Remote Sensing of Environment 154 (2014) 368–377

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Multivariate statistical analysis of asynchronous lidar data and vegetation models in a neotropical forest

Franklin B. Sullivan [a,*], Michael Palace [a], Mark Ducey [b]

[a] Earth Systems Research Center, Institute of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA
[b] Department of Natural Resources, University of New Hampshire, Durham, NH, USA

ARTICLE INFO

Article history:
Received 3 June 2013
Received in revised form 3 April 2014
Accepted 23 April 2014
Available online 4 June 2014

Keywords:
Multivariate analysis
Tropical forest change
Vegetation profile
Lidar

ABSTRACT

In this paper, we conduct multivariate analyses of similarity amongst lidar instruments and canopy models by exploiting the similarities of binned canopy height profiles to community datasets in that we treat profile bins as "species" to calculate distance matrices between sample units. Canopy profiles were derived from lidar data using the MacArthur–Horn transformation and from field data using a model we developed that uses two sets of allometric equations describing crown shape and tree height and a third from raw field data. We conducted a statistical comparison of seven asynchronous relative vegetation profiles (RVP) derived from different methodologies between the years of 2005 and 2012. We compared three airborne lidar datasets, three modeled profiles from field data, and one terrestrial lidar dataset using pairwise Mantel tests, multi-response permutation procedure (MRPP) and a permutation-based comparison using a similarity index for within plot comparisons. We used the results of MRPP to determine possible drivers of poor agreement between instruments and models and found a moderate relationship between within-plot variability (observed delta) and estimated above ground biomass ($r^2 = 0.273$, $p < 0.05$), which we attribute to poor model performance on low density/low biomass plots. In addition, correlation analysis of height metrics derived from RVPs resulted in weak correlations at low height percentiles and strong correlations at higher percentiles. Overall, we identified general similarity between lidar profiles using MRPP (lidar only $A = 0.202$, $p < 0.001$), but poorer agreement between lidar and modeled profiles (all profiles $A = 0.076$, $p < 0.001$). We attribute some of these differences to selection of canopy profile models and to approaches used for accounting for canopy occlusion in lidar transformations. In addition, we discuss the possibility of using Mantel tests to estimate temporal scales of vertical structure change in La Selva Biological Station, Costa Rica.

1. Introduction

Tropical forests are among the most structurally complex forests in the world (Whitmore, 1982). Tropical and neotropical forests are estimated to hold upwards of 60% of global terrestrial biomass stocks (Saatchi et al., 2011), and as such, quantifying spatial distribution, variability, and change in structure is crucial (Chave et al., 2004; Keller, Palace, & Hurtt, 2001). Field efforts to measure tropical forest structure and biomass at La Selva, Costa Rica have been prominent (Clark, 2002; Clark, Castro, Alvarado, & Read, 2004; Clark, Read, Clark, Cruz, Dotti & Clark, 2004; Treuhaft et al., 2010). Clark and Clark (2000) identified environmental controls, primarily soil characteristics and seasonal flooding, on forest structure by analyzing data from a network of 0.5 ha plots. Studies have shown that La Selva is anywhere between a net carbon sink (Phillips et al., 1998) or a slight source to carbon neutral (Clark, 2002). Although there is some ambiguity in the results of field studies on net primary productivity, relationships between forest structure and remote sensing data have proven robust (Asner et al., 2012; Clark, Castro, et al., 2004; Drake, Dubayah, Knox, et al., 2002; Palace, Keller, Asner, Hagen, & Braswell, 2008), allowing for spatially extensive aboveground biomass estimates.

Remote estimates of forest structure are a key advancement in forest management and carbon accounting (Frolking et al., 2009). These data have aided in quantifying standing biomass and biomass change across large spatial scales in regions where collecting data is difficult and time consuming or otherwise not possible (Means et al., 1999; Nelson, Short, & Valenti, 2004), especially in tropical settings (Clark, 2004; Clark, Castro, et al., 2004; Clark, Read, et al., 2004; Drake, Dubayah, Clark, et al., 2002; Drake, Dubayah, Knox, Clark, & Blair, 2002; Drake et al., 2003; Dubayah et al., 2010; Kellner, Clark, & Hofton, 2009; Kellner, Clark, & Hubbell, 2009; Treuhaft et al., 2010; Zhao, Popescu, & Nelson, 2009). In addition, efforts have been made to assess the ability of sensors to measure other structural characteristics, such as leaf area index, tree density, forest height and basal area (e.g. Chambers et al., 2007; Drake, Dubayah, Clark, et al., 2002; Drake, Dubayah, Knox, et al., 2002; Lefsky et al., 1999; Zhao et al., 2011). Light detection and ranging

* Corresponding author at: 462 Morse Hall ESRC, Durham, NH 03824, USA. Tel.: +1 603 862 4035.
E-mail address: franklin.sullivan@unh.edu (F.B. Sullivan).

http://dx.doi.org/10.1016/j.rse.2014.04.027
0034-4257/

(lidar) instruments flown aboard satellites and aircraft have been used to characterize the relative distribution of plant area within plots and footprints, based on the number or power of returns in assigned height intervals (Lefsky et al., 1999). Vertical complexity is thought to be an indicator of forest structure and heterogeneity, as evidenced by the use of metrics from power return distributions and relative vegetation profiles (RVPs) to estimate many structural characteristics, most notably biomass, based on empirical relationships with height quartiles, height of median energy and maximum return height (Drake, Dubayah, Clark, et al., 2002; Dubayah et al., 2010; Stark et al., 2012).

In addition to lidar data collected on aircraft, data have been collected using ground-based instrumentation. Terrestrial lidar has been used to estimate forest structure based on field measurements (Henning & Radtke, 2006; Hopkinson, Chasmer, Young-Pow, & Treitz, 2004; Yao et al., 2011), as well as to estimate RVPs (e.g. Hilker et al., 2012), similar to airborne lidar methods. Hilker et al. (2012) found good agreement between terrestrial and airborne lidar profiles in a thinned lodgepole pine forest; however, there have been fewer efforts to evaluate consistency between RVPs developed from terrestrial and airborne lidar in tropical forests.

Because of high costs and high demand of airborne lidar data collection, there have been few repeat flights of moderate- to high-spatial-resolution airborne lidar sensors over the same location. Establishing the ability of different sensors on different platforms to consistently measure vertical structure would aid in characterizing tropical forest change. La Selva Biological Station in particular provides a unique opportunity for a study of this nature. Here, we aimed to evaluate correspondence between two different forms of airborne lidar, a terrestrial lidar, and models based on allometric equations in dense tropical forest stands. Few studies have exploited the community-like data structure of canopy height profiles. In contrast to most canopy height profile comparisons, we conducted a series of statistical tests common in the field of community ecology to assess agreement between modeled canopy height profiles from field data, allometry, and four different lidar data sets collected using different methodologies over a 7 year period. Specifically, we used Mantel tests to characterize consistency of vertical structure variability across RVP types, and assessed the extent of agreement between RVPs within plots using multi-response permutation procedure (MRPP) and a more rigorous approach in which we measured Jaccard similarity and a permutation-based significance test.

Mantel tests have been used in other community and remote sensing research to measure correlations between, for example, spatial covariation of remote sensing metrics, plant composition, and bird habitat (e.g. Ranganathan, Krishnaswamy, & Anand, 2010), floristic patterns and climate variables such as rainfall and temperature (Häger, 2010) and canopy phenology (assessed by vegetation index time series from the Moderate Resolution Imaging Spectroradiometer) (MODIS) and plant biodiversity (e.g. Viña et al., 2012). MRPP, on the other hand, has been commonly used to evaluate grouping structure of species distributions based on cluster analysis or a priori group assignment. For example, Prado and Lewinsohn (2004) tested the apparent grouping structure of insect and plant associations in Brazil. In a remote sensing context, MRPP has been used to show that Normalized Difference Vegetation Index (NDVI) and species composition differed significantly across counties and ecoregions in North and South Carolina (He, Zhang, & Zhang, 2008). Lastly, permutation-based approaches have been applied to vegetation profile analysis previously. Drake, Dubayah, Knox, et al. (2002) used randomizations and an area of overlap test to assess differences between field and lidar profiles.

For this paper, we selected La Selva Biological Station as our study site because field and remote sensing data have been acquired frequently over a relatively short time-span. The high frequency of lidar data collection and availability of field data at this site compared to other candidate study sites made it the ideal location for a study of this nature. The primary objective of this study was to demonstrate the use of multivariate statistics commonly used in community ecology studies to assess relationships between RVPs. We sought to address this using asynchronous datasets collected using different methodologies by illustrating the use of the statistics for three secondary objectives: (1) to determine whether RVPs acquired using different airborne lidar systems were comparable; (2) to test if ground-based and airborne lidar RVPs were comparable; (3) to assess how well field data based RVP models and lidar derived RVPs correspond. Additionally, we demonstrate the utility of Mantel tests for lidar analyses and their potential uses for evaluating forest structural change. In addition, we conducted a correlation analysis using height metrics calculated from RVPs to compare our community-based results to traditional univariate approaches.

2. Methods

2.1. Study site and data sets

La Selva Biological Station is a tropical wet forest located at approximately 10.43°N and 84°W in the Atlantic lowlands of Costa Rica (Holdridge, Grenke, Haheway, Liang, & Tosi, 1975). La Selva has a mean annual rainfall of approximately 4300 mm and a mean annual temperature of 26 °C (Sanford, Paaby, Luvall, & Phillips, 1994). Within the forest, we randomly selected 20 sets of coordinates to establish as the center point of variable radius plots for field data collection and for terrestrial lidar data acquisition. Plot locations were constrained to <100 m of established trails, >50 m from rivers and water bodies, and >25 m from established study areas. Plots were geo-located using a high grade global position system. A site map is provided for reference in Fig. 1.

Field data were collected contemporary to terrestrial lidar scans (L4_TLS) in January 2012. On each plot, one TLS scan was collected at the established plot center. We used a FARO Focus 3D for our TLS scans, which has particular advantages for applications in forest structural measurements. First, the beam is relatively narrow (<5 mm at 50 m range). Beam width has been shown to be important for accurate reconstruction both of forest biometric properties and canopy attributes (Ducey et al., 2013). TLS returns numbered approximately 40 million per plot. For field sampling, variable radius plot sampling was employed to estimate vertical and overall stand structure (Bitterlich, 1984). Accordingly, overstory trees >10 cm diameter at breast height were tallied at plot center using a Spiegel–Relaskop with basal area factor 4 m^2 ha^{-1} and at four satellite plots 30 m from plot center in each of the cardinal directions. Total tree height, crown depth, and crown radius were measured for each tree included in the plot center tally. Satellite plots were used to estimate forest biometric properties using the basal area ratio approach (Marshall, Iles, & Bell, 2004). Biomass was estimated for each tree using general tropical biomass allometric equations developed by Brown (1997). Allometric equations were developed to estimate canopy shape of individual trees from a subsample of trees using a 16 m^2 ha^{-1} basal area factor. In order to generate RVPs from field data, canopy dimensions of sampled trees were estimated based on DBH using allometric equations from Asner et al. (2002) (M2_Asner) and derived from data collected for the current study (M3_LaSel), and plant area distributions for all sampled trees from a plot were aggregated into a single profile. In addition, RVPs were generated by aggregating profiles from plant area distributions based on direct measurements of canopy dimensions of sampled trees from this study (M1_raw).

Lidar data were collected using three different airborne instruments between the years of 2005 and 2009. Full waveform lidar data were collected using NASA's laser altimeter instrument, Laser Vegetation Imaging Sensor (LVIS) (Blair & Hofton, 1999), in March 2005. LVIS has a scan angle of approximately 12° and was flown at an altitude of 10 km, yielding a swath width of 2 km and 25 m footprints with 20 m