Remote Sensing of Environment 212 (2018) 212–230

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Geostatistical estimation of forest biomass in interior Alaska combining Landsat-derived tree cover, sampled airborne lidar and field observations

Chad Babcock[a,*], Andrew O. Finley[b], Hans-Erik Andersen[c], Robert Pattison[c], Bruce D. Cook[d], Douglas C. Morton[d], Michael Alonzo[e], Ross Nelson[d], Timothy Gregoire[f], Liviu Ene[g], Terje Gobakken[g], Erik Næsset[g]

[a] School of Environmental and Forest Sciences, University of Washington, Seattle, WA, United States
[b] Forestry Department, Michigan State University, East Lansing, MI, United States
[c] USDA Forest Service, Pacific Northwest Research Station, Seattle, WA, United States
[d] Code 618, Biospheric Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD, United States
[e] Department of Environmental Science, American University, Washington, DC, United States
[f] Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, United States
[g] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway

ARTICLE INFO

Keywords:
Lidar
Landsat
Forest biomass
Bayesian hierarchical models
Gaussian process
Cokriging
Block kriging
Carbon monitoring
Linear model for coregionalization
Small area estimation

ABSTRACT

Lidar provides critical information on the three-dimensional structure of forests. However, collecting wall-to-wall laser altimetry data at regional and global scales is cost prohibitive. As a result, studies employing lidar for large area estimation typically collect data via strip sampling, leaving large swaths of the forest unmeasured by the instrument. The goal of this research was to develop and examine the performance of a coregionalization modeling approach for combining field measurements, strip samples of airborne lidar and Landsat-based remote sensing products to predict aboveground biomass (AGB) in interior Alaska's Tanana Valley. The proposed modeling strategy facilitates mapping of AGB density across the domain. Additionally, the coregionalization framework allows for estimation of total AGB for arbitrary areal units within the study area—a key advance to support diverse management objectives in interior Alaska. This research focuses on characterization of prediction uncertainty in the form of posterior predictive coverage intervals and standard deviations. Using the framework detailed here, it is possible to quantify estimation uncertainty for any spatial extent, ranging from point-level predictions of AGB density to estimates of AGB stocks for the full domain. The lidar-informed coregionalization models consistently outperformed their counterpart lidar-free models in terms of point-level predictive performance and total (mean) AGB precision. Additionally, including a Landsat-derived forest cover covariate further improved precision in regions with lower lidar sampling intensity. Findings also demonstrate that model-based approaches not explicitly accounting for residual spatial dependence can grossly underestimate uncertainty, resulting in falsely precise estimates of AGB. The inferential capabilities of AGB posterior predictive distribution (PPD) products extend beyond simply mapping AGB density. We show how PPD products can provide insight regarding drivers of AGB heterogeneity in boreal forests, including permafrost and fire, highlighting the range of potential applications for Bayesian geostatistical methods to integrate field, airborne and satellite data.

1. Introduction

Coupling remote sensing data with field-based forest measurements via regression frameworks offers the potential to increase the precision of inventory estimates and provides a mechanism for mapping the spatial distribution of forest biophysical properties. A plethora of studies show strong relationships between lidar metrics and forest variables (Asner et al., 2009; Babcock et al., 2013; Finley et al., 2014b, 2017; Lim et al., 2003; Næsset, 2004, 2011). These findings have spurred investment in collecting lidar data for large areas from aircraft and satellites alike. Of particular interest is the use of lidar to assist in the estimation of forest inventory parameters in high-latitude terrestrial ecosystems. From a carbon monitoring perspective, forests in boreal systems may contain large stores of aboveground biomass (AGB) and

* Corresponding author.
E-mail address: babcoc76@uw.edu (C. Babcock).

https://doi.org/10.1016/j.rse.2018.04.044
Received 20 December 2017; Accepted 24 April 2018
Available online 10 May 2018

carbon, but the uncertainty associated with current estimates is extremely high (Bradshaw and Warkentin, 2015; Pan et al., 2011). Understanding that the taiga-tundra ecotone is one of the most vulnerable environmental systems to climate change and that its boreal forests can contribute substantially to the global carbon cycle, methods are needed to begin monitoring forest carbon stocks and fluxes for these systems (Gauthier et al., 2015; Magnani et al., 2007; Neigh et al., 2013).

Current approaches used by the United States Forest Service's (USFS) Forest Inventory and Analysis (FIA) program to quantify AGB and carbon stocks in temperate regions rely on extensive, spatially-balanced field plot probability samples to generate forest inventory estimates with acceptable levels of precision (Bechtold and Patterson, 2005; Woodall et al., 2015). In vast remote landscapes, implementing the estimation techniques used by the FIA in the contiguous United States becomes prohibitively expensive due to the high cost of collecting field inventory data in difficult-to-access boreal regions, e.g., interior Alaska (Barrett and Gray, 2011). A potential solution commonly put forward to reduce the expense of monitoring AGB in boreal forest systems is to augment sparse collections of field samples with remote sensing auxiliary data (Wulder et al., 2012). Lidar-derived measures of forest structure tend to be highly correlated with AGB field observations and, thus, are prime candidates to supplement boreal field campaigns. Additionally, passive sensors such as Landsat can be used to derive remote sensing data products correlated with forest AGB (Kumar et al., 2015). Methodologies leveraging relationships between field and lidar can potentially be further improved by incorporating Landsat-based products (Margolis et al., 2015; Pflugmacher et al., 2014; Powell et al., 2010; Zheng et al., 2004).

Here, we address two challenges encountered when attempting to estimate forest AGB for large areas using lidar coupled with other remote sensing information: 1) incomplete spatial coverage of remote sensing data; and 2) prediction uncertainty quantification. Incomplete spatial coverage is a common problem for studies using airborne or spaceborne lidar over sizable study domains (Andersen et al., 2011; Bolton et al., 2013; Nelson, 2010; Nelson et al., 2004). Model-based methodologies used to link field and lidar data to estimate and map AGB typically require laser altimetry information for the entire spatial domain of interest (Babcock et al., 2015, 2016; McRoberts et al., 2013). The expansive nature of boreal systems, make wall-to-wall collections of airborne lidar data unrealistic. Further, future spaceborne lidar systems are not designed to procure complete coverage information. Rather, these campaigns will collect data for relatively narrow bands along the orbital tracts of the sensors' host satellite (GEDI, 2014; ICESat-2, 2015). In order to glean any additional information provided by sampled remote sensing data in a statistically rigorous manner, estimation frameworks that can accommodate incomplete coverage auxiliary information are necessary.

The second issue examined here is the problem of obtaining useful estimates of uncertainty about forest AGB stocks using model-based statistical procedures—necessary for decision making with imperfect predictions of forest AGB. In design-based estimation frameworks, error is assumed to arise from the sampling design, which can be appropriately characterized when plots are selected probabilistically (Cochran, 1977; Thompson, 2002). In model-based inference, error is attributed to the underlying process by which the response, e.g., AGB, is generated (Gregoire, 1998; Ver Hoef, 2002). Studies attempting to estimate means and totals for areal units using ancillary data within a model-based paradigm need to specify frameworks that reliably accommodate the structure of the data to be modeled. It can be the case that modelers who attempt to use model-based forest inventory estimation approaches posit potentially unrealistic assumptions about the distributional characteristics of model errors, such as independent and identically distributed (*iid*) errors. In a spatial context, it is likely the field observations of AGB will be spatially autocorrelated. If the auxiliary information used in the model fails to fully account for the spatial dependence among field observations, model-based approximations of AGB uncertainty can be grossly underestimated (Cressie, 1993; Griffith, 2005).

Coregionalization models constructed within a Bayesian hierarchical framework offer a solution to both above-mentioned challenges (Gelfand et al., 2004). This class of multivariate spatial regression models is designed to predict multiple response variables simultaneously while leveraging spatial cross-correlation structures between error components of the responses. Further, the model can accommodate spatial misalignment, i.e., missing response variable measurements at some locations. If the lidar data is treated as an explanatory variable (used on the right-hand side of the model as in most lidar studies) predictions are only possible where lidar data is available. Within a coregionalization model, the lidar-derived metrics can be treated as additional response variables (moved to the left-hand side) and jointly predicted with the response of interest, e.g., AGB, across the entire landscape while explicitly modeling the spatially co-varying relationship among the predictions within and across locations (Finley et al., 2014a). A coregionalization framework also allows for the inclusion of wall-to-wall covariates derived from satellite data to assist in the joint prediction of forest AGB and lidar information.

When multivariate coregionalization models are estimated using a Bayesian hierarchical approach, uncertainty occurring at all levels of the model can be propagated through to prediction and subsequent estimation of means and totals for areal units (Berliner, 1996; Cressie and Wikle, 2011; Gelfand and Smith, 1990; Hobbs and Hooten, 2015). Forms of multivariate spatial prediction models have been in existence since the 1960s, e.g., cokriging (Matheron, 1963). These non-hierarchical implementations, however, struggle to effectively deal with uncertainty associated with spatial covariance parameters, e.g., spatial variances and decays (Diggle and Ribeiro, 2007, Section 7.1.1). Due to increased computational efficiencies gained by ignoring uncertainty in spatial variability, 'plug-in' spatial covariance parameters are used in cokriging interpolation routines available in popular GIS software packages. This limits their use for fully model-based predictive inference (Schelin and Sjöstedt-De Luna, 2010).

The development of inferential approaches for complex spatial prediction within a statistical framework is an active area of research. In a hierarchical modeling context, coregionalization frameworks can be constructed using random effects that arise from spatially correlated Gaussian processes and partition variability into spatial and non-spatial components (Banerjee et al., 2014; Cressie et al., 2009). When formulated as such, estimation approaches including Restricted Maximum Likelihood (REML) or Markov chain Monte Carlo (MCMC) become possible in frequentist and Bayesian paradigms of statistical model-based inference, respectively (Ver Hoef et al., 2004). There are advantages to choosing a Bayesian hierarchical approach to inference over counterpart frequentist methods. Access to the full posterior predictive distribution (PPD), a by-product of Bayesian inference, allows for easy posterior summarization of means or totals with associated uncertainty for the full spatial domain in addition to any sub-domains that may be of interest–even under back-transformation (Stow et al., 2006). Access to PPDs facilitate subsequent, i.e., post-model-fitting, analysis to inform ecological or management objectives while accounting for prediction uncertainty. However, these increases in flexibility come with substantial increases in computational demand.

The aim of this study is to develop and examine the performance of a statistical modeling framework that can 1) incorporate partial coverage lidar data and wall-to-wall Landsat products to improve AGB density prediction; and 2) accommodate spatially structured variability unaccounted for by covariates, thereby allowing for more reliable model-based characterizations of uncertainty (e.g., uncertainty intervals with intended coverage) and improved prediction accuracy. We look to the Tanana Inventory Unit (TIU) in interior Alaska to explore the potential for the proposed coregionalization model to estimate forest AGB stocking by coupling spatially sparse field inventory, partial coverage lidar and Landsat-derived tree cover data products in boreal