Geocarto International

ISSN: 1010-6049 (Print) 1752-0762 (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass

Tyas Mutiara Basuki , Andrew K. Skidmore , Patrick E. van Laake , Iris van Duren & Yousif A. Hussin

To cite this article: Tyas Mutiara Basuki , Andrew K. Skidmore , Patrick E. van Laake , Iris van Duren & Yousif A. Hussin (2012) The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass, Geocarto International, 27:4, 329-345, DOI: 10.1080/10106049.2011.634928

To link to this article: https://doi.org/10.1080/10106049.2011.634928

Geocarto International
Vol. 27, No. 4, July 2012, 329–345

Taylor & Francis
Taylor & Francis Group

The potential of spectral mixture analysis to improve the estimation accuracy of tropical forest biomass

Tyas Mutiara Basuki[a,b]*, Andrew K. Skidmore[a], Patrick E. van Laake[a], Iris van Duren[a] and Yousif A. Hussin[a]

[a]*ITC-Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands;* [b]*Forestry Research Institute of Solo, Jl. A. Yani – Pabelan, P.O. Box 295, Solo, Indonesia*

(*Received 21 June 2011; final version received 20 October 2011*)

A main limitation of pixel-based vegetation indices or reflectance values for estimating above-ground biomass is that they do not consider the mixed spectral components on the earth's surface covered by a pixel. In this research, we decomposed mixed reflectance in each pixel before developing models to achieve higher accuracy in above-ground biomass estimation. Spectral mixture analysis was applied to decompose the mixed spectral components of Landsat-7 ETM+ imagery into fractional images. Afterwards, regression models were developed by integrating training data and fraction images. The results showed that the spectral mixture analysis improved the accuracy of biomass estimation of *Dipterocarp* forests. When applied to the independent validation data set, the model based on the vegetation fraction reduced 5–16% the root mean square error compared to the models using a single band 4 or 5, multiple bands 4, 5, 7 and all non-thermal bands of Landsat ETM+.

Keywords: above-ground biomass; spectral mixture analysis; decomposition of mixed components; fraction endmembers; selective logging

1. Introduction

An accurate estimation of above-ground forest biomass, carbon stocks and change in biomass is essential to monitor impacts of forest management and policies. These require good quality spatial baseline information. Remote sensing has proven to be an important tool in the assessment of above-ground biomass at regional, national and global level (Brown 2002, Rosenqvist *et al*. 2003, Patenaude *et al*. 2005, UNFCCC 2009). The use of products from optical sensors, in combination with empirical models, is a commonly applied method to assess above-ground biomass. Using regression, a link may be made between spectral reflectance or vegetation indices and biomass (Foody *et al*. 2003, Lu *et al*. 2004, Okuda *et al*. 2004). The normalized difference vegetation index (NDVI) is the most frequently used compared to other vegetation indices (Lu *et al*. 2004). However, the results using this index for biophysical assessments in tropical forests are inconsistent. Accuracy

*Corresponding author. Email: basuki@itc.nl; tmbasuki@yahoo.com

ISSN 1010-6049 print/ISSN 1752-0762 online

http://dx.doi.org/10.1080/10106049.2011.634928
http://www.tandfonline.com

depends on which specific biophysical parameters need to be quantified and the characteristics of the study areas (Lu *et al.* 2004). Some of previous researchers found that NDVI is significantly correlated with above-ground biomass (González-Alonso *et al.* 2006, Zheng *et al.* 2007) while others found that it could not be applied to assess this parameter (Sader *et al.* 1989, Lu *et al.* 2004). These contradictory results can be attributed to saturation of the NDVI value at high biomass levels (Mutanga and Skidmore 2004, Okuda *et al.* 2004) as well as atmospheric contamination (Huete *et al.* 1994, Xiao *et al.* 2003).

To improve assessments, several models based on spectral reflectance or vegetation indices have been developed and tested. Lu *et al.* (2004) integrated forest inventory data, six reflective TM bands, and several vegetation indices to estimate above-ground biomass. They concluded that for forest with a complex stand structure, band TM5 and linear transformed indices were strongly correlated with the above-ground biomass. Steininger (2000) studied relationships between spectral reflectance generated from Landsat Thematic Mapper data in tropical forests of Brazil and Bolivia. It was observed that spectral band 5 was the best estimator for biomass in these Brazilian forests. However, it was restricted to dry weight biomasses estimates less than 150 ton ha^{-1} only.

The enhanced vegetation index (EVI) was developed to reduce atmospheric influences and improve signal sensitivity in high biomass regions (Huete *et al.* 1997) while the global environmental monitoring index (GEMI) was designed to minimize atmospheric and soil effects (Pinty and Verstraete 1992, McDonald *et al.* 1998). The latter index partially reduced background reflectance under sparse vegetation cover (McDonald *et al.* 1998). Modified soil adjusted vegetation index (MSAVI) and near infrared reflectance were also useful to retrieve biophysical parameters. Zheng *et al.* (2004) examined that MSAVI and infrared reflectance strongly correlated with above-ground biomass for pine forest. For lowland mixed *Dipterocarp* forests, however, the use of remote sensing based methods to estimate biomass or forest stand parameters highly varied in their results. Okuda *et al.* (2004) and Nssoko (2007) found very low correlation between above-ground biomass and Landsat reflectance values, NDVI, and EVI. In contrast to Okuda *et al.* (2004) and Nssoko (2007), Foody *et al.* (2003) obtained a relatively good coefficient of determination ($R^2 = 0.69$) when neural network was applied to predict above-ground biomass. In addition, Tangki and Chappell (2008) reported that average radiance in band 4 of Landsat-5 TM highly correlated ($R^2 = 0.76$) with biomass in the Ulu Segama Forest Reserve in Sabah, Malaysia, part of Borneo Island. But, the model used by Tangki and Chappell (2008) was not validated, which means that the applicability of the model for other areas is uncertain. Consequently, we have to look for a method to fill in this knowledge gap.

One of the limitations in estimating above-ground biomass using vegetation indices is that a signal recorded by a sensor in one pixel is in fact a spectral mixture of radiance or reflectance of all components on the earth's surface covered by that pixel. In spectral mixture analysis, each pixel is considered as a combination of multiple components weighted by relative surface abundance (Tompkins *et al.* 1997). This enables us to model an image as a linear mixture of a few basic pure spectral components known as endmembers (Tompkins *et al.* 1997, Small 2004, Adams and Gillespie 2006). Spectral mixture analysis has been used to determine and map urban vegetation abundances (Small 2001, 2003, 2005, Small and Lu 2006, Powell *et al.* 2007, Pu *et al.* 2008, Tooke *et al.* 2009). It was

also applied for long term monitoring of cover types over large areas using medium resolution imagery (Rogan *et al*. 2002, Hostert *et al*. 2003), coarse resolution imagery (DeFries *et al*. 2000) as well as combining medium and coarse resolution images (Uenishi *et al*. 2005).

Related to vegetation analysis, spectral mixture analysis is commonly applied to study vegetation abundance outside tropical forests. In tropical forests, this method is mostly used to assess canopy damage in Amazon forests (Souza and Barreto 2000, Souza *et al*. 2003, 2005). Souza *et al*. (2005) developed a normalized difference fraction index (NDFI). They combined it with a contextual classification algorithm to map canopy damage derived from logging, fires and other disturbances. This integrated method was successfully applied with an overall accuracy of 90.4%. In addition, Lu *et al*. (2003) used spectral mixture analysis in tropical forests for classifying successional and mature forests. It was observed that the spectral mixture analysis approach produced an overall classification accuracy of 78.2% and resulted in a 7.4% increase in accuracy, when compared to a maximum likelihood classification (Lu *et al*. 2003). Spectral mixture analysis is rarely applied for quantitative analysis in tropical forest. Lu *et al*. (2005) applied this method to estimate above-ground biomass in tropical forests. They found that spectral mixture analysis provided high relationship between the above-ground biomass and vegetation fraction in successional forest, but in the primary forest the correlation was very low.

Mixture of spectral components may happen in a selectively logged forest. The gaps in the forest canopy resulting from selective logging are smaller than the pixel size of optical sensors such as Landsat or ASTER. This means that pixels represent a mix of forest canopy cover, bare soil and other components that are found on the earth surface. Sist *et al*. (2003) studied the effect of conventional selective logging and reduced impact logging on canopy openness in *Dipterocarp* forests in East Kalimantan, Indonesia. Before logging, the mean canopy openness in conventional selective logging and reduced impact logging were 3.6% and 3.1%, respectively. After logging, it ranged from 17.5 to 20.7% in conventional selective logging and 4 to 18% in reduced impact logging. In the same areas, Vega (2005) found that felling a single tree in such forests created a gap fraction ranging from 30 to 100% for pixels with a resolution of 15 m.

Having identified the lack of accurate above-ground biomass assessments and the potential of applying spectral mixture analysis for quantitative analysis in selectively logged tropical forests, this study aims to test the potential of spectral mixture analysis to improve the estimation accuracy of above-ground biomass in *Dipterocarp* forests. These forests cover extensive areas in tropical South-East Asia, so it is necessary to obtain a suitable method for estimating its biomass. The improvement of the accuracy of biomass estimation can be expected because spectral mixture analysis has potential to obtain the proportion of vegetation from mixture spectral components in the study area, and thus the main component which contributes to above-ground biomass estimation.

In developing the models based on spectral mixture analysis, all of non-thermal bands of Landsat-7 ETM+ were used. Therefore, as the comparison to the proposed models, we generated a multiple regression based on the spectral band 1, 2, 3, 4, 5 and 7 of the Landsat-7 ETM+. In addition, moisture vegetation indices, a single band and multiple bands 4, 5 and 7 of Landsat-7 ETM+ were used as the estimators for the above-ground biomass.

2. Materials and methods

2.1 *Study area*

The research was conducted in Labanan concession forest, Berau Regency, East Kalimantan, Indonesia. The concession lies between 1°45′ and 2°10′ north latitude and between 116°55′ and 117°20′ east longitude, the situation map is illustrated in Figure 1.

The concession area is 83,240 ha. The forest type in the study area is called lowland *Dipterocarp* forest. It is dominated by the family of Dipterocarpaceae. According to Sist and Saridan (1998), this family contributes about 25% of the total tree density, 50% of the total tree basal area and 60.2% of the stand volume. The second most abundant family is Euphorbiaceae comprising 13.5% of the total tree density and 9.1% of the basal area. Selective logging was applied for commercial timber harvesting from 1976 to 2003.

2.2 *Image pre-processing*

A Landsat-7 ETM+ image (path 117 and row 59) acquired in 31 May 2003 was used to generate spectral reflectance and moisture vegetation indices. Image pre-processing was conducted to minimize atmospheric, geometric and radiometric errors. Haze removal was applied to reduce atmospheric effects. Geometric correction of the image using a first order polynomial was carried out using 15 ground control points (GCPs) collected from the study area (Wijaya 2005), resulting in a total root mean squared error (RMSE) less than a 0.5 pixel. The image was geo-rectified to the universal transverse mercator (UTM) coordinate system with datum WGS 1984 and zone 50 North and re-sampled to 32 m spatial resolution. Prior to derivation of the spectral indices and spectral mixture analysis, the digital numbers (DNs) of the image were converted to radiance and then to reflectance using an exoatmospheric model described by NASA (2005).

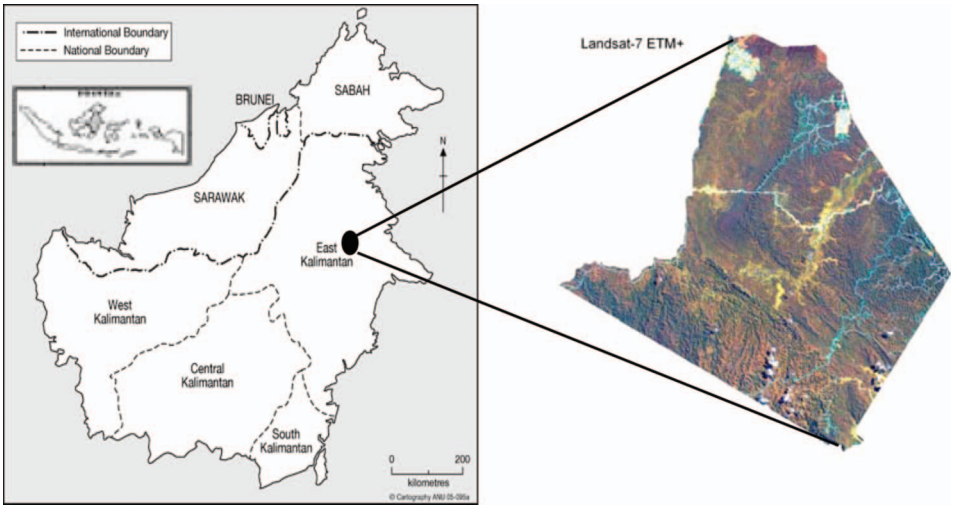


Figure 1. Location of the study area.

2.3 *Spectral mixture analysis*

The spectral mixture analysis method used in this study (Tompkins *et al.* 1997) is formulated below:

$$R_b = \sum_{i=1}^{m} f_i r_{i_b} + E_b \quad (1)$$

where R_b is the reflectance of a pixel at band b, f_i is the fractional abundance of endmember i (from a total of m endmembers), r_{i_b} is the reflectance at band b of endmember i and E_b is the error in band b of the model fit.

The spectral mixture analysis was run using environment for visualizing images (ENVI) 4.7 and interactive data language (IDL) 7.1 software. To reduce noise, reflectance bands 1, 2, 3, 4, 5 and 7 of Landsat 7 ETM+ were transformed using minimum noise fraction. Afterwards, the minimum noise fraction image was analysed to obtain the spectrally most pure pixel, using the pixel purity index (Boardman *et al.* 1995). The highest pixel purity index values were used to identify the location of sample endmembers on the original image (Souza *et al.* 2005). The sample endmembers were also evaluated by their location on the extremes of the image feature space by assuming that these represent the purest pixels in the images (Lu *et al.* 2003, 2005). Endmember selection is a crucial step, since this determines the success of the spectral unmixing. Improper choice of endmembers will cause negative and super positive (>1) fraction images. By prior knowledge of the study area and trial and error of endmember selection, it was found that for the study area the most suitable endmembers consisted of vegetation, soil and shade. Using the selected endmembers, spectral mixture analysis was applied resulting in the fraction images for vegetation, soil, shade and error fractions. The fraction images were investigated to identify fraction overflow or fractions that had values less than 0 or more than 1. The spectral mixture analysis was run iteratively until a minimum fraction overflow was obtained. The number of pixel having fraction overflow was quantify using a developed script and run by ENVI-IDL software. The most suitable model with an average fraction overflow of 3% was used for further analysis. Water bodies, settlements, shrubs and roads were masked because these features were not of interest in the study.

2.4 *Spectral indices*

Spectral indices including NDVI, simple ratio, EVI, advanced vegetation index, soil adjusted vegetation index, MSAVI, GEMI and atmospherically resistant vegetation index have been applied for above-ground biomass assessment in the study area previously (Nssoko 2007, Wijaya *et al.* 2010). In the current research, two single bands (band 4 or 5) and the moisture vegetation indices based on the combination of reflectance band 4 and 5 (MVI5) and the combination of reflectance band 4 and 7 (MVI7) were applied. The single bands and indices were chosen because these had a better performance for biophysical estimations in other tropical forests (Steininger 2000, Lu *et al.* 2004, Freitas *et al.* 2005, Tangki and Chappell 2008). The formulas for moisture vegetation indices of MVI5 and MVI7 as follows:

$$\mathrm{MVI5} = (\mathrm{NIR} - \mathrm{MIR5})/(\mathrm{NIR} + \mathrm{MIR5}) \quad (2)$$

$$\mathrm{MVI7} = (\mathrm{NIR} - \mathrm{MIR7})/(\mathrm{NIR} + \mathrm{MIR7}) \quad (3)$$

where, MVI5 = moisture vegetation index band 5; MVI7 = moisture vegetation index band 7; NIR = near infrared reflectance (band 4) of Lansad-7 ETM+; MIR5 = middle infrared of band 5 Landsat-7 ETM+; MIR7 = middle infrared of band 7 Landsat-7 ETM+.

In addition, multiple linear regression models were developed using all bands of Landsat-7 ETM+ (bands 1–5 and 7) and also from the first three bands which have the highest correlation with the total above-ground biomass. These multiple bands have not been applied for the current study area.

2.5 *Estimation of reference above-ground biomass*

Seventy seven (77) random sampling plots were identified based on a Landsat-7 ETM+ (31 May 2003). The plots were divided randomly into two groups, 50 plots were used for developing models and 27 plots for validations. The location of the plots distributed from flat on swampy areas and near rivers to undulating and hilly regions. The terrain had flat to steep slopes. The protected forests were located on hilly areas with steep slopes. The coordinates of the plots were recorded using GPS.

A circular plot with the size of 500 m^2 was used. A slope correction was applied, in case the plot was located on a slope. The radius of the circle ranged from 12.62 m to 15.78 m. Within each of 77 plots, tree diameters equal or greater than 10 cm were measured using a meter tape at breast height (DBH). In this study, woody plants which have DBH equal to or greater than 10 cm are defined as trees (Lu *et al*. 2005). The above-ground biomass in this study is restricted to the above-ground of the trees which have DBH $\geq$10 cm.

The above-ground biomass was estimated using a local allometric equation developed by Basuki *et al*. (2009). This equation was developed using a regression analysis of 122 trees with diameters ranging from 5 to 200 cm and consisting of 48 species. The equation to estimate above-ground biomass was:

$$\ln(AGB) = 2.196 \times \ln(DBH) - 1.201 \quad (4)$$

where, AGB is above-ground biomass in dry weight (kg/tree) and DBH in cm. The adjusted R^2 of the model is 0.963. The above-ground biomass was obtained by de-transforming *ln* above-ground biomass values. The descriptive statistics of the above-ground biomass of the sampling plots for the training and validation data are presented in Table 1.

Table 1. Descriptive statistics of the above-ground biomass (ton ha^{-1}) for the training (50 plots) and the validation (27 plots) data.

Above-ground biomass (ton ha^{-1})	Training samples (50 plots)	Validation samples (27 plots)
Minimum	102.62	102.26
Mean	343.79	329.20
Maximum	839.32	812.15
Standard deviation	156.61	191.98

2.6 *Statistical analysis*

Prior to development models, Pearson correlation coefficient was used to evaluate the strength of linear correlation between two variables. Multiple and simple linear regressions were employed to establish relationships between above-ground biomass and the independents variables. The independent variables consisted of spectral band 4; band 5; band 4, 5, 7; band 1, 2, 3, 4, 5, 7; MVI5; MVI7 and the vegetation, soil, and shade values generated from the fraction images. The values of the independent variables were calculated as the average of a 3 by 3 pixel window (Lu *et al*. 2004). Multicolinearity was inspected for multiple linear regressions models. Variance inflation factor (Brauner and Shacham 1998) was used to determine whether a model has multicolinearity between the independent variables. The most suitable model for estimating the above-ground biomass was selected based on the value of coefficient determination (R^2) of the regression (Lu *et al*. 2005) and for the model which minimized the RMSE on the validation data.

3. Results

3.1 *Fraction images*

The fraction images produced from the spectral mixture analysis highlight the heterogeneity of the forest cover in the study area (Figure 2). Due to dense vegetation, mature forests in the unlogged areas and advanced successional forests which were logged more than 20 years ago are shown in light grey. They also have higher vegetation fraction values when compared to forests logged in 2003 as illustrated in the vegetation fraction image in Figure 2(a). In the soil fraction image (Figure 2(b)), light grey areas where represent pixels with a high fraction of bare soil. On the fraction shade image (Figure 2(c)), unlogged forests have a darker colour indicating a lower fraction of shade compared to logged over forests.

3.2. Remote sensing based estimation of above-ground biomass

Table 2 shows the Pearson correlation between fraction images, spectral reflectance of Landsa-7 ETM+ and the above-ground biomass. It can be seen that vegetation and shade fraction have strong Pearson correlation with the above-ground biomass. For the Landsat-7 ETM+, band 4 has the highest correlation with the above-ground biomass followed by band 5 and 7. Therefore, these three bands together were also used to generate multiple linear regressions to predict the above-ground biomass.

The statistical analysis of the regression models and the independent validation are provided in Table 3. The model estimators based on the vegetation indices and a single band or band combination of the Landsat-7 ETM+ provides coefficient determination less than 0.6 (Table 3). Among these estimators, the use all non-thermal bands (bands 1–5 and 7) produces the highest R^2, and the lowest are found for moisture vegetation indices.

Compared to the estimators of band 4, band 5, and non-thermal band of Landsat-7 ETM+ as well as moisture vegetation indices, the models based on the fraction images improve the R^2 of the regression models and reduced the RMSE of the validation data as presented in Table 3. In this case, the multiple linear regressions based on the vegetation and soil fractions and the simple linear regression of the vegetation fraction have a similar coefficient of determination.

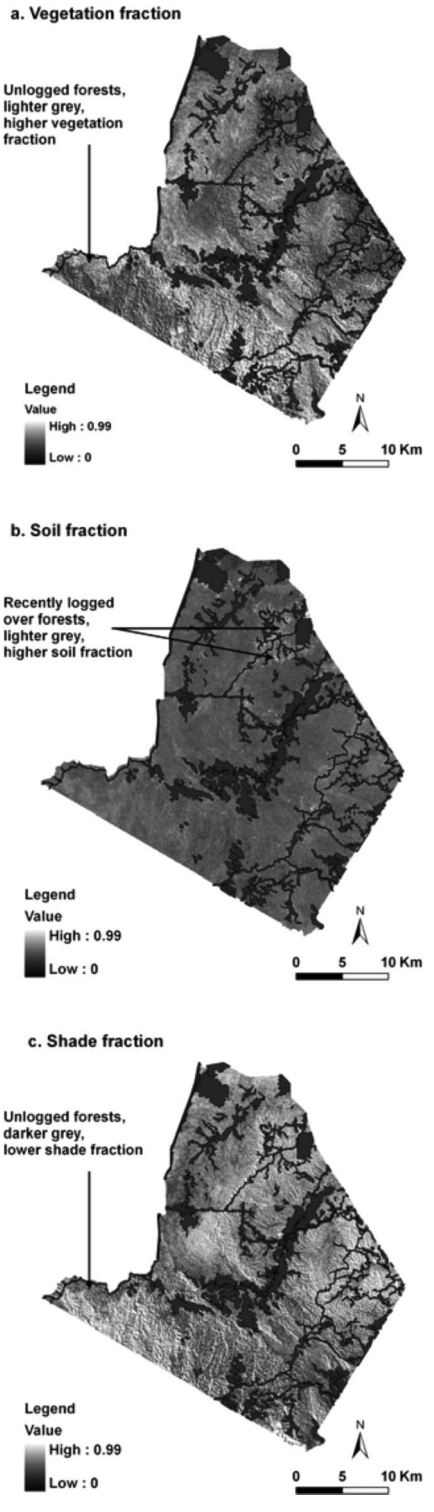


Figure 2. Fraction endmembers of Landsat-7 ETM+ produced by spectral mixture analysis, vegetation (a), soil (b) and shade (c).

Table 2. Pearson correlation at 95% confidence interval between the above-ground biomass (ton ha^{-1}) and the fraction endmembers or spectral reflectance of Landsat-7 ETM+ for 50 sample plots.

Fraction endmembers			Landsat-7 ETM+		
Fraction	Pearson correlation	*p*-value	Band	Pearson correlation	*p*-value
Vegetation	0.800	0.000	1	0.205	0.076^{ns}
Soil	−0.312	0.014	2	0.231	0.053^{ns}
Shade	−0.756	0.000	3	−0.097	0.252^{ns}
–	–	–	4	0.757	0.000
–	–	–	5	0.564	0.000
–	–	–	7	0.369	0.040

Note: ns, non-significant at 95% confidence interval.

Table 3. Linear regression between remote sensing based vegetation and soil fractions or the spectral reflectance of Landsat-7 ETM+ and the total above-ground biomass (ton ha^{-1}) using the training data (50 sample plots) and its RMSE of the independent validation data (27 sample plots).

	Regression model			
Independent variable	Equation	Adjusted R^2	Significance level of *p*-value	Validation RMSE (ton ha^{-1})
Vegetation fraction	AGB = −258.055 + 1350.342 × veg	0.632	0.000	130.4
Vegetation and soil fractions	AGB = −200.793 + 1303.543 × veg − 723.260 × soil	0.635	0.000	133.4
Reflectance band 4	AGB = −1663.499 + 8461.377 × ref4	0.565	0.000	137.8
Reflectance band 5	AGB = −1075.753 + 14811.499 × ref5	0.304	0.000	155.5
Reflectance band 4, 5, 7	AGB = −1658.158 + 8058.407 × ref4 + 2778.176 × ref5 − 5691.019 × ref7	0.548	0.000	137.5
Reflectance band 1, 2, 3, 4, 5, 7	AGB = −1272.352 + 3766.575 × ref1 + 8502.767 × ref2 − 33838.339 × ref3 + 7325.874 × ref4 + 5086.444 × ref5 − 4270.166 × ref7	0.582	0.000	144.0
MVI5	AGB = −390.551 + 1729.827 × MVI5	0.021	0.160^{ns}	–
MVI7	AGB = −1779.172 + 2759.609 × MVI7	0.041	0.085^{ns}	–

Note: veg, fraction of vegetation; soil, fraction of soil; ref, spectral reflectance; MVI5, moisture vegetation index band 4 and 5; MVI7, moisture vegetation index band 4 and 7; ns, non-significant at 95% confidence interval; RMSE, root mean squared error.

When all the fraction images were employed to predict above-ground biomass, multicolinearity occurred between the vegetation and the shade fraction, therefore the shade fraction was not applied to develop the models.

To test the applicability of the proposed regression models, validation was conducted using independent data from 27 plots and the results are also provided in Table 3. The validation was conducted for each regression model, except for MVI5 and MVI7 because these vegetation indices have low correlation with the above-ground biomass. Among the regression models, the proposed model developed from the vegetation fraction has the lowest RMSE of the validation data. The scatter plots of the measured and the predicted above-ground biomass using independent data are illustrated in Figure 3.

4. Discussion

The proposed models based on the fraction images have a higher adjusted R^2 compared to the moisture vegetation indices and the simple and multiple regression models of the Landsat-7 ETM+ bands. This indicates the importance of decomposing spectral reflectance of pixels in medium resolution images such as Landsat-7 ETM+ into fractions of vegetation, soil and shade. The decomposition is essential to obtain the proportion of the vegetation within the mixed components since this is the variable ultimately needed to estimate the above-ground biomass. Besides this, separation of shade from the vegetation is needed because in a complex forest structure, canopy shading is an important aspect influencing vegetation reflectance captured by a sensor, and consequently it could affect biomass estimation (Steininger 2000, Asner and Warner 2003).

The decomposition of mixed endmembers is useful for qualitative as well as quantitative analysis. Visual interpretation of fraction images helps to differentiate newly harvested areas from primary or mature forests. This phenomenon supports previous research undertaken by Lu *et al.* (2003) in Amazonian tropical forests. They applied spectral mixture analysis and concluded that secondary succession forests could be discriminated from mature forests due to different proportion of vegetation, shade and soil covered by a pixel. Besides that, Souza *et al.* (2005) found that intact forests had significantly higher green vegetation fractions than that of conventionally logged forests, as well as logged and burned forests.

The models explain 63% of the variation of the above-ground biomass. This coefficient of determination is lower than that obtained by Lu *et al.* (2005) in the successional forests in Brazil. In the successional forests, Lu *et al.* (2005) found that their regression model derived from a vegetation fraction image produced a coefficient of determination of around 0.80. In the primary forest, the shade fraction image only explained less than 20% of the variation of above-ground biomass. Possible reasons for the different results in the current study and the research conducted by Lu *et al.* (2005) are the differences in the forest stand structures and thus the determination of the endmembers which is site specific. The training data to develop our models consist of sample plots from the successional and primary forests, whereas Lu *et al.* (2005) developed the regression model separately from the successional and primary forest.

The proposed models are a major improvement compared to the conventional models. They have a higher coefficient of determination compared to the previous study conducted by Wijaya *et al.* (2010) at the same study area. In the current study, the fraction images explained 63% of the variation in the above-ground biomass. By contrast, Wijaya *et al.* (2010) found that the coefficient of determination between above-ground biomass with the common spectral indices e.g. SR, NDVI, EVI,

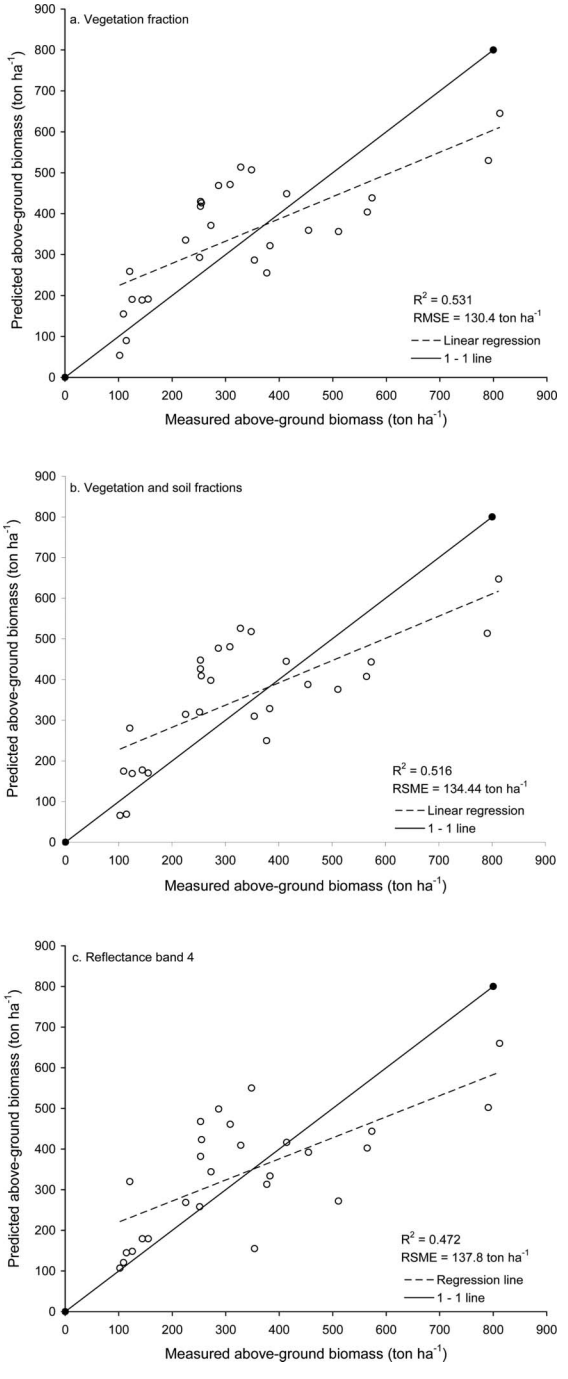
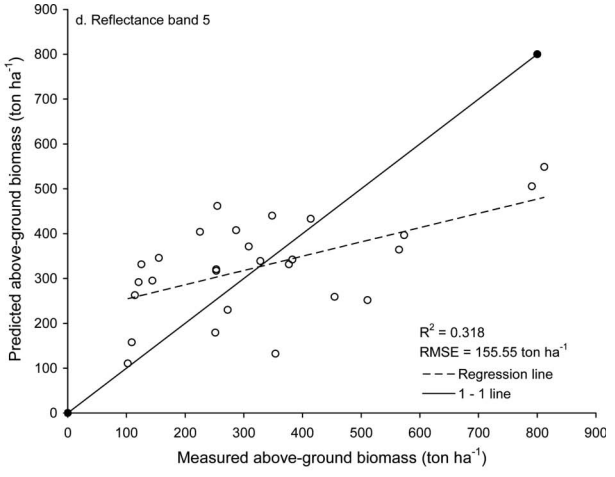
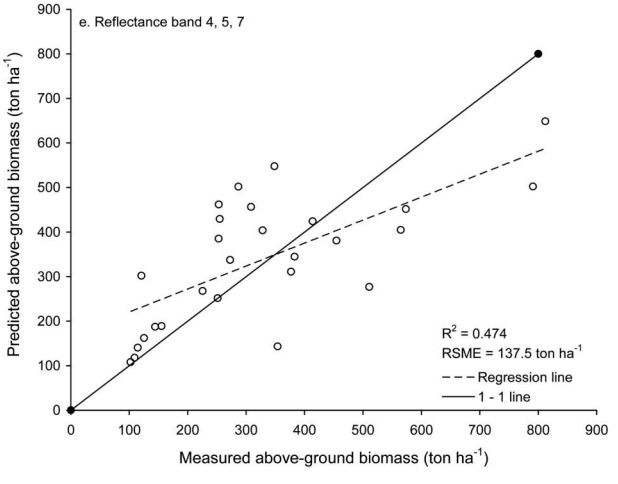


Figure 3. Scatter plots of the measured and the predicted values of above-ground biomass (ton ha^{-1}) using different independent variables applied to the validation data set. The measured total above-ground biomass was the result of applying an allometric equation. The predicted values were calculated using regression models including vegetation fraction (a), vegetation and soil fractions (b), reflectance of Landsat-7 ETM+ band 4 (c), band 5 (d), bands 4, 5, 7 (e), and bands 1, 2, 3, 4, 5 and 7 (f).





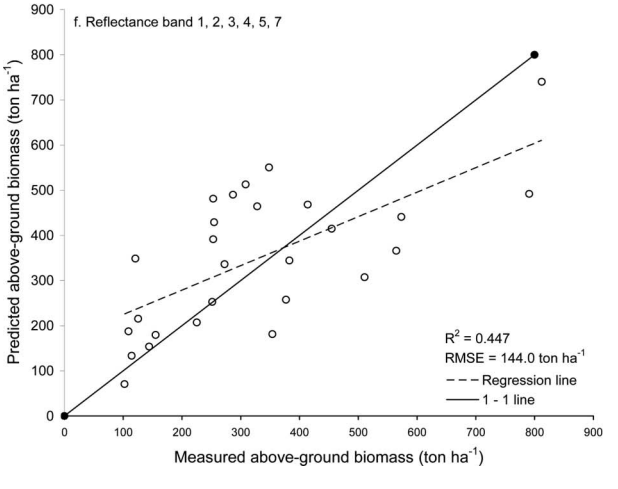


Figure 3. (*Continued*).

ARVI, SAVI, GEMI, Tasseled caps and grey level co-occurrence matrix mean (GLCM_mean) was less than 0.296. The improved accuracy of above-ground biomass estimation using the fraction images is because in the spectral mixture analysis the proportion of the vegetation, soil and shade has been determined before developing the regression models. Further analysis of the data revealed that the models based on fraction images reduced the RMSE compared to a single or multiple bands spectral reflectance of Landsat-7 ETM+ when applied to the independent validation data set (Table 3). Based on this solid analysis, we can conclude that breaking down images into vegetation fraction images provides the best predictor for biomass and therefore this methodology should be used in similar studies.

The use of all bands (1–5 and 7) of the Landsat-7 ETM+ produces a higher R^2 value of the regression model compared to band 4 and combination of band 4, 5 and 7. However, the improvement is not significant because the Pearson correlation between the visible bands and the above-ground biomass is low (Table 2).

The results of using single bands (e.g. band 4 or 5) for the current forest area are different from the research undertaken by Steininger (2000) and Tangki and Chappell (2008). These could be caused by differences in forest stand structure and the amount of biomass. The regression model of Steininger (2000) was based on secondary forest with a maximum dry weight of above-ground biomass of 150 ton ha^{-1}. Steininger found a strong coefficient of determination ($R^2 = 0.73$) between above-ground biomass and TM band 5. In our study, the maximum dry weight of the above-ground biomass was 839 ton ha^{-1} and originated from logged over and primary forests. Although Tangki and Chappell (2008) conducted their research in *Dipterocar*p forest, similar to our study area, but they reported a maximum dry weight of the above-ground biomass in undisturbed forest of around 500 ton ha^{-1}.

The results of using moisture vegetation indices as biomass estimators vary from one forest to another. In our forest area, these indices were not suitable to predict above-ground biomass. However, Boyd *et al.* (1999) and Lu *et al.* (2004) found a better correlation between moisture vegetation indices and biomass. Boyd *et al.* (1999) obtained R^2 of 0.22 when moisture vegetation index from near infra red and middle infra red of NOAA advanced very high resolution radiometer (AVHRR) was applied to estimate above-ground biomass in tropical forest in southern Cameroon. Lu *et al.* (2004) observed that moisture vegetation indices produced different correlation with above-ground biomass in Altamira, Bragantina and Penta de Pedras which are located in Brazilian Amazon. The moisture vegetation index of Landsat TM band 4 and 5 was used to estimate above-ground biomass with R^2 of 0.40, 0.18 and 0.19, respectively for Altamira, Bragantina and Penta de Pedras. The reason for these different results is because of variations in forests structure including tree height, species composition, and the amount of biomass and vegetation vigor. These phenomena imply that application of vegetation indices or spectral signature of Landsat TM or ETM+ depends on the biophysical characteristics of the forests and is site specific.

Looking at the validation, the scatter plots of the models derived from fraction images produce similar patterns (Figure 3(a) and (b)). The model derived from reflectance of band 4 yields over estimation for the above-ground biomass less than 354 ton ha^{-1} (Figure 3(c)). The regression model of bands 4, 5, 7 and the non-thermal band of Landsat-7 ETM+ show a similar pattern (Figure 3(e) and (f)). However, all the graphs show the increases of the measured above-ground biomass are

followed by the increase of predicted ones until an above-ground biomass reaches around 400 ton ha^{-1}. After that point, the predicted values are lower than the measured ones. These phenomena are probably saturation effects, where the increase above-ground biomasses are not followed by the increase of predictor variables.

The spectral mixture analysis has potential to improve the accuracy of the above-ground biomass estimation compared to the common vegetation indices or spectral reflectance. However, this methodology needs expertise and careful selection of endmembers as this is the crucial step in the analysis. In determining endmembers, especially using pixel purity index, the process is interactive in comparison to spectral reflectance or vegetation indices that directly derived from the image. The proposed models can be applied to estimate above-ground biomass estimation up to 400 ton ha^{-1}, above this value the saturation problem may be present.

In this research, only three endmembers are applied. Including non-photosynthetic vegetation such as branches and stems may be more appropriate for these forests. Additional non-photosynthetic vegetation endmembers have been successfully applied to map forest degradation (Souza *et al.* 2003) and to map canopy damage from selective logging and forest fires (Souza *et al.* 2005). For the forest in this study, however, determining such endmembers is difficult because the resulting fraction images were overflow having negative and super positive (>1) values.

5. Conclusions

Our findings demonstrate that spectral mixture analysis can increase the accuracy of above-ground biomass assessment for mixed *Dipterocarps* forests. The proposed models based on the spectral mixture analysis explain 63% of the variability the above-ground biomass. The proposed models improve the estimation accuracy of the above-ground biomass compared to more conventional models, such as spectral reflectance of band 4; band 5; bands 4, 5, 7; all non-thermal band and moisture vegetation indices. Decomposing mixed spectral components into fractions of individual components is essential before developing the regression models in selective logging forests. Endmembers selection is a critical step in the methodology. The problem in dense forest with complex stand structure and very high biomass is that saturation in light absorption occurs, which cannot be solved by applying this technique.

Acknowledgements

We would like to thank the staff of Labanan concession area, Bp. Ir. Dody Herika and his staff for their assistance during field campaign; Willem Nieuwenhuis, ITC, for helping with some scripts, and Dr. Nichola M. Knox, ITC, for her assistance in proof reading. This research is supported by a grant from the Netherlands Organization for International Cooperation in Higher Education (NUFFIC).

References

Adams, J.B. and Gillespie, A.R., 2006. *Remote sensing of landscapes with spectral images. A physical modeling approach*. UK: Cambridge Univ. Press, 362.

Asner, G.P. and Warner, A.S., 2003. Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Remote Sensing of Environment*, 87, 521–533.

Basuki, T.M., *et al*., 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. *Forest Ecology and Management*, 257, 1684–1694.

Boardman, J.M., Kruse, F.A., and Green, R.O., 1995. Mapping target signature via partial unmixing of AVIRIS data. *Summaries of the Fifth JPL Airborne Earth Science Workshop*. Washington, DC: JPL Publication 95-1, 23–26.

Boyd, D.S., Foody, G.M., and Curran, P.J., 1999. The relationship between the biomass of Cameroonian tropical forests and radiation reflected in middle infrared wavelengths (3.0–5.0 μm). *International Journal of Remote Sensing*, 20, 1017–1023.

Brauner, N. and Shacham, M., 1998. Identifying and removing sources of imprecision in polynomial regression. *Mathematics and Computers Simulation*, 48, 75–91.

Brown, S., 2002. Measuring carbon in forests: current status and future challenges. *Environmental Pollution*, 116, 363–372.

DeFries, R.S., Hansen, M.C., and Townshend, J.R.G., 2000. Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8 km AVHRR data. *International Journal of Remote Sensing*, 21, 1389–1414.

Foody, G.M., Boyd, D.S., and Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between region. *Remote Sensing of Environment*, 85, 463–474.

Freitas, S.R., Mello, M.C.S., and Cruz, C.B.M., 2005. Relationships between forest structure and vegetation indices in Atlantic rainforest. *Forest Ecology and Management*, 218, 353–362.

González-Alonso, F., *et al*., 2006. Forest biomass estimation through NDVI-composites. The role of remotely sensed data to assess Spanish forests as carbon sinks. *International Journal of Remote Sensing*, 27 (24), 5409–5415.

Hostert, P., Röder, A., and Hill, J., 2003. Coupling spectral unmixing and trend analysis for monitoring long-term vegetation dynamics in Mediterranean rangelands. *Remote Sensing of Environment*, 87, 183–197.

Huete, A., Justice, C., and Liu, H., 1994. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, 49, 224–234.

Huete, A.R., *et al*., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, 59, 440–451.

Lu, D., Moran, E., and Batistella, M., 2003. Linear mixture model applied to Amazonian vegetation classification. *Remote Sensing of Environment*, 87, 456–469.

Lu, D., Batistella, M., and Moran, E., 2005. Satellite estimation of aboveground biomass and impacts of forest stand structure. *Photogrammetric Engineering & Remote Sensing*, 71 (8), 967–974.

Lu, D., *et al*., 2004. Relationship between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, 198, 149–167.

McDonald, A.J., Gemmell, F.M., and Lewis, P.E., 1998. Investigation of the utility of spectral vegetation indices for determining information on coniferous forests. *Remote Sensing of Environment*, 66, 250–272.

Mutanga, O. and Skidmore, A.K., 2004. Hyperspectral band depth analysis for a better estimation of grass biomass (*Cenchrus ciliaris*) measured under controlled laboratory conditions. *International Journal of Applied Earth Observation and Geoinformation*, 5, 87–96.

NASA, 2005. *Landsat 7 science data users handbook* [online]. Available from: http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html [Accessed on 30 June 2005].

Nssoko, G.E., 2007. *Sensitivity of spectral vegetation indices to trees biomass in tropical rainforests: a case of Labanan concession area, East Kalimantan, Indonesia*. Unpublished thesis (MSc). ITC, Enschede, 48.

Okuda, T., *et al*., 2004. Estimation of aboveground biomass in logged and primary lowland rainforests using 3-D photogrammetric analysis. *Forest Ecology and Management*, 203, 63–75.

Powell, R.B., *et al*., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, 106, 253–267.

Patenaude, G., Milne, R., and Dawson, T.P., 2005. Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science and Policy*, 8, 161–178.

Pinty, B. and Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio*, 101, 15–20.

Pu, R., *et al*., 2008. Spectral mixture analysis for mapping abundance of urban surface components from the Terra/ASTER data. *Remote Sensing of Environment*, 112, 939–954.

Rogan, J., Franklin, J., and Roberts, D.A., 2002. A comparison of methods for monitoring multi temporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment*, 80, 143–156.

Rosenqvist, Å., *et al*., 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science and Policy*, 6, 441–455.

Sader, S.A., *et al*., 1989. Tropical forest biomass and successional age class relationship to vegetation index derived from LANDSAT TM data. *Remote Sensing Environment*, 28, 143–156.

Sist, P. and Saridan, A., 1998. Description of the primary lowland forest of Berau. *In*: J.-G. Bertault and K. Kadir, eds. *Sylvicultural research in a lowland mixed dipterocarp forest of East Kalimantan*. Cirad-Forêt Publication, 255.

Sist, P., *et al*., 2003. Reduced-impact logging in Indonesian Borneo: some results confirming the need for new silvicultural prescriptions. *Forest Ecology and Management*, 179, 415–427.

Small, C., 2001. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, 22 (7), 1305–1334.

Small, C., 2003. High spatial resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment*, 88, 170–186.

Small, C., 2004. The Landsat ETM+ spectral mixing space. *Remote Sensing of Environment*, 93, 1–17.

Small, C., 2005. A global analysis of urban reflectance. *International Journal of Remote Sensing*, 26 (4), 661–681.

Small, C. and Lu, J.W.T., 2006. Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis. *Remote Sensing of Environment*, 100, 441–456.

Souza, C., Jr. and Barreto, P., 2000. An alternative approach for detecting and monitoring selectively logged forests in the Amazon. *International Journal of Remote Sensing*, 21 (1), 173–179.

Souza, Jr. C., *et al*., 2003. Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models. *Remote Sensing of Environment*, 87, 494–506.

Souza, Jr., C.M., Roberts, D.A., and Cocrane, M.A., 2005. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. *Remote Sensing of Environment*, 98, 329–343.

Steininger, M.K., 2000. Satellite estimation of tropical secondary forest above-ground biomass data from Brazil and Bolivia. *International Journal of Remote Sensing*, 21 (6–7), 1139–1157.

Tangki, H. and Chappell, N.A., 2008. Biomass variation across selectively logged forest within a 225-km^2 region of Borneo and its prediction by Landsat TM. *Forest Ecology and Management*, 256, 1960–1970.

Tompkins, S., *et al*., 1997. Optimization of endmembers for spectral mixture analysis. *Remote Sensing of Environment*, 59, 472–489.

Tooke, T.R., *et al*., 2009. Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications. *Remote Sensing of Environment*, 113, 398–407.

Uenishi, T.M., *et al*., 2005. A land cover distribution composite image from coarse spatial images using an unmixing method. *International Journal of Remote Sensing*, 26 (5), 871–886.

UNFCCC, 2009. Cost of implementing methodologies and monitoring systems relating to estimates of emissions from deforestation and forest degradation, the assessment of carbon stocks and greenhouse gas emissions from changes in forest cover, and the enhancement of forest carbon stocks. Technical report.

Vega, B., 2005. Image fusion of optical and microwave data to assess criteria and indicator (C&I) related to forest encroachment, for certification process of sustainable forest management (SFM). Unpublished thesis (MSc). ITC, Enschede, 79.

Wijaya, A., 2005. Application of multi-stage classification to detect illegal logging with the use multi-source data: a case study in Labanan forest management unit, East Kalimantan, Indonesia. Unpublished thesis (MSc). ITC, Enschede, 64.

Wijaya, A., *et al*., 2010. Improved strategy for estimating stem volume and forest biomass using moderate resolution remote sensing data and GIS. *Journal of Forestry Research*, 21 (1), 1–12.

Xiao, X., *et al*., 2003. Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia. *Remote Sensing of Environment*, 84, 385–392.

Zheng, D., *et al*., 2004. Estimating aboveground biomass using Landsat 7 ETM^{+} data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment*, 93, 402–411.

Zheng, G., *et al*., 2007. Combining remote sensing imagery and forest age inventory for biomass mapping. *Journal of Environmental Management*, 85, 616–623.