Towards improved boreal biome-scale forest canopy cover product

Hadi
Dept. of Built Environment
Aalto University
Vaisalantie 8
02130 Espoo
Finland
hadi.hadi@aalto.fi

Lauri Korhonen
School of Forest Sciences
University of Eastern Finland
Yliopistokatu 7
80101 Joensuu
Finland
lauri.korhonen@uef.fi

Aarne Hovi
Dept. of Built Environment
Aalto University
Vaisalantie 8
02130 Espoo
Finland
aarne.hovi@aalto.fi

Petri Rönnholm
Dept. of Built Environment
Aalto University
Vaisalantie 8
02130 Espoo
Finland
petri.ronnholm@aalto.fi

Miina Rautiainen
Dept. of Built Environment
Aalto University
Vaisalantie 8
02130 Espoo
Finland
miina.a.rautiainen@aalto.fi

Abstract

We report the Landsat-based prediction accuracy for percent canopy cover (CC) of 250 forest plots in eight sites across Finland where reference CC data were obtained from field measurements. We tested empirical models based on individual spectral bands and spectral vegetation indices (SVI), and semi-empirical linear spectral mixture analysis (LSMA). Our leave-site-out cross validation results showed that fairly good prediction accuracy ($RMSE_{CV}$ 15.4% CC unit; $bias_{CV}$ 1.2%, not significant) can be achieved using the best empirical model i.e., zero-and-one inflated beta regression. Despite slightly higher overall error ($RMSE_{CV}$ 17.7%), LSMA seemed promising to overcome the saturation of empirical models at high CC (>80%) and overestimation at low CC by accounting for the influence of forest understory.

Keywords: canopy cover, global tree cover product, boreal forest, Landsat, beta regression, spectral mixture analysis

1 Introduction

The significance of global monitoring of forest cover is well recognized as FAO carries out the Global Forest Resources Assessment (FRA) every 5 years. The FRA however has two practical limitations. Firstly, it is based on national-level country reports which do not adequately reveal the detailed spatial distribution of selective clearcuts. Secondly, there may be inconsistency in methodology adopted by different countries, especially in defining forest area. Regarding the first limitation, satellite remote sensing offers the most feasible approach for the urgent task of global mapping and monitoring of forest at fine spatial resolution. The second limitation has led to the recommendation of shifting from mapping forest/non-forest discrete land use class to mapping continuous, unambiguous, and scalable canopy cover (CC) [1].

NASA and the Global Land Cover Facility has recently published the new fine resolution (30m) tree cover continuous fields (TCC) product based on Landsat archives [2]. The objectives of this study were to: (1) carry out the first validation of the Landsat TCC product in the boreal zone; and (2) test alternative retrieval methods to estimate boreal forest CC from Landsat multispectral satellite images.

2 Materials and Methods

2.1 Materials

Field measurements of canopy cover using Cajanus tube were carried out between 2005 and 2009 in eight forest sites in Finland (Fig 1) with a total of 250 plots (135 pine, 95 spruce, 20 birch). Each plot was approximately the size of one Landsat pixel. We used the freely available geometrically and atmospherically corrected Landsat 5 TM Surface Reflectance product [3]. The spectrum of Landsat pixel coincident with the field plot center was extracted.

Figure 1: Location of the study sites.

2.2 Methods

We tested several regression models namely linear, exponential, and Zero-and-One-Inflated Beta Regression (ZOINBR), as well as nonparametric Random Forest. As input to the models we used reflectance factors in individual bands and two spectral vegetation indices: NDVI and reduced simple ratio (RSR):

$$RSR = \frac{\rho_{NIR}}{\rho_{red}}\left(\frac{\rho_{SWIR,max} - \rho_{SWIR}}{\rho_{SWIR,max} - \rho_{SWIR.min}}\right) \quad (1)$$

The minimum and maximum reflectance values in shortwave infrared ($\rho_{SWIR,min}$ and $\rho_{SWIR,max}$) were determined from all spectra of the pooled data.

Our implementation of LSMA assumes the surface reflectance measured by the satellite over the forest area as a linear mixture of signal from the tree canopy (sunlit and shaded) and understory (sunlit and shaded):

$$\rho_{TM,forest,i} = fr_{c,i} * \rho_{c,i} + fr_{u,i} * \rho_{u,i} + \epsilon_i \quad (2)$$

where

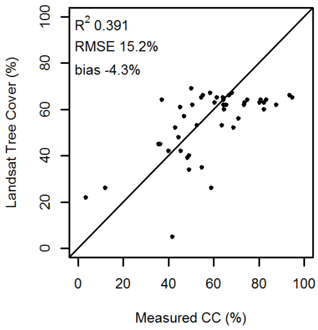
$\rho_{TM,forest,i}$: Landsat-derived surface reflectance in band i,
$fr_{c,i}$: fractional cover of tree canopy i.e., the canopy cover,
$fr_{u,i}$: fractional cover of understory,
$\rho_{c,i}$: reflectance of tree canopy endmember in band i,
$\rho_{u,i}$: reflectance of understory endmember in band i, and
ϵ_i: residual of spectra in band i;
with constraints $fr_{c,i} + fr_{u,i} = 1$ and $fr \geq 0$.

Leave-site-out cross validation scheme was chosen for accuracy assessment. We estimated $\rho_{c,i}$ and $\rho_{u,i}$ using CC of training data by substituting $fr_{c,i}$ = CC and $fr_{u,i}$ = (1-CC).

3 Results and Discussion

Validation of the latest TCC product for 2005 epoch revealed unsatisfactory accuracy with clear underestimation of CC in forest plots with high CC (>60-70%) (Fig 2).

Figure 2: Comparison between Landsat tree cover 2005 epoch and measured CC (n=48 plots from Suonenjoki field campaign 2005 and 2006).



Among the tested models for CC prediction, ZOINBR with inputs of all six bands provided the highest accuracy ($RMSE_{CV}$ 15.4%) (Fig 3d). Exponential RSR model (Fig 3b) performed better than NDVI (Fig 3a), but saturated at CC of approx. 80%, while Random Forest suffered the most from saturation i.e., also at low CC values (Fig 3c).

The highest accuracy achieved was reasonably good considering the large geographical area of the prediction. Our results are comparable to previous studies estimating CC from Landsat data [2, 4, 5]. Although differences in variance of residuals (normally distributed for all models, Shapiro-Wilk *p*-value>0.05) between models were not statistically significant (Bartlett test *p*-value=0.13), resolving saturation especially at low CC is imperative as internationally the tree cover threshold of 10% is used by FAO to classify forest/non-forest land cover. LSMA provided slightly lower accuracy (by 2.3%) but seemed more able to overcome the prediction saturation of fully empirical models (Fig 3e).

Wall-to-wall comparison between ZOINBR prediction and Landsat TCC (Fig 4) further confirmed the saturation of TCC in boreal forests.

Figure 3a-e: Leave-site-out prediction of tested models: (a) NDVI, linear regression; (b) RSR, exponential regression; (c) Random Forest; (d) ZOINBR (best); and (e) LSMA.

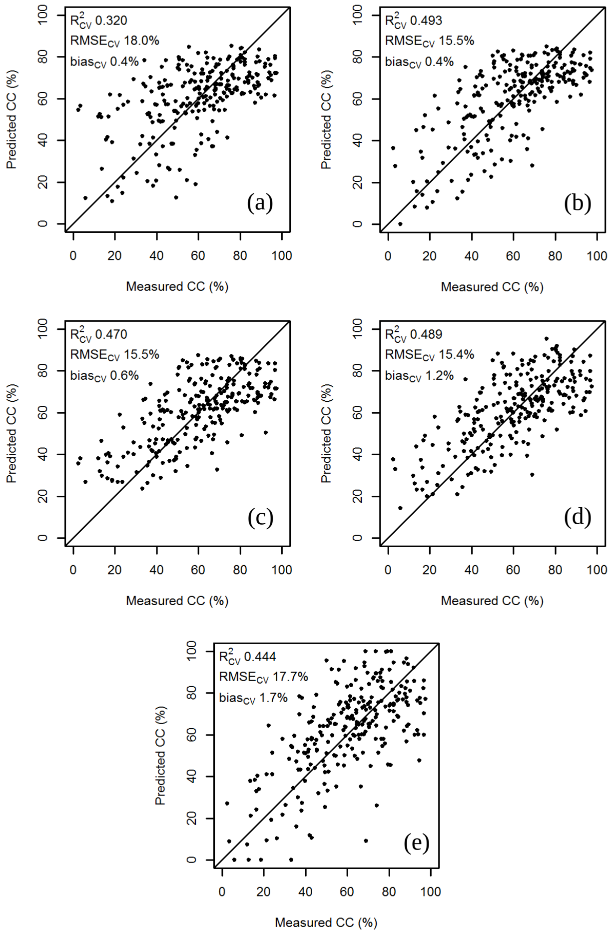
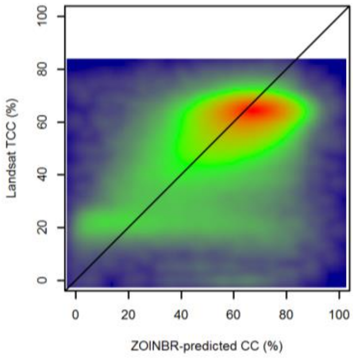


Figure 4: Pixel-by-pixel (n=145800) comparison between Landsat TCC and ZOINBR prediction in Suonenjoki. Colours indicate scatter density (high in red, low in blue).



4 Conclusion

We have shown that the regression tree-based global Landsat TCC underestimates CC in boreal forest, thus necessitating alternative methods to improve CC estimation from Landsat images. Improvements in methods which account for understory spectral contribution such as LSMA and physically-based models to resolve the prediction saturation of empirical models are much needed.

References

[1] J. O. Sexton *et al*. Conservation policy and the measurement of forests. *Nat. Clim. Chang.*, 6:192-196, 2016.

[2] J. O. Sexton *et al*. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *Int. J. Digit. Earth*, 6(5):427–448, 2013.

[3] J. G. Masek *et al*. A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geosci. Remote Sens. Lett.*, 3(1):68–72, 2006.

[4] A. M. S. Smith *et al*. A cross-comparison of field, spectral, and lidar estimates of forest canopy cover. *Can. J. Remote Sens.*, 35(5):447–459, 2010.

[5] M. Karlson *et al*. Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sens.*, 7(8):10017–10041, 2015.