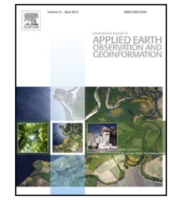
International Journal of Applied Earth Observation and Geoinformation 53 (2016) 118–127

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



The accuracy of large-area forest canopy cover estimation using Landsat in boreal region

Hadi [a,*], Lauri Korhonen [b], Aarne Hovi [a], Petri Rönnholm [a], Miina Rautiainen [a,c]

[a] Aalto University, School of Engineering, Department of Built Environment, PO Box 15800, 00076 Aalto, Finland
[b] University of Eastern Finland, School of Forest Sciences, PO Box 111, FI-80101 Joensuu, Finland
[c] Aalto University, School of Electrical Engineering, Department of Radio Science and Engineering, PO Box 13000, 00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 23 May 2016
Received in revised form 9 August 2016
Accepted 31 August 2016

Keywords:
Canopy cover
Landsat
Tree cover
Boreal
Forest

ABSTRACT

Large area prediction of continuous field of tree cover i.e., canopy cover (CC) using Earth observation data is of high interest in practical forestry, ecology, and climate change mitigation activities. We report the accuracy of using Landsat images for CC prediction in boreal forests validated with field reference plots (N = 250) covering large variation in latitude, forest structure, species composition, and site type. We tested two statistical models suitable for estimating CC: the beta regression (BetaReg) and random forest (RanFor). Landsat-based predictors utilized include individual bands, spectral vegetation indices (SVI), and Tasseled cap (Tass) features. Additionally, we tested an alternative model based on spectral mixture analysis (SMA). Finally, we carried out a first validation in boreal forests of the recently published Landsat Tree Cover Continuous (TCC) global product.

Results showed simple BetaReg with red band reflectance provided the highest prediction accuracy (leave-site-out $RMSE_{CV}$ 13.7%; R^2_{CV} 0.59; $bias_{CV}$ 0.5%). Spectral transformations into SVI and Tass did not improve accuracy. Including additional predictors did not significantly improve accuracy either. Non-linear model RanFor did not outperform BetaReg. The alternative SMA model did not outperform the empirical models. However, empirical models cannot resolve the underestimation of high cover and overestimation of low cover. SMA prediction errors appeared less dependent on forest structure, while there seemed to be a potential for improvement by accounting for endmember variability of different tree species. Finally, using temporally concurrent observations, we showed the reasonably good accuracy of Landsat TCC product in boreal forests (RMSE 13.0%; R^2 0.53; bias −2.1%), however with a tendency to underestimate high cover.

1. Introduction

Regular large-scale monitoring of forest cover with remote sensing is presently needed more than ever before; the world's political leaders collectively acknowledge the urgent importance of stopping deforestation and forest degradation in the international climate agreement reached in Paris in 2015. The boreal forests which account for 21% of world's forest area, while not experiencing serious deforestation threat, are highly vulnerable to global warming which increases disturbance e.g. from fire or insect outbreaks (Bonan, 2008). These disturbances potentially contribute CO_2 emissions and further enhance climate warming.

Forest cover mapping with remote sensing (RS) has been traditionally done by classification of forest and non–forest areas (or other pre-defined forest classes), which are then upscaled to a larger mapping unit to calculate the proportion of pixels or segments labelled as forest. However, the resulting maps are dependent on the forest definition, i.e. mainly the threshold of tree cover above which an area is classified as forest. The definitional discrepancies have been the main source of inconsistent satellite-based estimates between previous studies (Sexton et al., 2015). This has led to the recommendation of shifting from classification approach toward mapping canopy cover as a continuous field (Hansen et al., 2002; Sexton et al., 2015). In addition to allowing consistent forest area estimates, canopy cover maps would allow for monitoring changes in forest cover at scales smaller than the pixel size. Canopy cover is highly correlated with basal area and thus volume and biomass. Beyond practical forestry context, canopy cover is also an important ecological indicator related to

* Corresponding author.
E-mail address: hadi.hadi@aalto.fi (Hadi).

http://dx.doi.org/10.1016/j.jag.2016.08.009
0303-2434/

Table 1
Summary of study plots.

Site	No. of plots	Plot size (m)	Canopy cover (%)	Stand density (trees ha^{-1})	Mean tree height (m)	Mean DBH[a] (cm)	Basal area (m^2 ha^{-1})
Suonenjoki	98	25 × 24	2.5–96.8	257–15,920	1.8–29.2	1.4–36.6	0.3–44.8
Koli	30	30 × 30	43.7–97.5	433–2883	9.8–35.5	10.4–66.5	15.1–43.4
Tammela	7	r[b] = 12.5	24.0–74.0	552–7630	2.7–27.6	3.6–35.8	0.8–27.0
Joensuu	7	r = 12.5	17.8–87.3	383–4605	3.4–21.9	3.8–28.9	2.0–38.0
Rovaniemi	6	r = 12.5	13.5–92.4	682–12,470	1.3–14.4	0.5–19.1	1.0–26.0
Sodankylä	66	r = 12.5	6.0–78.2	95–17,080	0.4–24.4	0.0–34.6	0.0–34.0
Hyytiälä	24	Variable[c]	33.1–96.5	623–8700	2.2–33.1	1.1–47.0	0.8–54.1
Sotkamo	12	r = 12.5	45.2–94.6	965–8815	3.4–25.1	2.6–34.6	1.5–32.0
All	**250**	**Range**	**2.5–97.5**	**95–17,080**	**0.4–35.5**	**0.0–66.5**	**0.0–54.1**
		Mean ± SD	**60.3 ± 21.3**	**2575 ± 2446**	**15.0 ± 7.3**	**18.5 ± 10.1**	**19.5 ± 10.7**

[a] Diameter at-breast-height (1.3m).
[b] Circular plot with radius r.
[c] Mostly 30 × 30m, but also varied from 20 × 30m to 35 × 40m.

spatial heterogeneity of vegetated surface, habitat, rain and light interception, and forest microclimate (Jennings et al., 1999).

Canopy cover is defined as the proportion of ground covered by the vertical projection of the tree crowns (FAO, 2004; IPCC, 2003; Jennings et al., 1999), including small gaps inside the crown perimeter (Gschwantner et al., 2009). Unbiased measurement of canopy cover using vertical sighting tubes (Korhonen et al., 2006) is a time-consuming process. Thus, canopy cover (CC) is seldom included in traditional forest inventory, or is assessed visually either in the field or from aerial photos. Two implications therefore arise: firstly, remote sensing techniques as alternative to laborious field measurements of CC are much needed; and secondly, the accuracy of the remote sensing techniques should be assessed most reliably with unbiased vertical measurements of CC. Regarding the latter, use of aerial photos to generate reference CC data may be inaccurate (McIntosh et al., 2012), while visual estimates of reference CC from high spatial resolution images can vary substantially between human interpreters (Montesano et al., 2009).

Currently the most accurate remote sensing technique to estimate CC is the airborne LiDAR given its measurement principle and geometry (e.g., Griffin et al., 2008; Korhonen et al., 2011). However, LiDAR is not suited for continuous, large-scale CC monitoring. The planned NASA satellite LiDAR mission GEDI scheduled for launch in 2018 samples between 50N and 50S latitudes thus leaving the boreal region uncovered. Therefore, passive optical satellites remain an important data source for wall-to-wall CC monitoring. Among the operational optical Earth observation satellites, Landsat is one of the most promising data sources for global CC monitoring as they are freely available, have high spatial resolution and revisit frequency, and provide long term data record (continuity) spanning more than four decades.

CC has been globally generated in the past first at 1 km resolution with images from the AVHRR sensor on board NOAA satellite (DeFries et al., 1999), then at 500 m (Hansen et al., 2003) and 250 m (DiMiceli et al., 2011) resolution with MODIS sensor on board Aqua and Terra. More recently, Sexton et al. (2013) further rescaled the MODIS Vegetation Continuous Field (VCF) product to Landsat resolution (30 m) and showed good agreement between the two. The improvement in spatial resolution is a significant progress as previously validation efforts have been lacking due to uneasy sampling for reference ground measurements at MODIS coarse resolution. Importantly, to our knowledge the new Landsat Tree Cover Continuous field (TCC) product has not yet been validated in the boreal region.

The aim of this present study was to assess the capability of Landsat image data to estimate CC in boreal forests across wide geographical area with diverse forest structure, stand development, species composition, and site type. We utilized a large ground reference dataset (N = 250) of unbiased vertical measurements of ground reference CC collected in south, central, and north Finland.

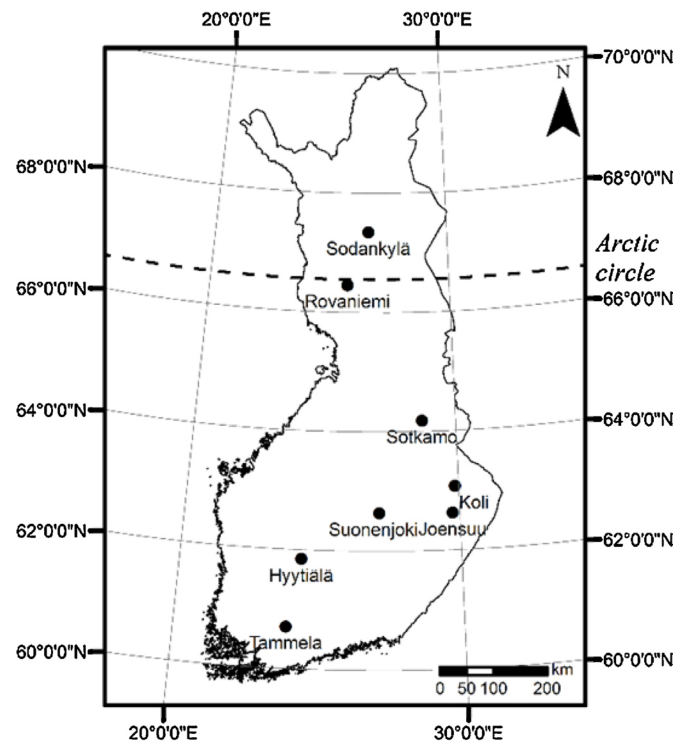


Fig. 1. The location of study sites.

The extensive dataset allowed us to evaluate a realistic scenario of predicting CC in a new area by leave-site-out cross validation procedure. We tested empirical regression methods suitable for proportion data (such as CC) and an alternative spectral mixture analysis for CC prediction. We also carried out a first validation of the new Landsat TCC global product in boreal forest using temporally consistent (same year, season) observations. Finally, we analysed the CC prediction errors to identify ways to further improve the prediction accuracy.

2. Materials

2.1. Field measurements

Field measurements of CC were carried out at eight forest sites in Finland (Table 1, Fig. 1) in 2005–2009. The 250 plots were located purposely to collect as diverse a dataset as possible from each study site and as a whole. The dataset comprised 135 Scots pine (*Pinus sylvestris* L.), 95 Norway spruce (*Picea abies* L. *Karst*) and 20 birch (*Betula* spp. L.) dominated plots. The site type varied with 119 xeric,