Remote Sensing of Environment 209 (2018) 90–106

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots

Giona Matasci[a,*], Txomin Hermosilla[a], Michael A. Wulder[b], Joanne C. White[b], Nicholas C. Coops[a], Geordie W. Hobart[b], Harold S.J. Zald[c]

[a] Integrated Remote Sensing Studio, Department of Forest Resources Management, University of British Columbia, 2424 Main Mall, Vancouver, BC, V6T 1Z4, Canada
[b] Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside Road, Victoria, BC, V8Z 1M5, Canada
[c] Department of Forestry and Wildland Resources, Humboldt State University, 1 Harpst St., Arcata, CA 95521, USA

ARTICLE INFO

Keywords:
Lidar
Landsat
Forest structure
Monitoring
Imputation
Random Forest

ABSTRACT

Passive optical remotely sensed images such as those from the Landsat satellites enable the development of spatially comprehensive, well-calibrated reflectance measures that support large-area mapping. In recent years, as an alternative to field plot data, the use of Light Detection and Ranging (lidar) acquisitions for calibration and validation purposes in combination with such satellite reflectance data to model a range of forest structural response variables has become well established. In this research, we use a predictive modeling approach to map forest structural attributes over the ~552 million ha boreal forest of Canada. For model calibration and independent validation we utilize airborne lidar-derived measurements of forest vertical structure (known as lidar plots) obtained in 2010 via a > 25,000 km transect-based national survey. Models were developed linking the lidar plot structural variables to wall-to-wall 30-m spatial resolution surface reflectance composites derived from Landsat Thematic Mapper and Enhanced Thematic Mapper Plus imagery. Spectral indices extracted from the composites, disturbance information (years since disturbance and type), as well as geographic position and topographic variables (i.e., elevation, slope, radiation, etc.) were considered as predictor variables. A nearest neighbor imputation approach based on the Random Forest framework was used to predict a total of 10 forest structural attributes. The model was developed and validated on > 80,000 lidar plots, with R^2 values ranging from 0.49 to 0.61 for key response variables such as canopy cover, stand height, basal area, stem volume, and aboveground biomass. Additionally, a predictor variable importance analysis confirmed that spectral indices, elevation, and geographic coordinates were key sources of information, ultimately offering an improved understanding of the driving variables for large-area forest structure modeling. This study demonstrates the integration of airborne lidar and Landsat-derived reflectance products to generate detailed and spatially extensive maps of forest structure. The methods are portable to map other attributes of interest (based upon calibration data) through access to Landsat or other appropriate optical remotely-sensed data sources, thereby offering unique opportunities for science, monitoring, and reporting programs.

1. Introduction

In Canada, forest ecosystems are a mosaic of trees, wetlands, and lakes, occupying an area of ~650 million ha (Wulder et al., 2008b), with a treed area of 347 million ha (Natural Resources Canada, 2016). The boreal forest, an important source of both renewable and non-renewable resources, occupies an area of 552 million ha (with 270 million ha of trees) and forms an east-west band across the country, representing a range of climatic, physiographic, and vegetation conditions (Brandt, 2009). To effectively implement sustainable management and development practices aiming at accommodating both conservation (e.g., preservation of wildlife habitats) and human use needs (e.g., building materials, fuels), boreal forests require comprehensive, timely, and accurate inventory and monitoring efforts. To this end, data collection campaigns are necessary to characterize and map forest structure, determining attributes such as canopy cover, height, biomass, stem volume as well as age, species, land-cover, and disturbance history (White et al., 2014).

The availability of accurate national forest structural information, often collected following sample-based inventories (Tomppo et al., 2010), is the foundation for satisfying a variety of science and policy information needs as well as for meeting national and international

---

[*] Corresponding author.
E-mail address: giona.matasci@gmail.com (G. Matasci).

https://doi.org/10.1016/j.rse.2017.12.020
Received 15 June 2017; Received in revised form 5 December 2017; Accepted 15 December 2017

reporting obligations (Canadian Council of Forest Ministers, 1995). However, there are important limitations of field plot-based measurements such as their cost, lack of spatial coverage, and long updating cycles. To cope with these field data collection issues, practitioners often relied upon photo plots, an expert-based interpretation of aerial imagery. For example, the Canadian National Forest Inventory (NFI) is based upon a 1% national sample as represented by 2 × 2 km photo plots established largely on a 20 × 20 km grid, supported by a subset of ground plots, collected on a panel-basis over a 10 year update cycle (Gillis et al., 2005).

More recently, Light Detection And Ranging (lidar) remote sensing technology (Baltsavias, 1999) has gained popularity as a means to obtain detailed 3-dimensional measurements of the structure of the canopy to represent forest conditions at a given place and time (Wulder et al., 2008a). As reviewed in Nelson (2013), this potential of using airborne laser-based acquisitions to study forested ecosystems was identified in the 1970s. More specifically, transects of airborne lidar data have been found to mitigate the costs of ground plot installation and offer spatially extensive and representative sampling of calibration and validation data to support the modeling of forest attributes (Wulder et al., 2012b). Wulder et al. (2012a) outline the concept of lidar plots, whereby samples of lidar are gathered (on a transect basis) to provide regional representation and spatially referenced data suitable for the development of such models.

At the same time, multispectral imagery from satellites platforms has been demonstrated as a source of data to provide spatially comprehensive characterizations of forest attributes over large areas with a level of spatial detail of relevance to the needs of forest inventory and sustainable forest management (Brosofske et al., 2014; Cohen et al., 2001; Woodcock et al., 1994). In particular, sensors of the Landsat mission such as Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) acquire reflectance products with suitable spectral and spatial resolutions that can be used as support to map vegetation conditions and dynamics (Cohen and Goward, 2004). A known limitation of medium resolution optical satellite imagery is radiometric saturation of the recorded signal when estimating vertically distributed attributes such as biomass or canopy height (Duncanson et al., 2010; Lu, 2006, 2005). When utilizing this type of data to characterize large areas, information on vertically distributed attributes can be obtained by leveraging time-series of images providing insights on forest development and succession through a reconstructed disturbance history (Pflugmacher et al., 2012). In forest ecosystems the temporal series of spectral information and related trends offers unique life-stage and succession insights to aid in the modeling of structural attributes such as stand height or biomass (Deo et al., 2017; Lu, 2006; Pflugmacher et al., 2012; Powell et al., 2010; Zald et al., 2014). The opening of the Landsat archive in 2008 (Woodcock et al., 2008) facilitated the implementation of studies utilizing the complete spatial and temporal depth of the Landsat archive (Hansen and Loveland, 2012; Wulder et al., 2008b). Additionally, there has been extensive development in routines to create composites free of atmospheric effects (Potapov et al., 2011; Roy et al., 2010; White et al., 2014). These composites can be used to detect and label change (Hermosilla et al., 2015a) as well as to uncover and quantify trends (Ju and Masek, 2016).

In recent years, there has been wide interest in developing methods relying on optical imagery to extrapolate forest structural data beyond lidar or field data coverage to represent an entire area of interest. Such approaches generally rely on statistical predictive modeling to relate localized measurements of forest conditions (e.g., lidar) and image-derived information covering broader areas (Wulder et al., 2012b). The forest/vegetation attributes of interest (canopy cover, tree height, diameter at breast height, basal area, biomass, stem volume, etc.) represent the response variables to be modeled, whereas features extracted from multispectral satellite images or other geospatial datasets such as Digital Elevation Models (DEM) or climatic layers constitute the predictor variables, or predictors. To implement these image-based spatial predictions, common methods include linear regression or Random Forest (RF) (Breiman, 2001). RF offers robust, accurate and scalable solutions to both regression and classification problems, allowing at the same time the user to gain insights on the model by means of implicitly produced variable importance measures. Application examples of RF can be found in both the remote sensing (Belgiu and Drăgu, 2016; Gislason et al., 2006) and forestry communities (Gleason and Im, 2012; Latifi et al., 2010). In forestry, another increasingly common approach is nearest neighbor (NN) imputation (Eskelson et al., 2009; Ohmann and Gregory, 2002). In contrast to regression approaches that can distort marginal distributions and covariation between Y-variables, imputation fills in missing data by substituting values from donor observations, with the underlying assumption that two locations with similar values of X-variables should be similar with respect to Y-variables. A major strength of imputation approaches is these donor-based methods are multivariate, non-parametric, and distribution-free (Eskelson et al., 2009).

Table 1 summarizes the key characteristics (type of input data, methods employed, forest attributes modeled, study area) of recent studies which combine lidar or field data and optical imagery to map forest structural attributes, recognizing that a number of studies also exist which produced carbon estimates in a laser profiling context (e.g., Nelson et al., 2017). The majority of these previous studies has tested methodologies over small areas (e.g., Ahmed et al., 2015). At the regional scale, Landsat imagery has been used in a number of studies to interpolate or extrapolate airborne lidar-based estimates of forestry productivity. Principally in forested areas these approaches have used a variety of statistical and model-based approaches to predict a range of attributes, most often height and aboveground biomass in either the US, Canada or Europe. Statistical approaches range from conventional regressions to more advanced ensemble methods like RF or regression trees such as in Hansen et al. (2016) who extrapolated Geoscience Laser Altimeter System (GLAS) tree height data with Landsat time-series in Sub-Saharan Africa. Profiling lidar data collected by the Portable Airborne Laser System (PALS) has also been used to provide high precision height measurements to be combined with GLAS pulses and Landsat-derived land-cover strata to produce local biomass and carbon estimates (Margolis et al., 2015; Neigh et al., 2013). Zald et al. (2016) applied an imputation model to map forest attributes over 50 Landsat WRS-2 scenes (forested ecozones of Saskatchewan) using a set of Landsat spectral, change and topographical predictor variables with reported accuracies in the 0.42–0.69 R^2 range when validating against independent lidar plots. Common to most of these approaches is the recognition that these technologies can inform forest management and reporting activities as well as to offer spatially explicit inputs to carbon accounting models (White et al., 2014). The level of spatial detail ultimately dictates the application and utility of a given structural map product. Studies that have been undertaken over large areas, have necessitated the use of more coarse spatial resolution imagery reducing the applicability below the regional scale. For example, Lefsky (2010) and Simard et al. (2011) both produced global tree height maps by intersecting GLAS height estimates with forest layers obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) images. Also relying on MODIS imagery, Beaudoin et al. (2014) produced Canada-wide estimates of a large number of forest attributes using NFI photo plot data for calibration and validation.

In this paper, building on the regional mapping effort by Zald et al. (2016), we present a methodological framework to combine wall-to-wall Landsat surface reflectance composites, forest change information, and descriptors of topography/location to map forest attributes (including canopy cover, height, aboveground biomass and stem volume) longitudinally across a continent. In so doing, we generate information products relating to forest structure at the unprecedented spatial resolution of 30 m for the entire 552 million ha Canadian boreal forest, representing 2010 conditions. We address, document, and communicate challenges related to data processing architecture, modeling