Remote Sensing of Environment 143 (2014) 26–38

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Influence of lidar, Landsat imagery, disturbance history, plot location accuracy, and plot size on accuracy of imputation maps of forest composition and structure

Harold S.J. Zald [a,*], Janet L. Ohmann [a], Heather M. Roberts [a], Matthew J. Gregory [a], Emilie B. Henderson [b], Robert J. McGaughey [c], Justin Braaten [a]

[a] Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA
[b] Institute for Natural Resources, Oregon State University, Portland, OR 97207, USA
[c] USDA Forest Service, PNW Research Station, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 17 July 2013
Received in revised form 18 December 2013
Accepted 19 December 2013
Available online 19 January 2014

Keywords:
Lidar
Landsat time series
Nearest-neighbor imputation
Disturbance
Forest composition and structure

ABSTRACT

This study investigated how lidar-derived vegetation indices, disturbance history from Landsat time series (LTS) imagery, plot location accuracy, and plot size influenced accuracy of statistical spatial models (nearest-neighbor imputation maps) of forest vegetation composition and structure. Nearest-neighbor (NN) imputation maps were developed for 539,000 ha in the central Oregon Cascades, USA. Mapped explanatory data included tasseled-cap indices and disturbance history metrics (year, magnitude, and duration of disturbance) from LTS imagery, lidar-derived vegetation metrics, climate, topography, and soil parent material. Vegetation data from USDA Forest Service forest inventory plots was summarized at two plot sizes (plot and subplot) and geographically located with two levels of accuracy (standard and improved). Maps of vegetation composition and structure were developed with the Gradient Nearest Neighbor (GNN) method of NN imputation using different combinations of explanatory variables, plot spatial resolution, and plot positional accuracy. Lidar vegetation indices greatly improved predictions of live tree structure, moderately improved predictions of snag density and down wood volume, but did not consistently improve species predictions. LTS disturbance metrics improved predictions of forest structure, but not to the degree of lidar indices, while also improving predictions of many species. Absence of disturbance attribution (i.e. disturbance type such as fire or timber harvest) in LTS disturbance metrics may have limited our ability to predict forest structure. Absence of corrected lidar intensity values may also have lowered accuracy of snag and species predictions. However, LTS disturbance attribution and lidar corrected intensity values may not be able to overcome fundamental limitations of remote sensing for predicting snags and down wood that are obscured by the forest canopy. Improved GPS plot locations had little influence on map accuracy, and we suggest under what conditions improved GPS plot locations may or may not improve the accuracy of predictive maps that link remote sensing with forest inventory plots. Subplot NN imputation maps had much lower accuracy compared to maps generated using response variables from larger whole plots. No single map had optimal results for every mapped variable, suggesting map users and developers need to prioritize what forest vegetation attributes are most important for any given map application.

1. Introduction

Forest management and conservation have grown increasingly complex, involving consideration of a wide array of ecological, economic, and societal values. Issues such as old growth conservation, wildlife habitat management, timber extraction, forest restoration, fuel reduction, and wildfire risk assessment often involve multiple interacting objectives, values, and threats (e.g. climate change, wildfires, and insect outbreaks) spanning broad spatial scales and long ecological gradients. In this complex policy and decision-making environment, quantitative information is required about forest vegetation conditions over large landscapes that is highly detailed with respect to multiple vegetation attributes, and spatially complete (i.e. mapped) (Spies et al., 2007).

Remotely sensed data are ideally suited to meet the need for spatially complete data about forests over large landscapes. Regional maps of forest cover are often based on multispectral satellite imagery (Cohen, Maiersperger, Spies, & Oetter, 2001; Hansen et al., 2003; Woodcock et al., 1994). However, maps from satellite imagery alone cannot provide

* Corresponding author at: 321 Richardson Hall, Oregon State University, Corvallis, OR 97331, USA. Tel.: +1 541 758 7759; fax: +1 541 737 1393.
E-mail addresses: harold.zald@oregonstate.edu (H.S.J. Zald), janet.ohmann@oregonstate.edu (J.L. Ohmann), heather.roberts@oregonstate.edu (H.M. Roberts), matt.gregory@oregonstate.edu (M.J. Gregory), emilie.henderson@oregonstate.edu (E.B. Henderson), bmcgaughey@fs.fed.us (R.J. McGaughey), justin.braaten@oregonstate.edu (J. Braaten).

0034-4257/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2013.12.013

the level of detail about forest composition and structure often required for many forest management and research applications. Measurements on field plot inventories often contain highly detailed ecological data, but only at sampled locations so they lack complete spatial coverage. As such, there is considerable interest in integrating field plots with remotely sensed data to generate maps with the spatial coverage of remotely sensed imagery and the ecological detail of field plots (Ohmann & Gregory, 2002; Tomppo, 1991; Tomppo, Goulding, & Katila, 1999).

One approach to integrating field plots and remotely sensed data is nearest-neighbor (NN) imputation, which has been widely used in forest inventory, monitoring, decision-support, and ecological research (Gjertsen, 2007; Moeur et al., 2011; Ohmann et al., 2012; Pierce, Ohmann, Wimberly, Gregory, & Fried, 2009; Reese et al., 2003; Spies et al., 2007; Tomppo et al., 2008; Wilson, Lister, & Riemann, 2012). Imputation is a method for filling in missing data by substituting values from donor observations (Eskelson et al., 2009). In forestry applications, imputation is used to estimate forest characteristics for large areas where a set of mapped explanatory variables are available for the entire spatial extent and these variables are related to a more detailed set of response variables only available for a limited sample of the study area. Response variables are usually measures of forest composition or structure derived from a sample of field plots, while mapped explanatory variables can include multispectral satellite imagery and other spatially complete datasets (i.e. climate, topography, etc.). In NN imputation, either a single donor observation (plot) can be chosen to fill in a given missing observation [$k = 1$], or multiple donor observations can be averaged to fill in a given missing observation [$k > 1$]. A major strength of NN imputation where $k = 1$ is the retention of the co-variance structure of multiple response variables, because each prediction links to a set of response values within a single plot.

As noted above, NN imputation mapping often relies on satellite imagery as mapped explanatory data (Ohmann & Gregory, 2002; Wilson et al., 2012). In particular, Landsat imagery (individual spectral bands and/or vegetation indices) is attractive for regional forest mapping due to its low cost, global coverage, long temporal record, and large scene-sizes, as well as spectral and spatial resolutions compatible with characterizing vegetation attributes (Cohen & Goward, 2004). However, Landsat and other passive optical sensors have limited sensitivity to vertical and below-canopy vegetation structure (Lu, 2006), and the information content in Landsat imagery is known to saturate in forests with high leaf area indices (Turner, Cohen, Kennedy, Fassnacht, & Briggs, 1999). These limitations of Landsat and other passive optical sensors pose problems for NN imputation mapping of forest attributes such as stand density, snags, and down wood (Eskelson, Temesgen, & Hagar, 2012; Pierce et al., 2009), which are important for carbon inventory and assessment, wildland fuels, and wildlife habitat.

Compared to Landsat and other passive optical sensors, Light Detection and Ranging (lidar) data can better represent the three-dimensional structure of forest canopies, and has been widely used to characterize vegetation cover and structure (see reviews by Dubayah & Drake, 2000; Lefsky, Cohen, Parker, & Harding, 2002; Reutebuch, Andersen, & McGaughey, 2005). Additionally, lidar does not suffer as much as Landsat imagery from declines in sensitivity and accuracy in forests with high leaf area index. The cost of lidar acquisition has declined dramatically over the past decade, such that lidar is increasingly available for large landscapes. Lidar has the potential to greatly improve NN imputation maps of forest structural attributes compared to maps developed using Landsat imagery or other passive optical sensors. Recent studies using lidar have had promising results at moderate spatial resolutions ($\leq$30 m pixels) and relatively small spatial extents (<60,000 ha); predicting presence/absence of snags and understory attributes (Martinuzzi et al., 2009), and imputation mapping of live tree structural attributes (Falkowski et al., 2010; Hudak, Crookston, Evans, Hall, & Falkowski, 2008). However, no published studies have determined if lidar can improve regional NN imputation mapping of forest attributes such as snag and down wood abundance, or species composition.

In addition to lidar data, advances utilizing the Landsat time series (LTS) may also improve accuracy of NN imputation maps. With the recent opening of the Landsat archive (Woodcock et al., 2008), there has been a proliferation of research in multi-temporal change detection and disturbance mapping (Huang et al., 2010; Kennedy, Yang, & Cohen, 2010; Masek et al., 2008). LTS disturbance metrics (such as time since, magnitude of, and duration of disturbance) may improve the accuracy of NN imputation indirectly, since many trends in forest composition and structure are closely related to disturbance history (Franklin et al., 2002; Oliver, 1980; Spies, 1991). This contrasts with lidar's direct characterization of forest structure. LTS disturbance metrics have been shown to have comparable predictive power to lidar for live basal area and aboveground biomass, and superior predictive power for dead basal area and aboveground biomass (Pflugmacher, Cohen, & Kennedy, 2012), suggesting many accuracy improvements that lidar can bring to imputation mapping might also be reached using LTS disturbance metrics. LTS disturbance metrics also have the advantage of complete spatial coverage and dramatically lower costs compared to lidar. LTS disturbance metrics have been used in NN imputation mapping of forests (Ohmann & Gregory, 2002), but no published studies have determined if LTS disturbance metrics can obtain predictions of comparable accuracy to lidar within the context of regional multivariate NN imputation maps of forest composition and structure. An additional advantage of LTS imagery for NN imputation mapping is it permits pixel-level normalization of multi-date images (Kennedy et al., 2010). This is an important consideration when minimization of year-to-year spectral variability and seamless multi-scene image mosaics are desired to relate to plot data collected over multiple years and across large spatial extents (Ohmann, Gregory, & Roberts, 2013).

For NN imputation and other methods linking field plots to remotely sensed data, accuracy of plot locations and plot size are important considerations. Studies relating lidar data to forest structure often do so using field plots geo-referenced using GPS receivers that manufacturers market as being capable of sub-meter accuracy when used under ideal conditions (Falkowski et al., 2010; Hudak et al., 2008; Kane et al., 2010; Pflugmacher et al., 2012). Although users and receiver manufacturers refer to GPS "accuracy", it is important to note that the "accuracy" statistic reported by GPS post-processing software is really "the precision of the solution" (i.e. a modeled estimate of geographic position), and computed GPS positions can still deviate from true geographic positions even when very high precision (aka accuracy) is reported. The accuracy of plot locations can be evaluated by comparing the GPS results with the "true" location obtained using high-order survey methods, but such comparisons are rarely made.

Unlike studies where research plots are geo-referenced using high precision GPS receivers, regional NN imputation mapping typically relies on existing plot networks such as the USDA Forest Service's Forest Inventory and Analysis Program (FIA) (Ohmann et al., 2012; Wilson et al., 2012). A variety of methods (i.e. recreational grade GPS receivers, map interpretation, and photo interpretation) have been used to determine the geographic locations of FIA plots. FIA plots located with recreational grade GPS receivers (the most common method used for locating FIA plots) have positional accuracy averaging 5–20 m, but some plots can have positional errors exceeding 20 m (Cooke, 2000; Hoppus & Lister, 2005). Additionally, FIA plots are comprised of multiple fixed-radius subplots, and NN imputation can be conducted using either larger plots (i.e. aggregates of subplots) or smaller individual subplots (McRoberts, 2009). Simulations suggest accuracy of plot locations and plot size can strongly impact the accuracy of lidar-derived estimates of forest biomass (Frazer, Magnussen, Wulder, & Niemann, 2011) and Landsat-derived estimates of forest area (McRoberts, 2010), but the impacts of plot location accuracy and plot size (in this study referring to whole plots versus individual subplots) on prediction accuracy have not been examined within the