Remote Sensing of Environment 190 (2017) 149–161

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Joint hierarchical models for sparsely sampled high-dimensional LiDAR and forest variables

Andrew O. Finley[a,*], Sudipto Banerjee[b], Yuzhen Zhou[c], Bruce D. Cook[d], Chad Babcock[e]

[a] Department of Forestry, Michigan State University, East Lansing, MI 48824, United States
[b] Department of Biostatistics, University of California, Los Angeles, CA, 90095-1772, United States
[c] Department of Statistics , University of Nebraska-Lincoln, Lincoln, NE, 68583-0963, United States
[d] Biospheric Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD 20742, United States
[e] School of Environmental and Forest Sciences, University of Washington, Seattle, WA 98195, United States

ARTICLE INFO

Article history:
Received 27 June 2016
Received in revised form 22 October 2016
Accepted 5 December 2016
Available online xxxx

Keywords:
Forest biomass
Uncertainty quantification
Functional analysis
Dimension reduction
Predictive process
Hierarchical models
Markov chain Monte Carlo

ABSTRACT

Recent advancements in remote sensing technology, specifically Light Detection and Ranging (LiDAR) sensors, provide the data needed to quantify forest characteristics at a fine spatial resolution over large geographic domains. From an inferential standpoint, there is interest in prediction and interpolation of the often sparsely sampled and spatially misaligned LiDAR signals and forest variables. We propose a fully process-based Bayesian hierarchical model for above ground biomass (AGB) and LiDAR signals. The process-based framework offers richness in inferential capabilities, e.g., inference on the entire underlying processes instead of estimates only at pre-specified points. Key challenges we obviate include misalignment between the AGB observations and LiDAR signals and the high-dimensionality in the model emerging from LiDAR signals in conjunction with the large number of spatial locations. We offer simulation experiments to evaluate our proposed models and also apply them to a challenging dataset comprising LiDAR and spatially coinciding forest inventory variables collected on the Penobscot Experimental Forest (PEF), Maine. Our key substantive contributions include AGB data products with associated measures of uncertainty for the PEF and, more broadly, a methodology that should find use in a variety of current and upcoming forest variable mapping efforts using sparsely sampled remotely sensed high-dimensional data.

1. Introduction

Coupling forest inventory with remotely sensed Light Detection and Ranging (LiDAR) datasets using regression models offers an attractive approach to mapping forest variables at stand, regional, continental, and global scales. LiDAR data have shown great potential for use in estimating spatially explicit forest variables over a range of geographic scales (Asner et al., 2009; Babcock et al., 2013; Finley et al., 2011; Iqbal et al., 2013; Muss et al., 2011; Næsset, 2011; Neigh et al., 2013). Encouraging results from these and many other studies have spurred massive investment in new LiDAR sensors and sensor platforms, as well as extensive campaigns to collect field-based calibration data.

Much of the interest in LiDAR based forest variable 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). In these, and similar initiatives, AGB is the forest variable of interest because it provides a nearly direct measure of forest carbon (i.e., carbon comprises ~50% of wood biomass, West, 2004). Most efforts to quantify and/or manage forest ecosystem services, e.g., carbon, biodiversity, and water, seek high spatial resolution wall-to-wall data products such as gridded maps with associated measures of uncertainty, e.g., point and associated credible intervals (CIs) at the pixel level. In fact several high profile international initiatives include language concerning the level of spatially explicit acceptable error in total forest carbon estimates, see, e.g., UN-REDD (2009) and UNFCCC (2015).

Many current LiDAR data acquisition campaigns focus on achieving complete coverage at a high spatial resolution over the domain of interest, e.g., resulting in a fine grid with each pixel yielding a high-dimensional LiDAR signal. In practice, a variety of non-statistical approaches are then used to characterize the LiDAR signals—effectively a dimension reduction step, Anderson et al. (2008), Gonzalez et al. (2010), Muss et al. (2011), Tonolli et al. (2011), Popescu and Zhao (2008), and Babcock et al. (2013). These signal characteristics serve as regressors in models where the outcome

* Corresponding author at: Department of Forestry, Michigan State University, East Lansing, MI 48824, United States.

http://dx.doi.org/10.1016/j.rse.2016.12.004
0034-4257/

forest variables are measured at a relatively small set of georeferenced forest inventory plots. The regression model is then used to predict the forest outcome variables at all LiDAR pixels across the domain. This approach works well for small-scale forest variable mapping efforts. However, next generation LiDAR acquisition campaigns aimed at mapping and quantifying variables over large spatial extents, such as ICESat-2 (Abdalati et al., 2010; ICESat-2, 2015), Global Ecosystem Dynamics Investigation LiDAR (GEDI) (GEDI, 2014), and NASA Goddard's LiDAR, Hyper-spectral, and Thermal (G-LiHT) imager (Cook et al., 2013; Stockton, 2014), will collect LiDAR data *samples* from the domain of interest, e.g., using transect or cluster designs. The designs specify point-referenced LiDAR sampling across the domain extent and also over forest inventory plot locations (again for regression model calibration). In such settings the primary objective is still delivery of high resolution wall-to-wall predictive maps of forest variables, but also corresponding maps of LiDAR signal predictions at non-sampled locations. Further, to inform future LiDAR collection sampling designs, there is interest in characterizing the spatial dependence of within and, more importantly, among LiDAR signals. This information can help guide LiDAR sampling strategies with the aim to maximize some information gain criterion; see, e.g., Xia et al. (2006), and Mateu and Müller (2012).

We propose a flexible framework to jointly model spatially misaligned LiDAR signals and forest inventory plot outcomes (e.g., AGB) that will *i*) automatically (i.e., no explicit variable selection step) extract information from the high-dimensional LiDAR signals to explain variability in the forest variable of interest, *ii*) estimate spatial dependence among and within LiDAR signals to improve inference and possibility help inform future LiDAR sampling strategies, and *iii*) provide full posterior predictive inference for both LiDAR signals and forest variables at locations where either one or neither of the data sources are available (i.e., wall-to-wall prediction).

Meeting these objectives is particularly challenging for several reasons. From a computational standpoint each LiDAR signal is high-dimensional and the signals as well as the forest inventory plots are observed at a potentially large number of locations. From a model specification standpoint there are several sources of dependence that should be accommodated, including *i*) within and between LiDAR signals, *ii*) between LiDAR signals and spatially proximate forest variable measurements, and *iii*) residual spatial dependence in the signals and forest variables. These dependencies often result from strong vertical and horizontal similarities in forest structure caused by past management and/or natural disturbances.

Our primary methodological contribution is the development of a modeling framework for high-dimensional misaligned data. Given the rich inference we seek (see preceding paragraph), our Bayesian hierarchical framework jointly models LiDAR signals and forest variables as a random process using latent Gaussian processes (GPs). This considerably enhances the computational burden of fitting them to datasets with a large number of spatial locations. The costs are exacerbated further by even a modest number of heights at which the LiDAR signal is observed. We achieve dimension reduction through bias-adjusted reduced-rank representations of the joint LiDAR-AGB process.

The manuscript is organized as follows. Section 2 provides an overview of the motivating dataset that comprises G-LiHT LiDAR and AGB measured at forest inventory plots on the Penobscot Experimental Forest (PEF) in Bradley, Maine. Section 3 describes the proposed hierarchical model for the joint LiDAR-AGB process. The details on Bayesian prediction and implementation are given in the Supplemental material. Section 4 offers an analysis of a synthetic dataset and PEF analysis. Finally, Section 5 concludes the manuscript with a brief summary and pointers toward future work.

2. Data

The PEF is a 1600 ha tract of Acadian forest located in Bradley, Maine (44° 52′ N, 68° 38′ W). The forest is divided into over 50 management units (MU)—delineated as black polygons in Fig. 1 (a)—that received management and monitoring since the 1950s (Sendak et al., 2003). Within each MU, different silvicultural treatments are implemented, e.g., unregulated harvest, shelterwood, diameter limit cutting, or natural regeneration. Following procedures described in Finley et al. (2014), AGB (Mg/ha) was calculated for each of 451 permanent sample plots (PSPs) across the PEF, shown as point symbols in Fig. 1 (a). The underlying surface in Fig. 1 (a) was generated by passing the point-referenced AGB through a deterministic surface interpolator. Due to MU specific harvesting and subsequent regrowth cycles, the surface exhibits patterns of spatial dependence with relatively strong homogeneity within MUs. For example, MU U7B—highlighted in Fig. 1 (a)—received a shelterwood harvest in 1978 with a final overstory harvest in 2003. This silvicultural treatment results in a MU with relatively young trees and even-aged composition with low AGB (indicated by a lighter surface color in Fig. 1 (a)). In contrast to U7B, C12 is characterized by older and larger trees, but also greater

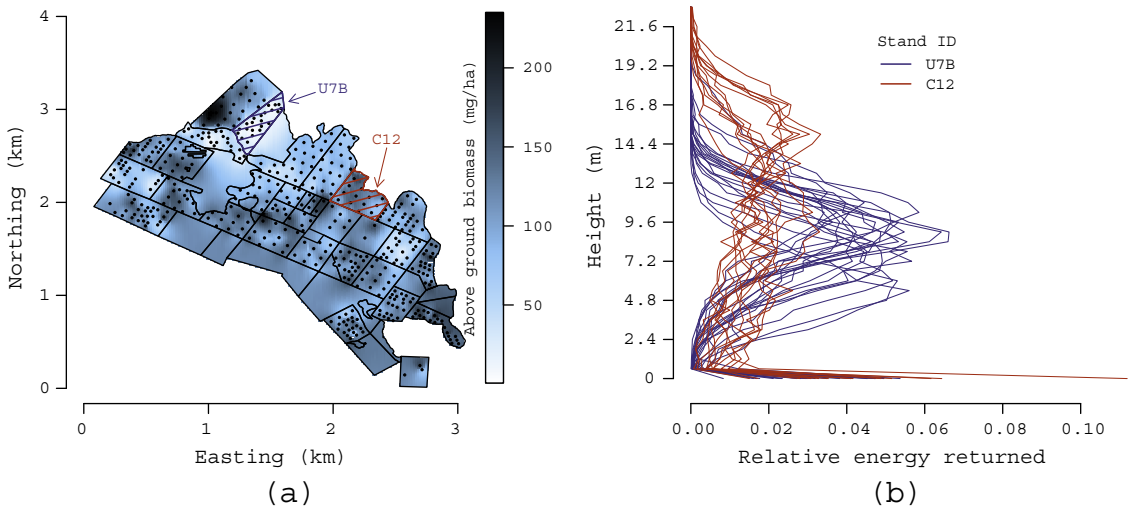

Fig. 1. (a) Penobscot Experimental Forest, Maine, with management units and forest inventory plot locations delineated as polygons and points, respectively. (b) G-LiHT LiDAR signals observed at forest inventory plots highlighted in (a).