvested (or disturbed in some other fashion) before forest growth begins to slow in the senescence phase. Also, most PSPs had fewer than five post-disturbance measurements, which may not be enough to appropriately characterize a sigmoidal growth curve. Still further, the length of the temporal record at the PEF may not be long enough to expose non-linearity in the system. For datasets where a sigmoidal growth curve can be seen, the framework described here can be adapted. Rather than proposing a linear trend for (1), one could consider a non-linear AGB growth model.

The PEF PSP network is spatially dense with many PSPs having a long temporal record of inventories, which is a luxury not common in most forest monitoring systems. In the future, we will extend the proposed modeling framework to large scale sparse PSP networks, e.g., FIA, to estimate and map AGB growth. Previous modeling exercises provide some encouraging results. For example, in an effort to map forest biomass using FIA data over an area in northern Minnesota, Banerjee and Finley (2007) considered spatial regression models with a suite of Landsat derived covariates and residual structure captured using a Gaussian process capable of resolving short- and long-range spatial dependence. Their results showed effective spatial long-range dependence of ~7.9 km (with 95% credible interval from 4.1 to 11.1 km). In a different modeling exercise, aimed at mapping forest biomass over large heterogeneous landscapes (forest/non-forest areas), Finley et al. (2011) estimated residual spatial dependence patterns extending as far as ~352 km (with 95% credible interval from 290 to 429 km) across Michigan's lower peninsula. Datta, Banerjee, Finley, and Gelfand (2015) reported similar residual spatial dependence patterns when modeling AGB using FIA data across the conterminous US. All three studies used a model akin to benchmark model 2 and only the most recent FIA PSP measurements. Results from each study showed improved fit and/or holdout set prediction when leveraging residual spatial dependence through estimation of $\mathbf{w}_0$. Given the results of our current study, we anticipate including multiple measurements on FIA PSPs might further improve prediction at new locations and fetch additional insight into forest growth rates. Even if there is little residual spatial dependence—either due to effective covariates or simply a high degree of short range variability in the forest variable of interest—then $\mathbf{w}_0$ will not improve predictive ability through borrowing information from proximate PSPs. However, this does not negate the useful role of $\mathbf{w}_1$ to capture space varying growth rates.

Extensions to the proposed model could incorporate functional forms that use covariates in place of γ and $\mathbf{w}_1$. For instance, multispectral data may be useful for explaining variability in AGB growth rates across an expansive study region.

As we have demonstrated the benefits and knowledge gaps of adopting our proposed model for estimating forest AGB and associated AGB accumulation, the wider implications should not be understated. Global interest in mitigating the emission of carbon dioxide necessitates accurate estimates of terrestrial AGB and associated sequestration/emission rates. Rigorous combination of spatially sparse and infrequent in-situ measurements of forest AGB with potentially more spatially intense and frequent remote sensing products such as LiDAR offers the opportunity to refine our monitoring of the earth's terrestrial carbon balance. Beyond reducing uncertainty and informing field campaigns, our proposed model offers the opportunity to rapidly incorporate remotely sensed information following large-scale disturbance events (e.g., wildfires or hurricanes) to inform the distribution of carbon emission to specific disturbances (e.g., Kyoto Protocol removal of non-anthropogenic carbon emissions).

Acknowledgment

Data for this study were provided by a unit of the Northern Research Station, U.S. Forest Service, located at the Penobscot Experimental Forest in Maine. Significant funding for collection of these data was provided by the U.S. Forest Service (USFS 15-JV-11242307-116). Andrew Finley was supported by National Science Foundation (NSF) DMS-1513481, EF-1137309, EF-1241874, and EF-1253225, as well as NASA Carbon Monitoring System grants.

References

Abdalati, W., Zwally, H., Bindschadler, R., Csatho, B., Farrell, S., Fricker, H., ... Webb, C. (2010). The icesat-2 laser altimetry mission. *Proceedings of the IEEE*, *98*, 735–751.

Albercht, A., Hanewinkel, M., & Bauhus, J. (2012). How does silviculture affect storm damage in forests of south-western Germany? Results from empirical modeling based on long-term observations. *European Journal of Forest Resources*, *131*, 229–247.

Anderson, J. E., Plourde, L. C., Martin, M. E., Braswell, B. H., Smith, M. L., Dubayah, R. O., ... Blair, J. B. (2008). Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sensing of Environment*, *112*, 1856–1870.

Asner, G., Hughes, R., Varga, T., Knapp, D., & Kennedy-Bowdoin, T. (2009). Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems*, *12*, 261–278.

Awadallah, A., Abbott, V., Wynne, R., & Nelson, R. (2013). Estimating forest canopy height and biophysical parameters using photon-counting laser altimetry. *Proc. 13th international conference on LiDAR applications for assessing forest ecosystems (SilviLaser 2013)* (pp. 129–136).

Babcock, C., Finley, A. O., Bradford, J. B., Kolka, R., Birdsey, R., & Ryan, M. G. (2015). Lidar based prediction of forest biomass using hierarchical models with spatially varying coefficients. *Remote Sensing of Environment*, *169*, 113–127.

Babcock, C., Matney, J., Finley, A., Weiskittel, A., & Cook, B. (2013). Multivariate spatial regression models for predicting individual tree structure variables using lidar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*, 6–14.

Banerjee, S., Carlin, C., & Gelfand, A. (2014). Hierarchical modeling and analysis for spatial data. *Chapman & Hall/CRC monographs on statistics & applied probability* (2nd ed.). Chapman & Hall/CRC.

Banerjee, S., & Finley, A. (2007). Bayesian multi-resolution modeling for spatially replicated data sets with application to forest biomass data. *Journal of Statistical Planning and Inference*, *137*, 3193–3205.

Banerjee, S., & Johnson, G. (2006). Coregionalized single- and multi-resolution spatially-varying growth curve modelling with application to weed growth. *Biometrics*, *61*, 617–625.

Bechtold, W. A., & Patterson, P. L. (2005). The enhanced forest inventory and analysis program: National sampling design and estimation procedures. *US Department of Agriculture Forest Service*. North Carolina: Southern Research Station Asheville.

Blackard, J., Finco, M., Helmer, E., Holden, G., Hoppus, M., Jacobs, D., Lister, A., Moisen, G., Nelson, M., Riemann, R., et al. (2008). Mapping us forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, *112*, 1658–1677.

Blair, J., & Hofton, M. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters*, *26*, 2509–2512.

Brissette, J., & Kenefic (2008). Study plan: Silvicultural effects on composition. *Structure and growth of northern conifers in the Acadian forest region: Revision of the compartment management study on the Penobscot Experimental Forest*. Newtown Square, PA: USDA Forest Service, Northern Research Station.

Brissette, J., Kenefic, L., Russel, M., & Puhlick, J. (2012). *Overstory tree and regeneration data from the "silvicultural effects on composition, structure, and growth" study at Penobscot Experimental Forest*. Newtown Square, PA: USDA Forest Service, Northern Research Station.

Carlin, C., & Louis, T. (2008). Bayesian methods for data analysis. *Chapman & Hall/CRC texts in statistical science* (2nd ed.). Chapman & Hall/CRC.

Chen, Q., Vaglio Laurin, G., Battles, J. J., & Saah, D. (2012). Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass. *Remote Sensing of Environment*, *121*, 108–117.

CMS (2010). NASA carbon monitoring system. (http://carbon.nasa.gov. Accessed: 5-22-2014).

Cook, B., Corp, L., Nelson, R., Middleton, E., Morton, D., McCorkel, J., ... Montesano, P. (2013). NASA goddards lidar, hyperspectral and thermal (g-liht) airborne imager. *Remote Sensing*, *5*, 4045–4066.

Cowles, M., & Carlin, B. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, *91*, 883–904 (URL: http://www.jstor.org/stable/2291683).

Datta, A., Banerjee, S., Finley, A., & Gelfand, A. (2015). Hierarchical nearest neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*. http://dx.doi.org/10.1080/01621459.2015.1044091 (in press).

Finley, A., Banerjee, S., & MacFarlane, D. (2011). A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas. *Journal of the American Statistical Association*, *106*, 31–48.

Finley, A. O., Banerjee, S., & Gelfand, A. E. (2015). spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statstical Software*, 63.

García, M., Riaño, D., Chuvieco, E., & Danson, F. M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using lidar height and intensity data. *Remote Sensing of Environment*, *114*, 816–830.

GEDI (2014). Global ecosystem dynamics investigation lidar. (http://science.nasa.gov/missions/gedi/. Accessed: 1-5-2015).

Gelfand, A. E., Kim, H. J., Sirmans, C., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*, 387–396.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC: Chapman & Hall/CRC Texts in Statistical Science.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472.

Gregoire, T. G., Ståhl, G., Næsset, E., Gobakken, T., Nelson, R., & Holm, S. (2011). Model-assisted estimation of biomass in a lidar sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, *41*, 83–95.

Hayashi, R., Weiskittel, A., & Sader, S. (2014). Assessing the feasibility of low-density lidar for stand inventory attribute predictions in complex and managed forests of northern Maine, USA. *Forests*, *5*, 363–383.

Healey, S., Patterson, P., Saatchi, S., Lefsky, M., Lister, A., & Freeman, E. (2012). A sample design for globally consistent biomass estimation using lidar data from the geoscience laser altimeter system (glas). *Carbon Balance and Management*, *7*, 10.

ICESat-2 (2015). Ice, cloud, and land elevation satellite-2. (http://icesat.gsfc.nasa.gov/. Accessed: 1-5-2015).

Iqbal, I., Dash, J., Ullah, S., & Ahmad, G. (2013). A novel approach to estimate canopy height using icesat/glas data: A case study in the New Forest National Park, UK. *International Jorunal of Applied Earth Observation and Geoinformation*, *23*, 109–118.

Jenkins, J., Chojnacky, D., Heath, L., & Birdsey, R. (2003). National-scale biomass estimators for United States tree species. *Forest Science*, *49*, 12–35.

Kampe, T., Johnson, B., Kuester, M., & Keller, M. (2010). Neon: The first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *Journal of Applied Remote Sensing*, 4.

Kao, R. H., Gibson, C. M., Gallery, R. E., Meier, C. L., Barnett, D. T., Docherty, K. M., ... Schimel, D. (2012). NEON terrestrial field observations: Designing continental-scale, standardized sampling. *Ecosphere*, 3. http://dx.doi.org/10.1890/ES12-00196.1 (art115. URL: http://www.esajournals.org/doi/abs/10.1890/ES12-00196.1).

Le Toan, T., Quegan, S., Davidson, M., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., et al. (2011). The biomass mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, *115*, 2850–2860.

Michener, W., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. *Ecological Informatics*, *6*, 13–24.

Muss, J., Mladenoff, D., & Townsend, P. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, *115*, 824–835.

Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing*, *32*, 473–501.

Nathoo, F. S. (2010). Space-time regression modeling of tree growth using the skew-t distribution. *Environmetrics*, *21*, 817–833. http://dx.doi.org/10.1002/env.1057 (URL: http://dx.doi.org/10.1002/env.1057).

Neigh, C., Nelson, R., Ranson, K., Margolis, H., Montesano, P., Sun, G., ... Andersen, H. (2013). Taking stock of circumboreal forest carbon with ground measurements, airborne and spaceborne lidar. *Remote Sensing of Environment*, *137*, 274–287.

Nyland, R. (2007). Silviculture: Concepts and applications. *McGraw-Hill series in forest resources*. Waveland PressInc.

Oliver, C., & Larson, B. (1996). Forest stand dynamics. *Biological resource management*. Wiley.

Ometto, J. P., Aguiar, A. P., Assis, T., Soler, L., Valle, P., Tejada, G., ... Meir, P. (2014). Amazon forest biomass density maps: Tackling the uncertainty in carbon emission estimates. *Climatic Change*, 1–16.

Phillips, O. L., Malhi, Y., Higuchi, N., Laurance, W. F., Nez, P. V., Vsquez, R. M., ... Grace, J. (1998). Changes in the carbon balance of tropical forests: Evidence from long-term plots. *Science*, *282*, 439–442.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*, 9899–9904.

Sendak, P. E., Brissette, J. C., & Frank, R. M. (2003). Silviculture affects composition, growth, and yield in mixed northern conifers: 40-year results from the Penobscot Experimental Forest. *Canadian Journal of Forest Research*, *33*, 2116–2128.

Smith, W. (2002). Forest inventory and analysis: A national inventory and monitoring program. *Environmental Pollution*, *116*(Supplement 1), S233–S242.

Stow, C. A., Reckhow, K. H., & Qian, S. S. (2006). A bayesian approach to retransformation bias in transformed regression. *Ecology*, *87*, 1472–1477.

Talbot, J., Lewis, S. L., Lopez-Gonzalez, G., Brienen, R. J., Monteagudo, A., Baker, T. R., ... Phillips, O. L. (2014). Methods to estimate aboveground wood productivity from long-term forest inventory plots. *Forest Ecology and Management*, *320*, 30–38.

UN-REDD (2009). The UN-REDD Programme. (http://www.un-redd.org/. Accessed: 5-22-2014).

Weiskittel, A. R., Hann, D. W., Kershaw, J. A., Jr., & Vanclay, J. K. (2011). *Forest growth and yield modeling*. John Wiley & Sons.

Zhang, G., Ganguly, S., Nemani, R. R., White, M. A., Milesi, C., Hashimoto, H., ... Myneni, R. B. (2014). Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. *Remote Sensing of Environment*, *151*, 44–56.