ARTICLE IN PRESS

RSE-09738; No of Pages 12

Remote Sensing of Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data

Matthew C. Hansen [a,*], Peter V. Potapov [a], Scott J. Goetz [b], Svetlana Turubanova [a], Alexandra Tyukavina [a], Alexander Krylov [a], Anil Kommareddy [a], Alexey Egorov [c]

[a] *University of Maryland, United States*
[b] *Woods Hole Research Center, United States*
[c] *South Dakota State University, United States*

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form 26 January 2016
Accepted 11 February 2016
Available online xxxx

ABSTRACT

Landsat time-series multi-spectral data, GLAS (Geoscience Laser Altimeter System) height data and a regression tree model were used to estimate tree height for a transect in Sub-Saharan Africa ranging from the Sahara Desert through the Congo Basin to the Kalahari Desert (+22 to −22° latitude and 23 to 24° longitude). Objectives included comparing the performance of Landsat 7- and 8-derived inputs separately and combined in mapping tree height at a regional scale, assessing the relative value of good observation counts and different Landsat spectral inputs for tree height estimation across a range of environments, and describing tree height distributions and discontinuities in Sub-Saharan Africa. A total of 5371 images were processed and per pixel quality assessed to create a set of multi-temporal metrics for the 2013 and 2014 calendar years for Landsat 7 only, Landsat 8 only and both Landsat 7 and 8 combined. Differences in performance were slight between different sensor inputs. However, performance generally improved with increasing numbers of good observations. Metrics derived from red reflectance data contributed most in estimating tree height. The regression tree algorithm accurately reproduced the LiDAR-derived height training data with an overall mean absolute error (MAE) for tree height estimation of 2.45 m using integrated Landsat 7 and 8 data. Significant underestimations were quantified for tall tree cover (MAE of 4.65 m for >20 m heights) and overestimations for low/no tree cover (MAE 1.61 for <5 m heights). Resulting tree distributions were found to be discontinuous with a primary dry seasonal woodlands cluster of 5–10 m in height, a second cluster of primarily dry evergreen forest tree cover from 11–17 m, and a third cluster of humid evergreen forest tree cover ≥18 m. The integration of Landsat 7 and 8 and forthcoming Sentinel 2 time-series optical data to extend the value of LiDAR forest structure measurements is recommended.

1. Introduction

The Operational Land Imager (OLI) onboard the Landsat 8 spacecraft represents a significant advance in earth observation monitoring capability (Loveland and Irons, this issue). Additional spectral bands, enhanced bandwidths, increased signal to noise, greater radiometric detail, and other technical advances make Landsat 8 the most robust in the Landsat series. While Landsat 8 data are demonstrably technically superior to preceding Landsat sensors, we believe that the integration of data from different Landsat missions in order to provide denser time-series information to be critical to advancing monitoring capabilities. The high cadence, high spatial detail information domain outlined by Cihlar (2000) will most likely be met through a virtual constellation approach. Currently, Landsat 7 and 8 are available for integrated use. The Landsat-like spectral bandwidths of the Sentinel 2 satellites (Drusch et al., 2012) promise a further enrichment of 10–30 m time-series multi-spectral data.

With the opening of the Landsat archive, numerous research efforts have begun to exploit time-series of Landsat imagery in characterizing land change (Kennedy, Cohen, & Schroeder, 2007; Huang et al., 2009; Potapov et al., 2012; Zhu & Woodcock, 2014). However, the uneven acquisition of Landsat imagery through time and across space and missions poses a significant challenge to developing turnkey land cover characterization models such as those of the MODIS Land Science Team (Justice et al., 2002). Landsat data richness varies geographically as a function of the application and variation of acquisition strategies, cloud cover, and length of orbit over land. Fig. 1 shows annual growing season coverage for the Landsat 7 record through 2012 and illustrates the historic paucity of data for many parts of Africa, especially the dry tropics. The most data rich period of Landsat starts in 2013 with the combined Landsat 7 and 8 image archive. While variations in data richness due to cloud cover will always impact passive optical sensors such as Landsat, the integration of 8-day repeat Landsat 7 and 8 and

---

[*] Corresponding author.

http://dx.doi.org/10.1016/j.rse.2016.02.023
0034-4257/

Please cite this article as: Hansen, M.C., et al., Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data, *Remote Sensing of Environment* (2016), http://dx.doi.org/10.1016/j.rse.2016.02.023

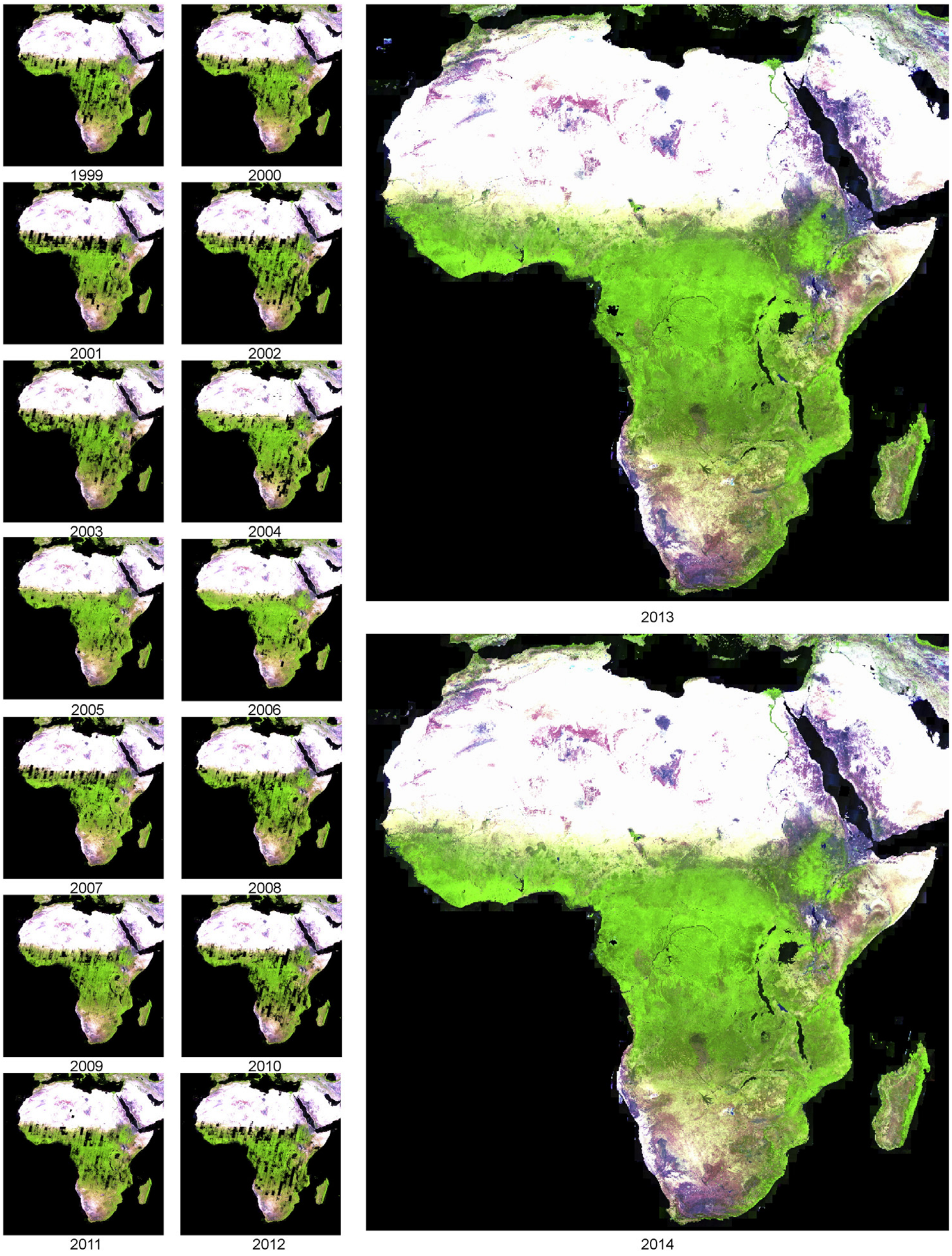

Fig. 1. Annual growing season median normalized reflectance from Landsat 7 Enhanced Thematic Mapper data (ETM+) from 2000 to 2014. Black areas in the interior of Africa represent data gaps, i.e. no viable land observations for that year.

forthcoming Sentinel 2 data will likely result in a MODIS-like turnkey large area mapping and monitoring capability.

One approach to exploiting time-series observations for mapping land cover is to convert all viable data for a given study period into multi-temporal metrics. Metrics are statistical derivatives of time-series observations that are not tied to a specific time of year and have been used with AVHRR, MODIS and Landsat data (DeFries, Hansen, & Townshend, 1995; Hansen, DeFries, Townshend, & Sohlberg, 2000; Hansen et al., 2003; Potapov et al., 2012, 2015). Examples include peak greenness or range of near-infrared reflectance. Metrics have the advantage of tolerating unequal data richness; for example, quantiles can be calculated for a population of 10 or 100 input observations. Since metrics are not tied to a specific date of year, they are appropriate for regional to global scale mapping. Africa, in particular, is a suitable region for the application of metrics due to the mirrored phenology north and south of the equator; metrics reorder input time-series observations, normalizing phenological variation without respect to the timing of phenology and have been shown to perform better than time-sequential data in mapping African tree cover (Hansen, Townshend, DeFries, & Carroll, 2005). In this study, we assess the value of metrics derived from Landsat 7 and 8 data streams to map tree height in Sub-Saharan Africa.

For large area mapping, we employ decision tree models, either classification trees for categorical land cover themes, such as forest and non-forest, or regression trees for continuously varying themes, such as percent tree cover. Tree models are robust, distribution-free algorithms that are easily implemented and transparent in their operation. Tree models have been used for the suite of land cover products from the MODIS Land Science Team, (Friedl et al., 2002; Hansen et al., 2002) and more recently in our global forest study using Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data (Hansen et al., 2013). Hanan et al. recently posited that classification and regression tree (CART) models can produce misleading results due to the possibility that they “impose discontinuities” in data where there “may be none

Please cite this article as: Hansen, M.C., et al., Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data, *Remote Sensing of Environment* (2016), http://dx.doi.org/10.1016/j.rse.2016.02.023