Taylor & Francis
Taylor & Francis Group

Check for updates

Constrained Functional Regression of National Forest Inventory Data Over Time Using Remote Sensing Observations

Md Kamrul Hasan Khan[a], Avishek Chakraborty[a], Giovanni Petris[a], and Barry T. Wilson[b]

[a]Department of Mathematical Sciences, University of Arkansas, Fayetteville, AR; [b]Northern Research Station, USDA Forest Service, St Paul, MN

ABSTRACT

The USDA Forest Service uses satellite imagery, along with a sample of national forest inventory field plots, to monitor and predict changes in forest conditions over time throughout the United States. We specifically focus on a 230,400 ha region in north-central Wisconsin between 2003 and 2012. The auxiliary data from the satellite imagery of this region are relatively dense in space and time, and can be used to learn how forest conditions changed over that decade. However, these records have a significant proportion of missing values due to weather conditions and system failures that we fill in first using a spatiotemporal model. Subsequently, we use the complete imagery as functional predictors in a two-component mixture model to capture the spatial variation in yearly average live tree basal area, an attribute of interest measured on field plots. We further modify the regression equation to accommodate a biophysical constraint on how plot-level live tree basal area can change from one year to the next. Findings from our analysis, represented with a series of maps, match known spatial patterns across the landscape. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

1. Introduction

The national forest inventory (NFI) of the United States, conducted under the Forest Inventory and Analysis (FIA) program of the USDA Forest Service, is designed to provide consistent and unbiased strategic-level information about the status and trends of the Nation's forest resources (Bechtold and Patterson 2005). Given the geographic size and wide distribution of this population, the inventory is conducted annually using a probability sample with a base sampling intensity of one field plot per 2400 ha. An active area of forestry research involves the use of auxiliary data that can be collected quickly and inexpensively, such as from satellite imagery, to model the relationship between these auxiliary data and the field plot data in an effort to improve the precision of population estimates, particularly for smaller domains within the larger population.

Since satellite-based optical sensors, such as Landsat 7's ETM+ instrument, detect reflectance from the Earth's surface, these data are expected to be closely correlated with land cover. Kauth and Thomas (1976) developed a linear transformation of the original Landsat Multispectral Scanner bands, named the tasseled cap (TC) transformation. Comparable transformations have since been developed for the Landsat Thematic Mapper (Crist and Cicone 1984), Enhanced Thematic Mapper Plus, and Operational Land Imager sensors. The TC features are related to growing vegetation, soil moisture, and overall surface brightness, and are correlated with the phases of vegetation development over time. With the advent of the data policy of 2008, granting unrestricted access to the entire USGS archive of Landsat, dense time series of TC features can now be used readily and economically to model forest dynamics. Numerous studies have shown TC features derived from Landsat imagery to be useful for mapping land cover (Yuan et al. 2005), as well as several forest characteristics such as growing stock volume (Zheng et al. 2014) and biomass (Karlson et al. 2015). In the current work, live tree basal area, measured in square meters per hectare (m^2/ha), is the forest characteristics of interest. It is a simple measure of tree size and represents the cross-sectional area of the stem based on the diameter of the tree at a certain height, see Section 2 for details of FIA requirement for measuring basal area. It is closely correlated with other measures of tree size, such as volume and biomass, but does not require additional measurements like tree form, height, or wood density. While extracting the data for basal area from FIA database, filters are used within the SQL query to limit the data retrieval to measurements of live trees alone, excluding basal area of standing dead trees. We note that, in the rest of the article, any mention of basal area actually implies live tree basal area, even if not explicitly stated.

The goal of this article is to build a novel and flexible hierarchical model leveraging the relationship between the TC features, for which we have observed values for the majority of the population units at regular time intervals, and the basal

area data, for which we have relatively few observed values only from the FIA sample of field plots at certain years, to make annual predictions of the latter for every population unit. Our dataset comes from Landsat 7 ETM+ imagery, collected during the decade 2003–2012, from a 230,400 ha square-shaped region in north central Wisconsin. We also use a relatively small dataset of 146 pairs of basal area measurements from the same period, consisting of two 5-year FIA measurement cycles 2003–2007 and 2008–2012. This region is particularly interesting from a modeling perspective because, around middle of that decade, a tornado touched down near the eastern edge of the region, causing extensive damage to forest within a long, narrow swath, an example of natural canopy disturbance common within this region (Stueve et al. 2011). So, it is important to see if the proposed methodology can successfully reproduce the dynamic pattern along a tornado trajectory—undisturbed forest during the initial years, removal of live tree basal area due to the tornado and its gradual regeneration in the following years, as well as for more common anthropogenic disturbance like harvest.

We turn to highlighting the original contributions of this article. A motivating earlier work on hierarchical spatial predictive model for NFI data can be found in Finley, Banerjee, and MacFarlane (2011). The current work differs significantly from that in terms of the scope and objective of modeling as well as structure of the data. Emphasis of Finley, Banerjee, and MacFarlane (2011) lies in building a low-rank multivariate spatial process that is computationally efficient for predicting forest variables over a large landscape. On the contrary, we consider a univariate response, live tree basal area, over a relatively smaller region and focus on two key ideas. First, unlike the majority of studies that use remote sensing imagery either at a single point in time or as a composite of images over the study duration, we use the entire time series of monthly images over the decade of the study period. The temporal patterns of the TC features derived from this imagery, being correlated with land cover, are shown to be informative in distinguishing deciduous from evergreen forests, different tree species from one another, as well as predicting tree biomass (Wilson, Lister, and Riemann 2012; Wilson, Knight, and McRoberts 2018). Hence, in the regression equation for basal area, we propose to use the TC features as functional predictors (Ramsay and Dalzell 1991; Morris 2015). The resultant predictive model is used to depict the variation in live tree basal area across the region as well as to infer about changes between successive years during the period. To account for zero basal area measurements, we use, in the same vein as Finley, Banerjee, and MacFarlane (2011), a two-component mixture of a log-Gaussian functional regression model and a point mass at zero, with the mixture weights determined by a probit functional regression.

Our second key idea is to tweak this functional regression setting to accommodate an important biophysical constraint on the development of basal area; over two consecutive time intervals, a substantial decrease in live tree basal area is far more likely than a substantial increase. As explained later, this characteristic is not necessarily shared by the TC covariates, so the model needs to incorporate a correction factor. For this purpose, we introduce latent indicators that identify situations where such correction becomes necessary and encourages the model

stochastically to adapt in the proper direction. A Markov chain Monte Carlo (MCMC) scheme is developed for this setting and, by use of likelihood-based and predictive diagnostics, we demonstrate the benefit of incorporating the constraint inside the hierarchy.

Importantly, in our case, the stack of monthly TC feature imagery over the study period has many missing values both in space and time, due either to the failure of the scan line corrector or the presence of cloud cover. This issue does not arise in most other studies, since it is common practice to use either a single image with minimal cloud cover or a composite image constructed over a longer time period to fill in any gaps in the data record. Hence, before proceeding with the functional predictor regression mentioned above, first we have to fill in the missing parts of these images using a spatiotemporal regression.

The rest of the article is organized as follows. Section 2 describes the study area as well as the datasets of satellite imagery and NFI field plots. A constrained functional regression framework is developed in Section 3, for predicting variation in basal area using the series of TC feature imagery as covariates. In Section 4, the models developed in preceding section are implemented in R (R Core Team 2019) and we present outputs from the data analysis including results from model comparison and validation. Finally, Section 5 discusses some important aspects of our work and outlines the scope of possible extensions.

2. Data Description

Our region of interest lies in the southeastern part of the Web-Enabled Landsat Data (WELD) tile H20V05 (Roy et al. 2010), located near Langlade, Shawano, and Menominee counties in north-central Wisconsin, USA. Figure 1 shows its position within the tile and the state in the left panel. The right panel shows a variety of land covers and uses, such as agricultural fields (orange pixels), deciduous (purple) and evergreen (green) forests, developed land (gray), as well as scattered water bodies (white) and wetlands (pink). The study area falls within the Laurentian Mixed Forest Province of the USDA Forest Service National Hierarchical Framework of Ecological Units (Cleland et al. 2007; McNab et al. 2007) with the town of Antigo located in the northwest. The province experiences a continental climate, with some maritime influence from nearby Great Lakes (Superior and Michigan). This leads to moderately long winters and warm summers, when most of the precipitation occurs. This landscape was shaped by past glaciation with a mix of boreal and broadleaf deciduous forests, lakes and wetlands, along with grasslands that have since been converted to agricultural fields. The eastern half of the study area is predominantly forested. On June 7, 2007, a tornado, with estimated winds of 225–255 km/hr, touched down and followed a 65 km long and 0.75 km-wide northeasterly path through parts of Menominee, Langlade, and Oconto counties.

The auxiliary data used in the study were dense Landsat time series images from the WELD project. WELD imagery are composites of high fidelity data, determined on a pixel-by-pixel basis, from all Landsat 7 ETM+ scenes collected over the compositing period. These composite images have been processed
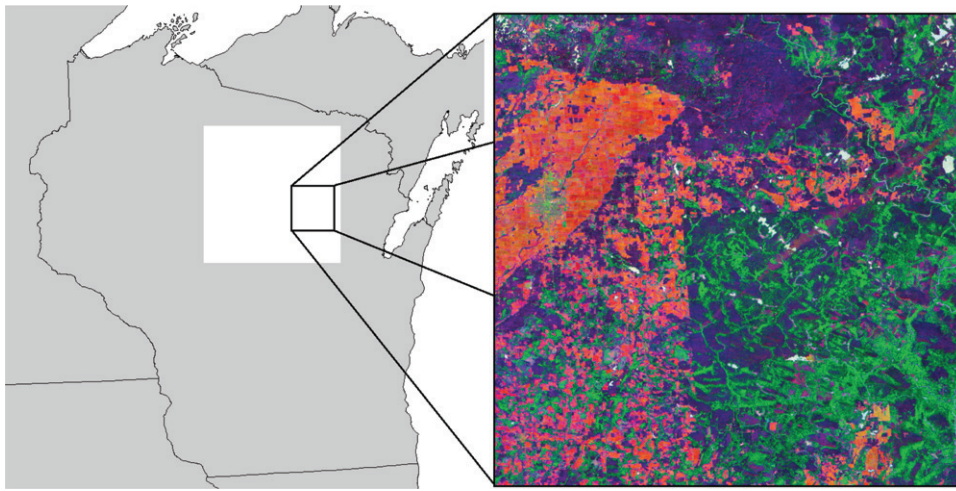
Figure 1. (Left) In the map of Wisconsin, the study area (the smaller square) is shown in the southeastern part of the WELD tile H20V05 (the larger square). (Right) Spatial variability of the land cover within the study area.
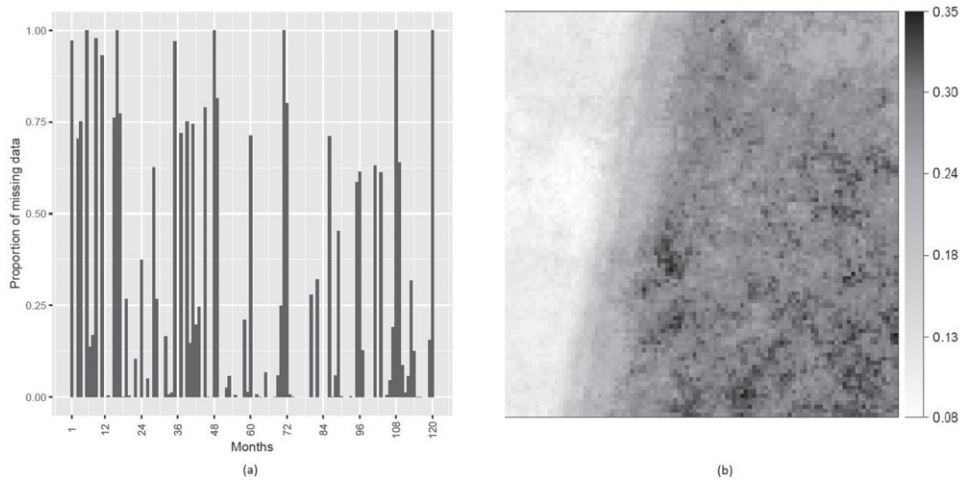
Figure 2. Proportion of missing TC observations within the study area during 2003–2012: (a) temporally, for each month and (b) spatially, for each gridcell.

for the contiguous United States and Alaska over the decade of 2003–2012. The composite scenes have been ortho-rectified, transformed to top-of atmosphere reflectance and mosaicked into 5000 × 5000 pixel tiles at the native 30-meter pixel resolution using the Albers Equal Area projection with origin at 23°N and 96°W. Our study area, a part of tile H20V05, consists of 1600 × 1600 pixels. The WELD monthly composites for the entire decade of 2003–2012 from this area were used for the study. For each monthly composite, the reflectance values from ETM+ were transformed to the first three TC components: brightness (TC1), greenness (TC2), and wetness (TC3) (Huang et al. 2002). The monthly TC features were then compiled into individual stacks. To control the size of the dataset, we aggregated every 16 × 16 adjacent pixels into a single gridcell of area 23.04 ha by taking the average of the pixels with available data. At this scale, our study area consists of 10,000 gridcells, an area 230,400 ha in size, Easting between 532,330 and 579,750 m, and Northing between 2,452,240 and 2,499,760 m. However, each TC component has a high frequency of missing values (∼23%) with 100% missing data for six months (April 2004, June 2003, November 2011, and December in 2006, 2011, and 2012), due either to completely missing records in the WELD archive or to

pixels that were masked out because of the presence of clouds, snow, or artifacts of sensor failure. Figure 2 represents the temporal and spatial patterns of missingness, in panels (a) and (b), respectively. The latter exhibits a strong linear boundary, with larger values in the east and smaller values in the west, due to the flight track of the descending polar orbit of the Landsat satellites. Areas to the west of the image are in the zone of overlap (sidelap) between neighboring Landsat scenes, while areas to the east have no overlap. Therefore, there are fewer pixel observations in the eastern portion of the image, meaning that it is more likely that data will be missing. Hence, to make the imagery complete, we need to perform a model-based filling-in of missing TC observations.

NFI data from the FIA program were collected on 146 field plots inside the study area during 2003–2012. An FIA field plot is a cluster of 4 circular subplots, each with a radius of 7.32 m, with the centers of three subplots located 36.58 m from the center of a central subplot, with one subplot due north of the central subplot, and the other two arranged by equal angles of 120°. All trees on the subplot having a diameter of 12.7 cm or larger at 1.3716 m above ground are measured. Additionally, each subplot contains a circular micro-plot with

a radius of 2.07264 m. All trees with a diameter between 2.54 cm and 12.7 cm at 1.3716 m above ground, which are defined as saplings, are measured on a micro-plot. For the purposes of this study, all trees measured on the micro-plot or subplot were used to compute the live tree basal area per hectare value for the plot. The plots are spatially distributed, each randomly located within a hexagonal tessellation of the landscape, with each hexagon being 2400 ha in size. In the case of Wisconsin, these hexagons were further subdivided in two, resulting in polygons 1200 ha in size. Each of these field plots was measured and then remeasured during that decade, spanning two 5-year FIA measurement cycles 2003–2007 and 2008–2012. Overall, in the dataset, positive live tree basal area measurements were recorded in 206 plots and a basal area measurement of zero was recorded in the remaining 86 plots indicating no basal area, according to FIA criterion, was found in those plots. These plot-level measurements were calculated by multiplying each condition's plot proportion by its corresponding basal area per hectare value, then summing across all conditions. The basal area measurements and TC features are linked spatially by matching the geographic coordinates of the centers of the field plots with the gridcells.

3. Model for FIA Study

Our objective is to use the TC feature imagery to predict the spatiotemporal variation in forest inventory measurements of live tree basal area during the decade. Live tree basal area is a nonnegative random variable; it takes positive values when live trees, large enough to meet the FIA's diameter threshold, are present on forested land, and is zero otherwise. Similar to Finley, Banerjee, and MacFarlane (2011), we propose a two-level hierarchical model, where the first level separates the occurrence of zeros and next level estimates the variation within nonzero measurements.

We begin with notations. The spatial domain of the study is denoted by $\mathcal{W}$ and is partitioned into disjoint areal units (or gridcells) $W_1, W_2, \ldots, W_S$ such that the feature imagery are available at the resolution of these areal units. Let $\mathcal{T}$ be the temporal domain and $X_s(\tau)$ represents the value of a feature variable at any gridcell W_s at time $\tau \in \mathcal{T}$. The dataset for each of these features consists of composite values over disjoint time intervals within $\mathcal{T}$, so we partition $\mathcal{T} = \bigcup_{t=1}^{T} D_t$. For the current application, TC features were reported monthly, so $D_1, D_2, \ldots, D_T$ represent consecutive months spanning $\mathcal{T}$. Define $X_{st} = \int_{\tau \in D_t} X_s(\tau) d\tau$ as the summary feature over W_s during D_t. As Section 2 indicates, some of these X_{st} values are missing and how to fill in those missing values is discussed in Appendix A.1 in the supplementary materials.

To link the TC feature values to basal area measurements at any gridcell, we construct a partition of $\mathcal{T}$ that has a coarser resolution than the partition $\{D_t : t = 1, 2, \ldots, T\}$. This can be achieved by collapsing every d adjacent time intervals of the former partition in a single unit. The level of aggregation d is determined by the user. The data analysis in this article uses $d = 12$, implying annual aggregation. It is adequate to model basal area at that time scale since reporting on forest resources is typically done on an annual basis, if not longer, because uncertainty can exceed annual change estimates. Consequently, we can write $\mathcal{T} = \bigcup_{k=1}^{T_y} \Delta_k$ where $T_y = T/d$ is the number of units in the coarsened partition $\{\Delta_1, \Delta_2, \ldots, \Delta_{T_y}\}$.

Let B_{sk} be the average basal area over time window Δ_k at gridcell W_s for $1 \leq k \leq T_y$ and $1 \leq s \leq S$. For the convenience of model development, we introduce a pair $(y_{sk}^{(1)}, y_{sk}^{(2)})$ corresponding to B_{sk}. $y_{sk}^{(1)}$ is a latent continuous variable whose negative values corresponds to the absence of basal area, implying $B_{sk} = 0$. When $y_{sk}^{(1)}$ is positive, B_{sk} is positive and is equal to $\exp\left(y_{sk}^{(2)}\right)$. It is adequate to propose models for $y_{sk}^{(r)}$, $r = 1, 2$, as B_{sk} can be written as a deterministic function of the pair as follows:

$$B_{sk} = 1(y_{sk}^{(1)} > 0) \exp\left(y_{sk}^{(2)}\right), \tag{1}$$

where, for any event A, $1(A)$ represents the binary indicator that becomes 1 only when A occurs. Below, in Section 3.1, we discuss modeling of $y_{sk}^{(1)}$ and $y_{sk}^{(2)}$ using functional regression. Section 3.2 discusses incorporation of a stochastic constraint in the model for $y_{sk}^{(2)}$ to better resemble how basal area can change temporally. Derivation of posterior distributions, estimation using MCMC and inference using posterior predictive distributions are described in Section 3.3.

3.1. Functional Predictor Regression for Basal Area

Functional predictor regression involves regression of a scalar response on a set of functional covariates. This framework, discussed in Morris (2015), is adopted below for modeling of $y_{sk}^{(r)}$ for $r \in \{1, 2\}$. For simplicity of notation, we omit the superscript r for now and reintroduce it when necessary. Hence, y_{sk} is the continuous response at gridcell W_s at time window Δ_k that we want to model using the functional auxiliary data $\{X_s(\tau) : \tau \in \Delta_k\}$ of the TC features over the same time window. The functional regression model can be written as

$$y_{sk} = \beta_0 + \int_{\tau \in \Delta_k} X_s(\tau)\beta(\tau)d\tau + \epsilon_{sk}, \tag{2}$$

where β_0 is the intercept and $\beta(\tau)$ is the functional regression coefficient. The error term ϵ_{sk} has a zero-mean normal distribution, is uncorrelated in space and time, and, accounts for two sources of noise: (i) the pure error of the functional regression and (ii) the approximation error due to using the FIA measurement from a given month of a year as representative of the average value for that entire year.

Since each Δ_k consists of d adjacent units from $\{D_t : t = 1, 2, \ldots, T\}$, we can replace the indexing variable t for the finer partition with a unique pair (k, l) such that $\Delta_k = \bigcup_{l=1}^{d} D_{kl}$. As an example, if $d = 12$, for any month D_t in the finer partition, k represents the year it belongs to and l represents position of that month within that year. With this notation, we can rewrite (2) as

$$y_{sk} = \beta_0 + \sum_{l=1}^{d} \int_{\tau \in D_{kl}} X_s(\tau)\beta(\tau)d\tau + \epsilon_{sk}.$$

Now, we let $\beta(\tau) \equiv \beta_l$ if $\tau \in D_{kl}$. This amounts to two assumptions. First, the coefficients do not change within a single unit of the finer partition D_{kl}. This is reasonable as X-data are available at that resolution only. Second, the coefficients vary only based on l, the position of D_{kl} within Δ_k, and are independent of k. This is equivalent to assuming that the functional linear model does not have dynamic coefficients. If d is chosen as 12, these two assumptions together imply that coefficients are constant within a month and have an annual periodicity, that is, the weight of TC features from a specific month, in predicting the average basal area for that year, does not change from one year to the next. Consequently, we can simplify the model for y_{sk} as

$$y_{sk} = \beta_0 + \sum_{l=1}^{d} \beta_l \int_{\tau \in D_{kl}} X_s(\tau) d\tau + \epsilon_{sk} = \beta_0 + \sum_{l=1}^{d} X_{skl}\beta_l + \epsilon_{sk},$$

(3)

where X_{skl} corresponds to the definition of X_{st} with t replaced by the equivalent (k, l) pair as mentioned above.

Since we want to use all TC covariates in the model for y_{sk}, we extend (3) as

$$y_{sk} = \beta_0 + \sum_{\text{TC}=1}^{3} \sum_{l=1}^{d} X_{skl}^{(\text{TC})} \beta_l^{(\text{TC})} + \epsilon_{sk}.$$

(4)

Using matrix notation, we can simplify the expression in the right-hand side of (4) as $X_{sk}^T \beta$ where X_{sk} and β are two column vectors that concatenate X_{skl} values and β_l parameters, respectively, for all TC covariates and for $l = 1, 2, \ldots, d$. Now, we apply (4) separately to $y_{sk}^{(r)}$ for $r \in \{1, 2\}$ to hierarchically extend (1) as

$$B_{sk} = 1(y_{sk}^{(1)} > 0) \exp\left(y_{sk}^{(2)}\right), \quad y_{sk}^{(r)} = \beta_0^{(r)} + X_{sk}^T \beta^{(r)} + \epsilon_{sk}^{(r)},$$

$$r = 1, 2, \quad \epsilon_{sk}^{(1)} \sim \mathcal{N}(0, 1), \ \epsilon_{sk}^{(2)} \sim \mathcal{N}(0, \sigma^2),$$

(5)

where the variance of $\epsilon_{sk}^{(1)}$ is fixed at 1 for identifiability and $\mathcal{N}$ stands for univariate normal distribution. It is easy to see that marginalizing $y_{sk}^{(r)}$ for $r = 1, 2$ would result in a two-component mixture model for B_{sk} with the components being a point mass at zero and a lognormal regression; their weights are determined by a probit model. To handle the sparse data size relative to number of covariates in (5), we use a shrinkage prior for covariate effects, the horseshoe (HS) prior from Carvalho, Polson, and Scott (2009). The advantages of this prior are that it has (i) a heavy tail suitable to retain significant covariate effects and (ii) a spike to infinity at the origin that shrinks the insignificant covariate effects. Since we assign this prior separately on β coefficients of $y_{sk}^{(1)}$ and $y_{sk}^{(2)}$ models, for notational simplicity, we again omit superscript r in the description of the prior and reintroduce it in the discussion of posterior distributions. The usual hierarchical specification of the HS prior is as follows: if β_j is the effect of j-th component of X_{sk}, then

$$\beta_j \sim \mathcal{N}(0, \lambda_j^2 \zeta^2); \ \lambda_j \sim \mathcal{C}^+(0, 1); \ \zeta \sim \mathcal{C}^+(0, 1) \quad \text{for}$$

$$j = 1, 2, \ldots, p,$$

where $\mathcal{C}^+$ denotes the standard Cauchy distribution truncated to $\mathbb{R}^+$, λ_j is called the local shrinkage parameter (specific to β_j), ζ is called the global shrinkage parameter (common to all

components of β) and p is the number of entries in X_{sk}. For conjugacy of posterior distributions, we set the prior distributions as $\beta_0 \sim \mathcal{N}(\alpha_{\beta_0}, v_{\beta_0}^2)$, $\sigma^2 \sim \text{IG}(a_0, b_0)$, with fixed values of hyperparameters, where IG denotes the Inverse-Gamma distribution.

3.2. Incorporating a Biophysical Constraint on Temporal Change

Live tree basal area is related to the presence of tree cover. A forested landscape would be expected to accrue live tree basal area slowly, over a period of many years, without any sudden increases. This landscape might similarly experience a slow decrease in basal area as trees age, become unproductive, and die. However, a rapid decrease is also possible due to harvest, storm damage or other agents of forest change. Hence, in our setting, $y_{sk}^{(2)}$ can abruptly decrease from one time window to the next in a gridcell, but the opposite is far more unlikely.

Variation in the time series of satellite imagery, on the other hand, would be subject to a different set of constraints. Fluctuations in seasonal weather patterns would be expected to result in greater variability in both TC brightness and wetness due to snow and rain. TC greenness is highly correlated with tree foliage, not directly with the basal area of the stem, which is expected to be more variable both seasonally and annually. Since the models in the current study are dynamic, constraints must be placed on how these models transition from one time step to the next to follow known biophysical patterns. Hence, we are proposing to *rectify* the model for $y_{sk}^{(2)}$ in (5) such that it discourages sudden increase in the mean response at any gridcell compared to the previous year. However, no such adjustment is deemed necessary in the model for $y_{sk}^{(1)}$, binary indicator for presence/absence of basal area, because a sudden year-over-year change from zero to nonzero is possible when trees meet the FIA's diameter threshold for measuring basal area.

Below, we present the modified equation for $y_{sk}^{(2)}$ along with a discussion of how it achieves the above-mentioned purpose:

$$y_{sk}^{(2)} \sim \mathcal{N}(\mu_{sk}, \sigma^2), \ 1 \leq s \leq S, 1 \leq k \leq T_y,$$

$$\mu_{sk} = z_{sk}(\beta_0^{(2)} + X_{sk}^T \beta^{(2)}) + (1 - z_{sk})\mu_{s(k-1)},$$

$$z_{s1} = 1, \ z_{sk} \sim \text{Ber}\left(\Phi(\mu_{s(k-1)} - \beta_0^{(2)} - X_{sk}^T \beta^{(2)})\right),$$

$$2 \leq k \leq T_y,$$

(6)

where Φ and Ber denote the standard normal cumulative distribution function and the Bernoulli probability distribution, respectively. At the initial time window Δ_1, the mean response depends solely on the TC imagery from the same window as no information is available from past. For the second time window onward, we introduce a binary parameter z_{sk} to select from two different choices for the mean structure: one evaluated only from the explanatory variables at the current time window and the other being the mean response from the previous time window. Moreover, the probability of selecting an option increases as it gets smaller compared to the other choice and vice-versa. Hence, if the covariates suggest a large increase in μ_{sk} relative to the preceding interval, $z_{sk} = 1$ becomes unlikely implying μ_{sk} stays at $\mu_{s(k-1)}$. On the other hand, if the covariates suggest a decrease or a small increase, that is relatively more likely to be

accepted. Thus, $(1 - z_{sk})$ acts as a stochastic inhibitor of a sudden increase in the $y_{sk}^{(2)}$. The unconstrained specification of (5) can be viewed as a special case of (6) with $z_{sk} = 1$ for $1 \le k \le T_y$.

3.3. Posterior Estimation and Inference

Using the observed B_{sk} values, we first draw the model parameters a posteriori and then sample from the posterior predictive distributions for unobserved B_{sk} values. The MCMC steps for unconstrained regression model of (5) is discussed first. Subsequently, we present a detailed derivation of the posterior distributions from the constrained model of (6).

3.3.1. Unconstrained Model
The posterior distributions for β_0, $\{\beta_j : j = 1, 2, \dots, p\}$ and σ^2 are standard, due to conjugacy of their prior distributions. For the hyperparameters within the HS prior, we use the data augmentation technique of Makalic and Schmidt (2016) to rewrite them as follows:

$$\zeta^2 | \xi \sim \text{IG}(0.5, 1/\xi),$$
$$\xi \sim \text{IG}(0.5, 1),$$
$$\lambda_j^2 | v_j \sim \text{IG}(0.5, 1/v_j),$$
$$v_j \sim \text{IG}(0.5, 1), \; j = 1, 2, \dots, p.$$

Inverse-Gamma being a conjugate prior for variance parameters, sampling the posterior distributions becomes standard. The form of the posterior distributions are mentioned in Appendix A.2 in the supplementary materials.

3.3.2. Constrained Model
We expand the expression for μ_{sk} in (6) as follows:

$$\mu_{sk} = z_{sk}(\beta_0^{(2)} + X_{sk}^T \beta^{(2)}) + (1 - z_{sk})\mu_{s(k-1)}$$
$$= z_{sk}(\beta_0^{(2)} + X_{sk}^T \beta^{(2)}) + (1 - z_{sk})z_{s(k-1)}$$
$$(\beta_0^{(2)} + X_{s(k-1)}^T \beta^{(2)}) + (1 - z_{sk})(1 - z_{s(k-1)})\mu_{s(k-2)}$$
$$\vdots$$
$$= \sum_{j=1}^{k} \left[\prod_{i=j+1}^{k} (1 - z_{si}) \right] z_{sj}(\beta_0^{(2)} + X_{sj}^T \beta^{(2)})$$
$$= \sum_{j=1}^{k} G_{skj}(\beta_0^{(2)} + X_{sj}^T \beta^{(2)}), \tag{7}$$

where

$$G_{skj} = \left[\prod_{i=j+1}^{k} (1 - z_{si}) \right] z_{sj}, \; 1 \le j \le k, \tag{8}$$

with the convention that the empty product $\prod_{i=k+1}^{k} (1 - z_{si}) = 1$. As z_{sk} can be either 0 or 1, G_{skj} is also binary. Below, we prove that

Result 1. For fixed s and k, exactly one member of the binary sequence $\{G_{skj}\}_{j=1}^{k}$ is 1.

Proof. We can rewrite,

$$G_{skj} = \begin{cases} (1 - \max_{j+1 \le i \le k} z_{si})z_{sj} & \text{for } j = 1, 2, \dots, (k-1), \\ z_{sk} & \text{for } j = k. \end{cases}$$

Suppose $j^\star = \max\{j : z_{sj} = 1\}$. ($j^\star$ exists since $z_{s1} \equiv 1$.) Now, the possible cases are:

- For $j > j^\star$, $G_{skj} = 0$ because G_{skj} is a multiple of z_{sj}.
- For $j < j^\star$, $G_{skj} = 0$ because $\max_{j+1 \le i \le k} z_{si} = 1$.
- For $j = j^\star$, $G_{skj} = 1$ because if $j^\star = k$, $z_{sk} = 1$ and if $j^\star < k$, $\max_{j+1 \le i \le k} z_{si} = 0$.

□

Therefore, (7) can be rewritten as

$$\mu_{sk} = \beta_0^{(2)} + X_{s, \arg\max_{j \le k}\{G_{skj}\}}^T \beta^{(2)}. \tag{9}$$

Now, we discuss the posterior sampling of z_s. Let $z_s = (z_{s1}, z_{s2}, \dots, z_{sT_y})^T$ with $z_{s1} = 1$ for $1 \le s \le S$. Accounting for two possible states for each of the remaining $(T_y - 1)$ binary z_{sk} variables, the vector z_s can be any one of the 2^{T_y-1} combinations. For the current dataset, $T_y = 10$ implies 512 possible binary sequences. Suppose $\widetilde{z}_s$ denotes one such candidate combination for z_s. Given $z_s = \widetilde{z}_s$, we use (8) and (9) to construct the candidate mean response $\widetilde{\mu}_s = (\widetilde{\mu}_{s1}, \widetilde{\mu}_{s2}, \dots, \widetilde{\mu}_{sT_y})^T$. Let us define $\mathcal{H} = \{s : \exists \, k \text{ such that } B_{sk} > 0\}$, the subset of indices corresponding to areal units with at least one nonzero basal area observation. It follows that we can update z_s for $s \in \mathcal{H}$ using a multinomial distribution with the probability given by

$$P(z_s = \widetilde{z}_s) \propto \left[\prod_{\{k: \, B_{sk} > 0\}} \phi(y_{sk}^{(2)} | \widetilde{\mu}_{sk}, \sigma^2) \right]$$
$$\times \left[\prod_{k=2}^{T_y} \widetilde{p}_{sk}^{\widetilde{z}_{sk}} (1 - \widetilde{p}_{sk})^{1 - \widetilde{z}_{sk}} \right],$$

where $\widetilde{p}_{sk} = \Phi(\widetilde{\mu}_{s(k-1)} - \beta_0^{(2)} - X_{sk}^T \beta^{(2)})$ and ϕ denotes the normal density function. Once we update z_s, we can also update $\arg\max_{j \le k}\{G_{skj}\}$ for $1 \le k \le T_y$. Readily, $\beta_0^{(2)}$ and σ^2 can be updated from standard distributions similarly as in the unconstrained setting of Section 3.3.1. However, the updating step for $\beta^{(2)}$ is unlike the usual regression since $\beta^{(2)}$ appears in two places within the hierarchy: (i) in the expression of μ_{sk} and (ii) in the probit regression of z_{sk} for $2 \le k \le T_y$. Using the data augmentation approach of Albert and Chib (1993) for z_{sk}, $\beta^{(2)}$ can also be simulated from a multivariate Gaussian distribution. The simulation of hyperparameters of the HS prior follows exactly as in Section 3.3.1. Further details of posterior distributions are included in Appendix A.3 in the supplementary materials.

For $s \in \mathcal{H}$, z_s vector is sampled during each MCMC iteration that we use in (9) to determine μ_{sk} and subsequently generate missing $y_{sk}^{(2)}$ measurements from their posterior predictive distribution. On the other hand, for $s \notin \mathcal{H}$, z_{sk} does not appear in the likelihood for any k, so we follow the structure of (6)

to sequentially generate $y_{s1}^{(2)}, y_{s2}^{(2)}, \ldots, y_{sT_y}^{(2)}$ using the parameter values from the current MCMC draw. Additionally, using the parameter draws from the unconstrained functional regression for $y_{sk}^{(1)}$, we generate binary indicators for nonzero status of B_{sk}. Posterior samples of B_{sk} values at unobserved (s, k) combinations are set either at 0 or an exponential of corresponding samples of $y_{sk}^{(2)}$ depending on the indicators being 0 or 1, respectively.

4. Analysis of Basal Area Data

We begin our analysis by filling in the missing parts of TC imagery in Section 4.1. These complete images are then used as functional covariates for predicting basal area. In Section 4.2, we do likelihood-based as well as cross-validation analyses to explore whether use of the constrained approach from Section 3.2 leads to improved predictive performance. In Section 4.3, we compare the proposed method against some of the commonly used nonparametric approaches from FIA literature. Finally, in Section 4.4, we present maps related to spatiotemporal prediction of basal area from the proposed model.

4.1. Filling in Missing Landsat Imagery

We model TC features in logarithmic scale. In the original dataset, TC1 has strictly positive values whereas both TC2 and TC3 can be positive as well as negative. Prior to applying the logarithmic transformation, we first translate and rescale the measurements of TC2 and TC3 to the same range as that of TC1. Hence, in case of brightness, X_{st} in Section 3 represents natural logarithm of original values and, for greenness and wetness, it represents logarithm of linear transformations of original values. Some more technical material pertaining to model selection for X_{st} has been collected in the Appendix in the supplementary materials. Specifically, the interested reader will find mention of different candidate models for X_{st} in Appendix A.1.1 in the supplementary materials; the details about the MCMC sampler in Appendix A.1.2 in the supplementary materials; the graphical and numerical summaries used for model comparison and selection in Appendix A.1.3 in the supplementary materials. Once we identify the best performing model, we use that to fill in the missing TC values. For every such missing value, we plug-in the empirical median of the corresponding posterior predictive distribution and measure the uncertainty of prediction by empirical width of the corresponding 90% highest posterior density (HPD) interval. In Figure 3, we present 10-year averaged monthly maps of the complete TC imagery in the original scale of these measurements.

Both the spatial and monthly variation seen in the panels of Figure 3 can be explained by differences in land cover across the study area, as well as seasonality in vegetation and weather. TC1 is related to overall surface albedo. Snow-covered fields have larger albedo values than forests. This pattern is most
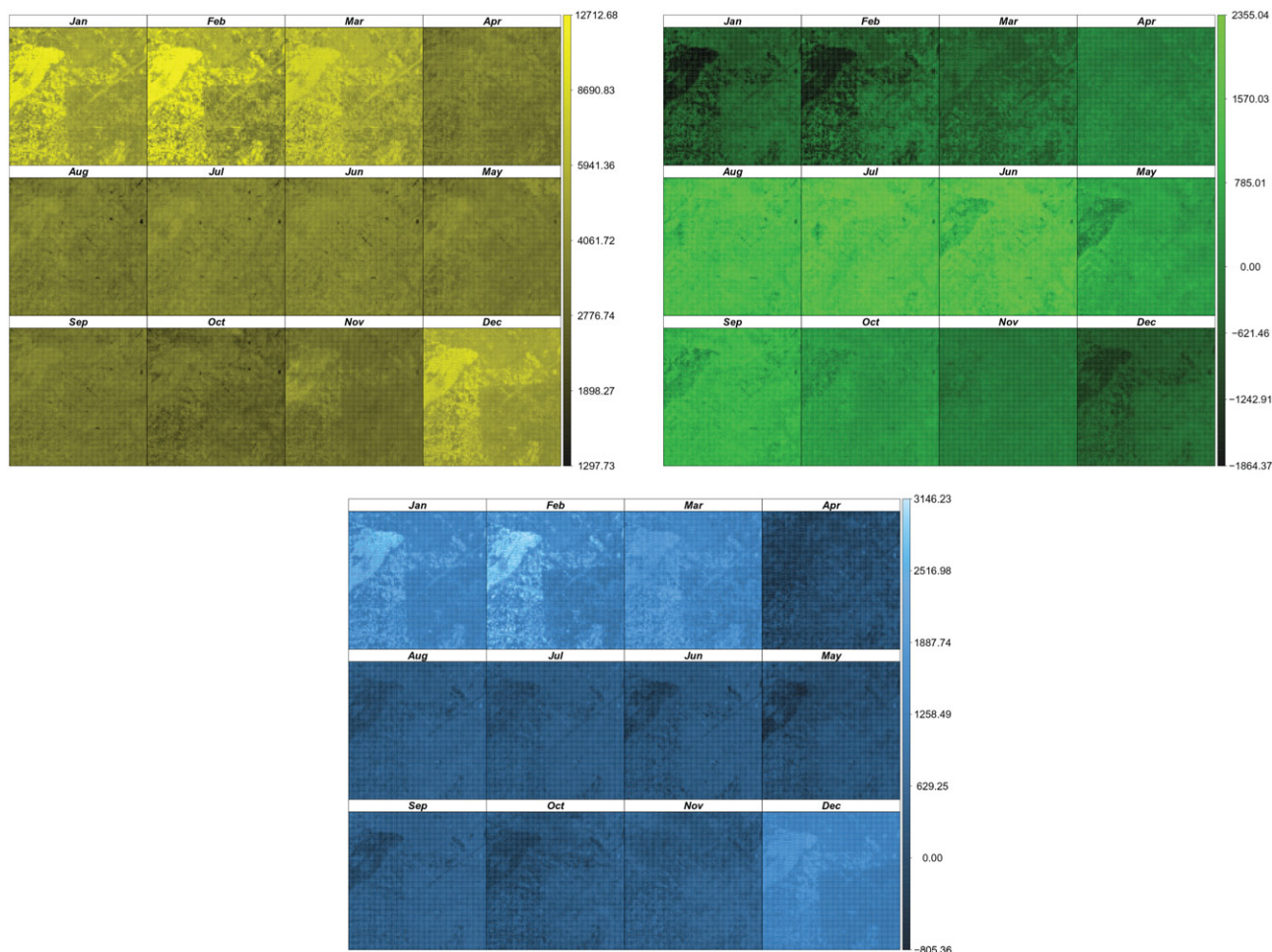

Figure 3. Monthly map of 10-year average TC features after filling in the missing values: (top left) TC1 brightness, (top right) TC2 greenness, and (bottom) TC3 wetness.

obvious for the TC1 panels during the winter months, with the western portion of the image being brighter than the eastern portion. The contrast in brightness between forests and agricultural fields is much smaller during the summer months, in the absence of snow. TC2 is related the amount of photosynthetically active vegetation on the land surface. During the winter months, agricultural fields have been previously harvested and are dormant. Similarly, deciduous trees have shed their leaves, while evergreen trees retain their needles. During the spring, fields are planted and crops begin to grow and deciduous trees leaf-out. Conversely in autumn, crops are harvested and deciduous trees begin to drop their leaves. These patterns are visible in the TC2 panels, with an overall increase in greenness during the spring and summer months. Also, the larger greenness values of evergreen trees stand out in the southeast potion of the image during the winter months. Finally, TC3 is related to soil moisture, as well as the presence of water and snow. As with brightness, wetness values are larger during the winter months. However, TC3 shows less contrast between forests and agricultural fields during the growing season for the study area, approximately April through October. Overall, these diagrams indicate that it is certainly possible for different land cover types to exhibit nearly identical values of TC brightness, greenness, or wetness at a specific month of an year, such as an agricultural field might be mistaken for a young stand of trees at the peak of the growing season. That reinforces the proposed approach in Section 3.1 for considering the TC features as functional covariates because if we consider the series of monthly TC values over an year, it is much less likely that two different land covers have nearly identical series of those values over the course of an entire year. For example, late in the autumn, the harvested agricultural field devoid of green vegetation will have a much different TC profile than a young stand of evergreen trees, or even a young stand of deciduous trees that exhibit a different rate of change in TC metrics across months.

It is also of interest to explore how percentages of missing data in a month influence predictive uncertainty of TC values at those missing locations. Figure 4 shows bar plots of the predictive uncertainty for all TC features. Here, we distribute the months in seven different groups (as shown in the diagram) based on the missing proportion of monthly TC data and then compute the average of 90% HPD interval width over missing cells across all months within each group.

Figure 4 indicates that, as expected, for all TC variables the uncertainty goes up with the increase in the percentage of monthly missing data and, cells with no available data in a month exhibit the highest uncertainty. Moreover, among these features, TC1 shows relatively larger predictive uncertainty than the rest. This is due to the fact that, in the transformed scale, the TC1 values are more evenly distributed across the range than TC2 and TC3 values, which are mostly concentrated within the upper half of the range. It should also be noted that this comparison of predictive uncertainty was carried out using the log-transformed TC values (the scale they were modeled at) instead of the original TC values. This was necessary because an exponential transformation, to take them back to the original scale, would make the uncertainty dependent on the mean, confounding the effect of extent of missingness with high and low values of these features.

4.2. Comparison of Constrained and Unconstrained Models

We start with a comparative evaluation of the two models for $y_{sk}^{(2)}$ with and without the constraint. The comparison is based on the 206 available nonzero basal area measurements. The MCMC is run for 60,000 iterations, discarding the first 10,000 iterations and thinning the rest at every 10th draw. Predictive performance under each model is examined using two different criteria. First, we calculate log-likelihood and Bayesian predictive information criterion (BPIC, Ando 2007; Li, Jun, and Zeng 2017). The model with the lowest BPIC is considered to be the best. Additionally, cross-validation is performed using the holdout method by repeatedly splitting the data into training sets and test sets, training the model on the former, and evaluating the predictive accuracy of the model on the latter. Every time we move 10 randomly chosen nonzero basal area observations to the test set and repeat the procedure 36 times such that all available observations are considered in the test set at least once and at most twice. We use three measures for assessment: (i) absolute bias, computed as the difference between the hidden test observations and their corresponding posterior medians; (ii) uncertainty, measured as the width of 90% HPD interval and (iii) empirical coverage, calculated as the proportion of test observations that are within the corresponding 90% HPD interval. These measures are averaged across replications. The
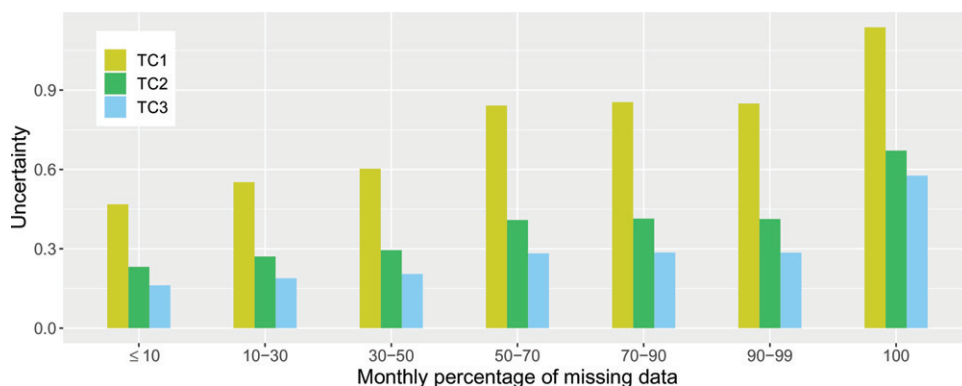

Figure 4. Averaged predictive uncertainty (90% HPD interval width) of missing TC values grouped by monthly percentage of missing data, in log-transformed scale.

Table 1. Influence of the constraint on the model for $y^{(2)}$.

Comparison statistic	Unconstrained model	Constrained model
Likelihood-based model comparison		
LL[a]	−3.969	39.093
BPIC	38.304	−107.254
Cross-validation		
Absolute bias[b]	8.620	8.389
Uncertainty[b]	31.529	45.184
Empirical coverage	0.828	0.917

[a] LL refers to posterior median of log-likelihood. [b] Bias and uncertainty are measured in m²/ha.
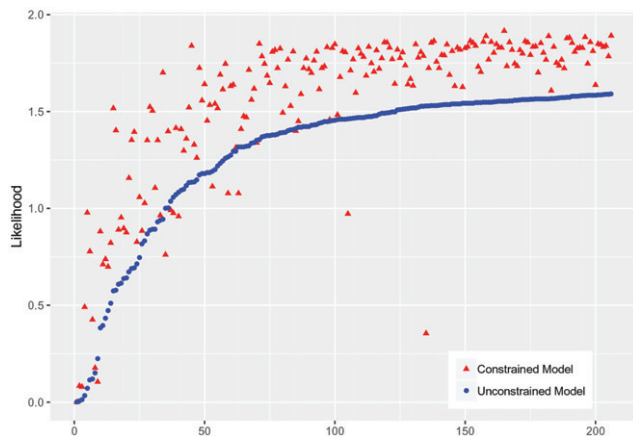
Figure 5. Observation-wise posterior median likelihood for $y^{(2)}$ with and w/o constraint.

model that attains the desired coverage level with minimum bias and uncertainty is deemed to be the best. The summary output is presented in Table 1.

Both the likelihood-based criterion as well as cross-validation show significant gains in the predictive performance for the constrained model over the unconstrained model. The former has a remarkably lower BPIC than the latter. Figure 5 represents posterior median likelihood values for each available $y_{sk}^{(2)}$ observation using both unconstrained and constrained models. It should be noted that, for the convenience of visualizing their difference, the observations are ordered according to increasing likelihood estimates from the unconstrained model. The figure clearly indicates an increase in likelihood for most observations following incorporation of the constraint. Furthermore, although the mean predictive uncertainty is lower in the unconstrained model (due to absence of the spatiotemporal sequence of indicator variables z_{sk}), only about 83% data lie inside the respective 90% HPD intervals. In contrast, the constrained model attains the desirable coverage with similar absolute bias, implying wider predictive intervals have actually been beneficial in this case.

Since the comparison measures show the constrained model as the superior one for modeling $y_{sk}^{(2)}$, we use it inside the hierarchical framework of (5). We also investigate the predictive accuracy of the model for $y_{sk}^{(1)}$ in identifying the zero and nonzero values of B_{sk} by computing area under the ROC curve (AUROC). We do this by running a cross-validation with a test set of 15 observations randomly chosen from all (zero and

nonzero) basal area observations and replicating the procedure 36 times such that all available observations are considered in the test set at least once and at most twice. The AUROC turns out to be 0.868 indicating satisfactory detection accuracy of zero and nonzero basal area plots. We use the above-mentioned test sets to simultaneously perform a predictive assessment for the combined hierarchical model for B_{sk}, consisting of unconstrained and constrained functional regression equations for $y_{sk}^{(1)}$ and $y_{sk}^{(2)}$, respectively. The predictive performance for the combined model is reported using three previously defined measures, averaged across replications: an empirical coverage rate of 93% for the 90% credible interval along with an average absolute bias of 7.787 m²/ha and average uncertainty of 37.104 m²/ha. Note that these numbers correspond to validation analysis of the combined hierarchical model for all zero and nonzero B_{sk} measurements whereas the numbers shown in Table 1 corresponds to the model for $y_{sk}^{(2)}$ applied on the subset of nonzero measurements only.

4.3. Comparison Against Existing Nonparametric Approaches

In the literature, there exist several nonparametric approaches that have been used for predictive mapping of forest attributes. These include random forests (Baccini et al. 2008), stochastic gradient boosting (Moisen et al. 2006), artificial neural network (Foody, Boyd, and Cutler 2003), generalized additive models (Frescino, Edwards, and Moisen 2001), nearest neighbors algorithm (Eskelson et al. 2009), and support vector regression (Chen and Hay 2011). Recently, convolutional neural networks, a deep learning method, have been used with high resolution imagery for forest classification and prediction of continuous forest variables (Chang et al. 2019). Shataee et al. (2012) gave a detailed comparison between some of these approaches in imputing several forest attributes such as tree volume, basal area, and number of stems using thermal and reflective bands of multispectral imagery. Moisen et al. (2006) also presented a comparative analysis for predictive mapping of basal area.

Below, we evaluate the proposed model's predictive performance relative to three competing approaches: support vector machine (SVM), stochastic gradient boosting (SGB), and generalized additive model (GAM), implemented in R using e1071 (Meyer et al. 2019), gbm (Greenwell et al. 2019), and gam (Hastie 2019) packages, respectively. In all of these methods, first the data on nonzero versus zero basal area are modeled using a binary regression and then, the data on logarithm of nonzero basal area measurements are fitted using a continuous regression. In GAM, the predictors are included in the model after smoothing with default parameters. While implementing these algorithms from above-mentioned packages, we retained their default settings as much as possible, see the files Nonparametric_Models_for_y1_cv.R and Nonparametric_Models_for_y2_cv.R in the supplementary materials for details.

To compare these models using cross-validation, we generated the test sets exactly as mentioned in Section 4.2. For classification of zero and nonzero test observations, we computed AUROC using probabilities predicted under each model.

Table 2. Model comparison against nonparametric methods.

Criterion	Proposed model	SVM	SGB	GAM
AUROC	0.868	0.854	0.845	0.796
Relative MSPE	0.794	0.812	0.894	1.679
Correlation	0.497	0.435	0.410	0.306

For nonzero basal area measurements, we compared the test response values against their point predictions. For SVM, SGB, and GAM, these predicted values were obtained using functions from above-mentioned packages. For the proposed Bayesian approach, empirical medians from respective posterior predictive distributions were used as predicted response values. We used them to calculate (i) relative mean square prediction error (relative MSPE), measured as the ratio of MSPE (average of squared differences between true and predicted basal area values) obtained using the model and MSPE obtained by naively using the mean of training data for prediction, and (ii) correlation coefficient between test responses and their predicted values. A smaller relative MSPE value (less than one) and a larger positive correlation coefficient are indicative of a better model fit. Table 2 displays those statistics, computed using true and predicted response values pooled from 36 replications of test set.

Table 2 shows that, across all criteria, the proposed model uniformly provides the highest predictive accuracy. SVM achieves the second best accuracy followed by SGB, and GAM is vastly outperformed in all cases. Moreover, none of these competing algorithms can easily be amended if one wants to incorporate the constraint on temporal change of basal area from Section 3.2. An added benefit of the proposed approach is that, for every point prediction, its reliability can readily be estimated using posterior predictive uncertainty, to be presented in the following section.

4.4. Posterior Predictive Analyses of Basal Area

Now that we have completed a detailed validation and comparison analyses for the proposed model, we move to the

spatiotemporal prediction of basal area. As mentioned in the end of Section 3.3.2, we collect posterior samples of B_{sk} values for all location-year combinations barring the ones with available basal area values. For the latter, we retained the original measurements. Figure 6 shows gridcell-wise posterior median of yearly average basal area values across the study area. The spatial and temporal patterns observed therein correspond well to known land cover and disturbances. The linear boundaries of Menominee County are quite visible in the southeastern portion of the image. This heavily forested county, with corresponding large basal area values throughout the years, is the least populous in Wisconsin and corresponds to the extent of the Menominee Indian Reservation. The land and forest management practices of the Menominee Tribe differ dramatically from those of neighboring communities. There are large tracks of undisturbed forest, with few and relatively small harvested areas. As mentioned before, the dynamic pattern of the model predictions along the tornado trajectory is of particular interest. The time series of basal area maps depicts a relatively stable landscape up to 2006. Since the tornado occurred around the middle of 2007, the yearly average basal area panel for 2007 shows its path with a lighter shade, reflecting an average of the dense forest and near-complete removal of live tree basal area before and after the event, respectively. The panels for 2008 to 2012 exhibit a period of gradual recovery along the tornado swath. We note that, at the chosen scale of aggregation (16×16 pixels), the exact tornado path is less than 2 gridcells wide, on an average, with a significant fraction of the swath partially encompassing the pixels along its edges.

Also of interest is the reliability of model prediction, quantified using the posterior uncertainty, the width of the 90% HPD interval. We present 10-year averaged uncertainty maps using absolute and relative scales in panels (a) and (b), respectively, in Figure 7. For the 146 gridcells that have two recorded basal area measurements each, uncertainty averages were calculated based on the posterior predictions for the remaining eight years. The relative uncertainty is computed using the ratio of uncertainty to the median basal area. Since very small values of median basal
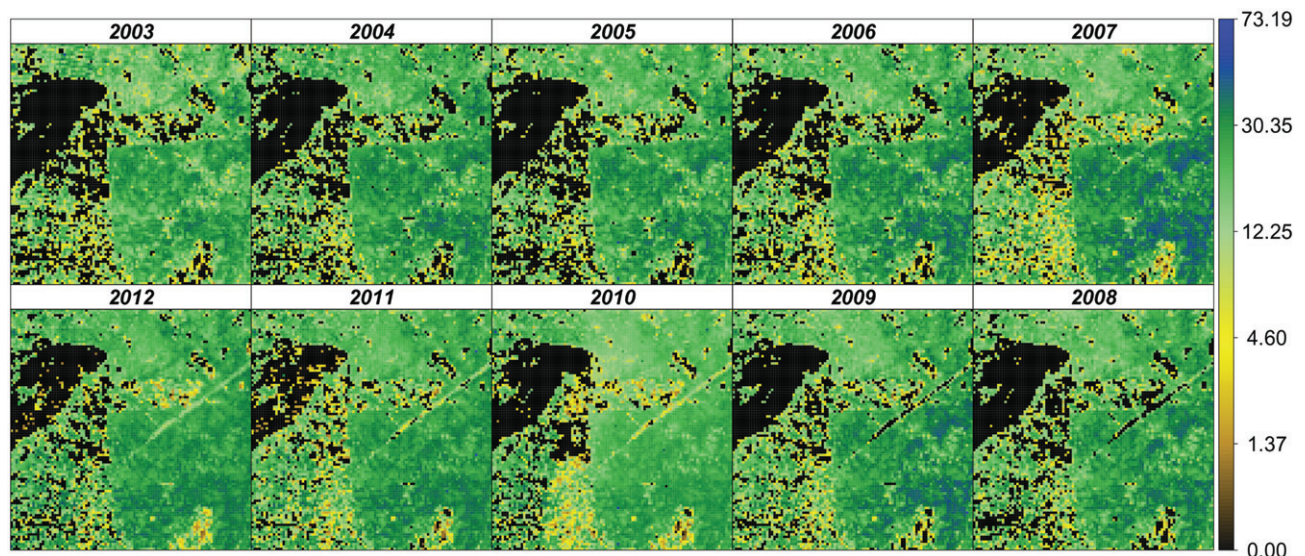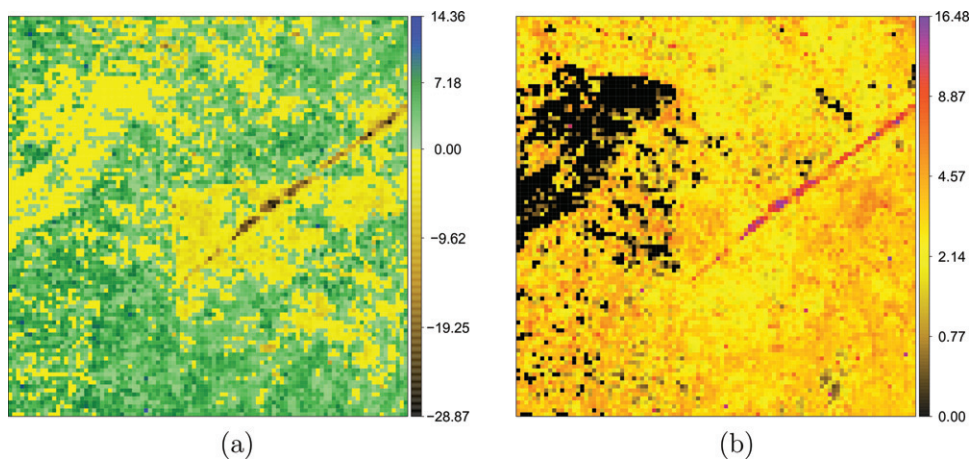

Figure 6. Posterior median of yearly average live tree basal area during 2003–2012 (in m²/ha). Panels are to be read year-wise left to right in the top row, then right to left in the bottom row.
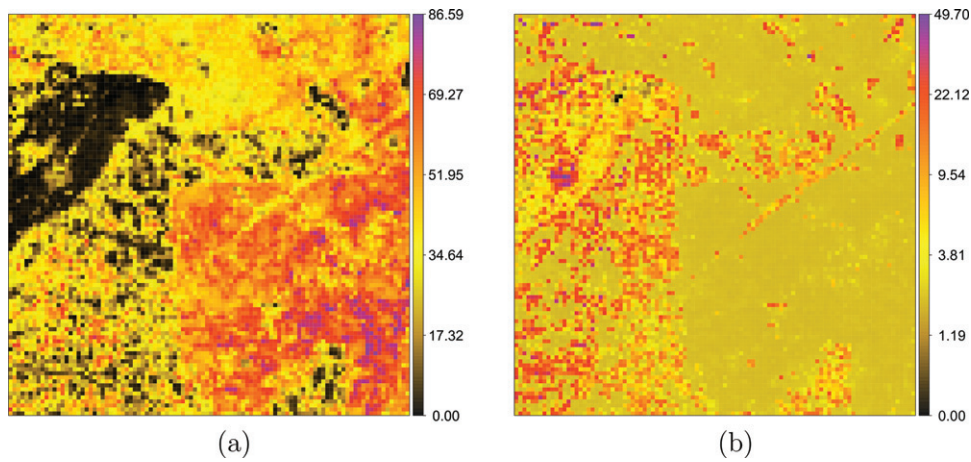
Figure 7. Maps of posterior uncertainty in absolute and relative scales: 10-year averaged (a) 90% HPD interval width (in m^2/ha) and (b) ratio of 90% HPD interval width and median for yearly average live tree basal area.

Figure 8. Maps of yearly variations (in m^2/ha) in posterior estimates: (a) change in estimated live tree basal area averaged over 2008–2012 compared to its average from the period 2003–2007 and (b) standard deviation of yearly average live tree basal area estimates.

area will lead to overly inflated values of relative uncertainty, we mitigate the inflation using a linear approximation of the ratio *only* for observations below a threshold of 0.001th quantile of the nonzero median basal area values. Panel (a) clearly shows heteroscedasticity in the uncertainty of the model predictions, with higher uncertainty associated with the dense forests of the Menominee Indian Reservation. This is a characteristic of the log-Gaussian distribution whose uncertainty increases proportionally with median. That is why, in relative terms in panel (b), the uncertainty appears to be essentially constant across those forested areas. The greatest relative uncertainty is associated with the town of Antigo, near the left edge of the figure, about two-thirds of the way from the bottom to top edge. This uncertainty is likely due to the pattern of tightly interspersed streets, lawns, and trees associated with urban and suburban settlements, which is exaggerated by the scale of aggregation (16 × 16 pixels), with resulting cells being predominately mixed due to the relatively smaller size of features in these land uses.

Now, we focus on exploring between-year variation in the posterior estimates of yearly average basal area during 2003–2012. Figure 8(a) depicts the change (can be positive or negative) in five-year average basal area during the remeasurement period of 2008–2012 relative to the its value during initial measurement period of 2003–2007. The impact of the tornado is readily

apparent in the image, with the initially large basal area values being reduced to zero directly along the tornado's path. The image also suggests a smaller reduction in basal area near the path, possibly due to damage from high winds. There are also some smaller areas to the southeast of the path that correspond with known harvest activity or regrowth from previous harvests, during the remeasurement period, that appear as brown/yellow and blue/green pixels, respectively. Standard deviation of 10 yearly average basal area values from each gridcell is plotted in Figure 8(b). As expected, the largest between-year variability is observed along the tornado trajectory that experiences near-complete removal of the live tree basal area around the middle of the study period and gradual regeneration in the following years. This variability is also due to the presence of gridcells along the edge of the tornado swath with mixed land covers. Agricultural fields around the town of Antigo mostly show no year-to-year variation, due to near-complete absence of any live tree basal area at those gridcells, as corroborated by Figure 6.

5. Conclusion

We have developed a hierarchical approach for utilizing time series of satellite imagery for dynamic prediction of live tree basal area. Through use of functional predictor regression, we

have established that the seasonal variation in these measurements provides significant information about live tree basal area distribution and can be used to complement the relatively sparse (in space and time) data collected from FIA plots. This represents a significant addition to the FIA modeling literature that has so far mostly considered remote sensing measurements for a single point in time or summarized as a composite image over a period of time. Augmenting one more level in the hierarchy, we adjust the regression equation to address the biased nature of temporal variation in basal area. The model fit is shown to improve significantly after this modification. It should be noted that, unlike the model for TC features, we did not use spatial and/or temporal random effects in the regression equation for basal area. That is due to the scarcity of available basal area measurements. At any given year, we have measurements for around 0.3% of the gridcells, that are significantly distant from each other (due to sampling design) across the study area. Temporally, for any gridcell, we have data for at most two of the ten years, separated by a period of around five years (the time difference between measurement and remeasurement in the same plot). In case of datasets that are significantly denser across spatial and temporal scales, inclusion of spatial as well as temporal random effects and associated gain in model fit can be explored.

The hierarchical structure developed for TC features (Appendix A.1 in the supplementary materials) as well as for basal area (Section 3) can be generalized to any region of interest with any spatial or temporal scale, even if different from ours. For example, if one wants to summarize the TC features seasonally or model the variation in basal area distribution at a multi-year scale instead of a year-to-year basis, that can easily be accommodated in the present setting. The work of Wilson, Lister, and Riemann (2012), using only a vegetation index from Moderate Resolution Imaging Spectroradiometer sensor, efficiently modeled forest conditions across a range of climatic and topographic conditions, as well as forest types. Given that the Landsat ETM+ sensor was used in the current study, we have not only a similar measure of the seasonality of growing vegetation (i.e., TC2 greenness), but also seasonal information related to surface albedo (TC1 brightness) and ground moisture (TC3 wetness). All of these TC components would be sensitive to the presence of trees, regardless of their environment.

It should be noted that the annual average basal area maps, presented in Figure 6, are based entirely on prediction from TC features, without using any information on regional distribution of forested and non-forest land. There is a fundamental limitation in determining forest land use (i.e., FIA plots with conditions defined as forest, where basal area measurements were taken) using remote sensing imagery alone, which reflects land cover. However, if one is interested in basal area prediction over forested land only, this can be obtained by limiting our posterior predictive analysis to gridcells on forested land, provided a binary forest land-use map is already available for the entire region.

There are certain practical sources of bias in this analysis that need to be discussed. As trees mature, the greenness is expected to increase. However, forests with closed canopy, after a certain level of growth, do not look very different from above even as they keep maturing, potentially causing the greenness to change

minimally. We argue that the proposed model is robust to this source of bias since it allows each month's greenness to have its own effect parameter on the yearly average basal area. As the growth and maturity of forests are linked to the time of the year, these effect parameters can also change accordingly. Another source of bias arises from the fact that FIA reports basal area from trees meeting a certain diameter threshold on forested land. As discussed in Section 2, the threshold varies between the subplots and the micro-plots within them, the latter being more accurate with a lower diameter threshold covering the saplings. Consequently, basal area from the young trees below that threshold or from trees outside the FIA's definition of forested land, if any, is not taken into consideration. Hence, in the data, (i) any plot with a reported basal area value of zero can potentially have some basal area from unaccounted trees and, (ii) a nonzero basal area measurement recorded in any plot may be lower than its actual value, due to presence of such unaccounted trees. Since TC values should be correlated with the basal area seen in a cell by the ETM+ sensor, inference based on this data can potentially underestimate the probability to have a nonzero basal area and the amount of basal area present, in a gridcell. Notably, in some instances, bias may also occur in the opposite direction when young trees get overtopped by the canopies of larger trees and are not detected by the ETM+ sensor. Furthermore, in this analysis, we ignore the uncertainty of the GPS measurement of the location of the FIA field plot as well as any georeferencing errors in the satellite imagery. As common with most of the existing literature on FIA modeling, we assume that these sources of error have negligible influence on the final inference. Accounting for them in the model requires making additional assumptions or using external information, and there exists scope of further research in this direction.

Now that we have established the feasibility of the proposed approach based on the association between basal area and TC features, its natural extension would be to scale up the analysis to the entire H20V05 WELD tile, about ten times as large as our study area. This will necessitate focusing on the computational aspects of the model. More specifically, the best candidate Model to fill in the missing values in the TC feature imagery involves spatiotemporal random effects, and, based on its runtime reported in Table A.1.1 of Appendix A.1.3 in the supplementary materials for the current study area, use of low-rank approximations need to be explored to control its computational cost in case of the entire tile. Finally, in addition to the TC features, auxiliary data on other factors that determine site productivity for trees, such as climate, landscape position and soil characteristics (Wilson, Lister, and Riemann 2012) can be utilized within this functional covariate setting to search for the most efficient and comprehensive predictive model for forest inventory.

Supplementary Materials

Appendix: Additional details on modeling and data analysis. (.pdf file)
Codes: All R codes (along with instructions) to run the models from this article and the appendices to generate included tables and figures. (.zip file)
Data: The TC dataset (.Rdata file) can be downloaded from the link included in the ACC form. The actual basal area dataset cannot be

shared due to the USDA Forest Service policy on data confidentiality. So, a simulated dataset of hypothetical basal area measurements (.csv file) is provided along with the code used to generate it.

Funding

References

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679. [1173]

Ando, T. (2007), "Bayesian Predictive Information Criterion for the Evaluation of Hierarchical Bayesian and Empirical Bayes Models," *Biometrika*, 94, 443–458. [1175]

Baccini, A., Laporte, N., Goetz, S., Sun, M., and Dong, H. (2008), "A First Map of Tropical Africa's Above-Ground Biomass Derived From Satellite Imagery," *Environmental Research Letters*, 3, 045011. [1176]

Bechtold, W. A., and Patterson, P. L. (2005), "The Enhanced Forest Inventory and Analysis Program-National Sampling Design and Estimation Procedures," Tech. Rep. SRS-80, US Department of Agriculture, Forest Service, Southern Research Station, Ashville, NC. [1168]

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009), "Handling Sparsity via the Horseshoe," in *Artificial Intelligence and Statistics*, pp. 73–80. [1172]

Chang, T., Rasmussen, B. P., Dickson, B. G., and Zachmann, L. J. (2019), "Chimera: A Multi-Task Recurrent Convolutional Neural Network for Forest Classification and Structural Estimation," *Remote Sensing*, 11, 768. [1176]

Chen, G., and Hay, G. J. (2011), "A Support Vector Regression Approach to Estimate Forest Biophysical Parameters at the Object Level Using Airborne Lidar Transects and QuickBird Data," *Photogrammetric Engineering & Remote Sensing*, 77, 733–741. [1176]

Cleland, D., Freeouf, J., Keys, J., Nowacki, G., Carpenter, C., and McNab, W. (2007), "Ecological Subregions: Sections and Subsections for the Conterminous United States," General Technical Report WO-76D. [1169]

Crist, E. P., and Cicone, R. C. (1984), "A Physically-Based Transformation of Thematic Mapper Data—The TM Tasseled Cap," *IEEE Transactions on Geoscience and Remote Sensing*, 3, 256–263. [1168]

Eskelson, B. N., Temesgen, H., Lemay, V., Barrett, T. M., Crookston, N. L., and Hudak, A. T. (2009), "The Roles of Nearest Neighbor Methods in Imputing Missing Data in Forest Inventory and Monitoring Databases," *Scandinavian Journal of Forest Research*, 24, 235–246. [1176]

Finley, A. O., Banerjee, S., and MacFarlane, D. W. (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. [1169,1171]

Foody, G. M., Boyd, D. S., and Cutler, M. E. J. (2003), "Predictive Relations of Tropical Forest Biomass From Landsat TM Data and Their Transferability Between Regions," *Remote Sensing of Environment*, 85, 463–474. [1176]

Frescino, T. S., Edwards, T. C., Jr., and Moisen, G. G. (2001), "Modeling Spatially Explicit Forest Structural Attributes Using Generalized Additive Models," *Journal of Vegetation Science*, 12, 15–26. [1176]

Greenwell, B., Boehmke, B., Cunningham, J., and GBM Developers (2019), "gbm: Generalized Boosted Regression Models," R Package Version 2.1.5. [1176]

Hastie, T. (2019), "gam: Generalized Additive Models," R Package Version 1.16.1. [1176]

Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, G. (2002), "Derivation of a Tasselled Cap Transformation Based on Landsat 7 at-Satellite Reflectance," *International Journal of Remote Sensing*, 23, 1741–1748. [1170]

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., and Mattsson, E. (2015), "Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest," *Remote Sensing*, 7, 10017–10041. [1168]

Kauth, R. J., and Thomas, G. (1976), "The Tasselled Cap—A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat," in *LARS Symposia, Paper 159*, available at *https://docs.lib.purdue.edu/lars_symp/159/*. [1168]

Li, Y., Jun, Y., and Zeng, T. (2017), "Deviance Information Criterion for Bayesian Model Selection: Justification and Variation," Working Paper, available at *https://ink.library.smu.edu.sg/soe_research/1927/*. [1175]

Makalic, E., and Schmidt, D. F. (2016), "High-Dimensional Bayesian Regularised Regression With the BayesReg Package," arXiv no. 1611.06649. [1173]

McNab, W., Cleland, D., Freeouf, J., Keys, J., Jr., Nowacki, G., and Carpenter, C. (2007), "comps. 2007. Description of ecological Subregions: Sections of the Conterminous United States [CD-ROM]," Gen. Tech. Report WO-76B, US Department of Agriculture, Forest Service, Washington, DC. [1169]

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2019), "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien," R Package Version 1.7-3. [1176]

Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., and Edwards, T. C., Jr. (2006), "Predicting Tree Species Presence and Basal Area in Utah: A Comparison of Stochastic Gradient Boosting, Generalized Additive Models, and Tree-Based Methods," *Ecological Modelling*, 199, 176–187. [1176]

Morris, J. S. (2015), "Functional Regression," *Annual Review of Statistics and Its Application*, 2, 321–359. [1169,1171]

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [1169]

Ramsay, J. O., and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Series B, 53, 539–572. [1169]

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., Loveland, T. R., Vermote, E., and Zhang, C. (2010), "Web-Enabled Landsat Data (WELD): Landsat ETM+ Composited Mosaics of the Conterminous United States," *Remote Sensing of Environment*, 114, 35–49. [1169]

Shataee, S., Kalbi, S., Fallah, A., and Pelz, D. (2012), "Forest Attribute Imputation Using Machine-Learning Methods and ASTER Data: Comparison of k-NN, SVR and Random Forest Regression Algorithms," *International Journal of Remote Sensing*, 33, 6254–6280. [1176]

Stueve, K. M., Perry, C. H., Nelson, M. D., Healey, S. P., Hill, A. D., Moisen, G. G., Cohen, W. B., Gormanson, D. D., and Huang, C. (2011), "Ecological Importance of Intermediate Windstorms Rivals Large, Infrequent Disturbances in the Northern Great Lakes," *Ecosphere*, 2, 1–21. [1169]

Wilson, B. T., Knight, J. F., and McRoberts, R. E. (2018), "Harmonic Regression of Landsat Time Series for Modeling Attributes From National Forest Inventory Data," *ISPRS Journal of Photogrammetry and Remote Sensing*, 137, 29–46. [1169]

Wilson, B. T., Lister, A. J., and Riemann, R. I. (2012), "A Nearest-Neighbor Imputation Approach to Mapping Tree Species Over Large Areas Using Forest Inventory Plots and Moderate Resolution Raster Data," *Forest Ecology and Management*, 271, 182–198. [1169,1179]

Yuan, F., Sawaya, K. E., Loeffelholz, B. C., and Bauer, M. E. (2005), "Land Cover Classification and Change Analysis of the Twin Cities (Minnesota) Metropolitan Area by Multitemporal Landsat Remote Sensing," *Remote Sensing of Environment*, 98, 317–328. [1168]

Zheng, S., Cao, C., Dang, Y., Xiang, H., Zhao, J., Zhang, Y., Wang, X., and Guo, H. (2014), "Retrieval of Forest Growing Stock Volume by Two Different Methods Using Landsat TM Images," *International Journal of Remote Sensing*, 35, 29–43. [1168]