

# A first map of tropical Africa's above-ground biomass derived from satellite imagery

A Baccini<sup>1</sup>, N Laporte<sup>1</sup>, S J Goetz<sup>1</sup>, M Sun<sup>1</sup> and D. Huang<sup>2</sup>

<sup>1</sup> The Woods Hole Research Center, 149 Woods Hole Road, Falmouth, MA 02540, USA

<sup>2</sup> Atmospheric Sciences Division, Brookhaven National Laboratory, Upton, NY 11973, USA

E-mail: [abaccini@whrc.org](mailto:abaccini@whrc.org)

Received 6 May 2008

Accepted for publication 19 November 2008

Published 10 December 2008

Online at [stacks.iop.org/ERL/3/045011](http://stacks.iop.org/ERL/3/045011)

## Abstract

Observations from the moderate resolution imaging spectroradiometer (MODIS) were used in combination with a large data set of field measurements to map woody above-ground biomass (AGB) across tropical Africa. We generated a best-quality cloud-free mosaic of MODIS satellite reflectance observations for the period 2000–2003 and used a regression tree model to predict AGB at 1 km resolution. Results based on a cross-validation approach show that the model explained 82% of the variance in AGB, with a root mean square error of 50.5 Mg ha<sup>-1</sup> for a range of biomass between 0 and 454 Mg ha<sup>-1</sup>. Analysis of lidar metrics from the Geoscience Laser Altimetry System (GLAS), which are sensitive to vegetation structure, indicate that the model successfully captured the regional distribution of AGB. The results showed a strong positive correlation ( $R^2 = 0.90$ ) between the GLAS height metrics and predicted AGB.

**Keywords:** biomass, carbon, MODIS, lidar, GLAS, regression tree, Random Forest, REDD

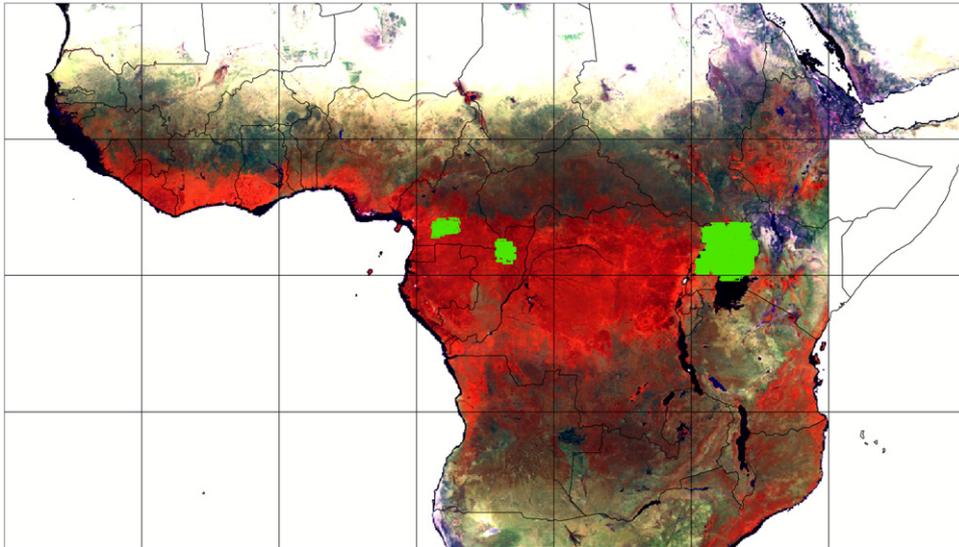
## 1. Introduction

Forests contain about 80% of global terrestrial above-ground carbon stocks (biomass), and play an important role in the global carbon cycle (Houghton 2005). Tropical forests are a strong carbon sink (Stephens *et al* 2007) and tropical deforestation contributes about one fifth of total anthropogenic CO<sub>2</sub> emissions to the atmosphere (Houghton 2007). Refining these estimates requires improved knowledge of the density and spatial distribution of forest biomass across the globe, particularly in high biomass tropical forest ecosystems.

Africa has the second largest block of rainforest in the world, next to the Amazon basin, but is the least known in terms of carbon stocks and rates of forest conversion. The existing biomass estimates are derived from national or partial forest inventories that provide precise and accurate estimates at the plot or local level, but much less accurate information over broader spatial scales. This is partly because Africa is diverse in terms of the wide range of ecosystems it includes,

which range from xeric shrublands in the Transvaal region of the south and the Sahelian zone in the north to the dense humid forests of the Congo Basin countries (White 1983). It is also due in part to the very wide range of biomass estimates associated with these diverse ecosystems (e.g. Gibbs *et al* 2007), degradation occurring in the Congo Basin associated with industrial logging (Laporte *et al* 2007) and deforestation for agriculture (Hansen *et al* 2008). But mostly, Africa is least known because it has often been a difficult place to work as a result of political instability, a diversity of languages and cultures, and limited infrastructure to support scientific research (much more limited than, for example, countries of the Amazon Basin or Southeast Asia). As a result, few countries in Africa have forest inventories and many are obsolete.

The United Nations Framework Convention on Climate Change (UNFCCC) recognized the important role of deforestation in the carbon cycle and discussions have been initiated to reduce emissions from deforestation and



**Figure 1.** Pixel mosaic of MODIS NBAR data (RGB 2, 6, 1). The black lines define the MODIS tiles and country borders. The green dots indicate the location of biomass data.

degradation (REDD) in developing countries. The suggested schemes for carbon credit allocation based on deforestation (Mollicone *et al* 2007) or carbon stock baselines (Gurney and Raymond 2008) require accurate estimates of carbon stock.

There have been no comprehensive studies that used remotely sensed data to map the spatial distribution of forest biomass for Africa. The most recent estimates are derived from studies based on applying field measurements to forest cover type classes (Gibbs *et al* 2007). While this is an approach that has utility and has been used frequently in the past, it can miss information on the variability of forest biomass density within cover type classes. As a result, there are no detailed maps on the amount and spatial distribution of carbon in the region.

Remote sensing has been extensively used as a basis to infer forest structure and above-ground biomass (AGB) (Dobson 2000, Saatchi *et al* 2007, Houghton *et al* 2007, Baccini *et al* 2004, Blackard *et al* 2008, Zhang and Kondragunta 2006, Zheng *et al* 2004, Lu 2006). Although remotely sensed observations do not directly measure biomass, the radiometry is sensitive to vegetation structure (crown size and tree density), texture and shadow, which are correlated with AGB, particularly in the short wave infrared bands (of which the MODIS sensors have 4). Consequently, remotely sensed spectral reflectance measurements can be useful predictors of biomass (Gemmell 1995, Shugart *et al* 2000, Puhr and Donoghue 2000). Most recently, lidar (light detection and ranging) remote sensing has been used to successfully characterize vegetation vertical structure and height, and to infer AGB (Lefsky *et al* 2005, Drake *et al* 2002).

In this paper, we describe mapping AGB across tropical Africa using MODIS observations and extensive field measurements. The approach leverages a combination of field data that provide accurate information at the plot level, and remote sensing data that are continuous in space over large areas. A mosaic of best-quality MODIS observations provided

cloud-free spectral reflectance data for the entire region. Field measurements were then used to calibrate a regression tree model that estimated AGB for each 1 km<sup>2</sup> pixel as a function of the spectral information derived from MODIS data. The results were cross-validated using a reserved set of field data, as well as independent lidar measurements from the Geoscience Laser Altimeter System (GLAS).

## 2. Data and methods

### 2.1. Study area

The study area encompasses about 20 million km<sup>2</sup> of tropical Africa, covered by 19 MODIS tiles (figure 1). The region is characterized by a diverse range of moist tropical forest, seasonal and semi-arid woodland, savanna, and wetland forests (Laporte *et al* 1998, White 1983).

### 2.2. MODIS data

The MODIS Nadir bidirectional reflectance distribution function adjusted reflectances (NBAR) product (MOD43B4.V4) have 1 km spatial resolution and a composited 16 day temporal resolution. The data have been corrected for solar and view geometry, atmospheric attenuation, and screened for cloud cover (Schaaf *et al* 2002). We used seven bands designed for land applications with wavelength from 459 to 2155 nm and analyzed ten 16-day periods of NBAR data for each year between 2000 and 2003, in the process developing a mosaic of best-quality observations (figure 1). The MODIS NBAR products we used are already cloud screened as part of the production process (Schaaf *et al* 2002). We leveraged this by using 4 years of data (2000–2003) and focusing on filling any remaining gaps associated with clouds by temporally compositing with high-quality screened data.

### 2.3. Lidar GLAS measurements

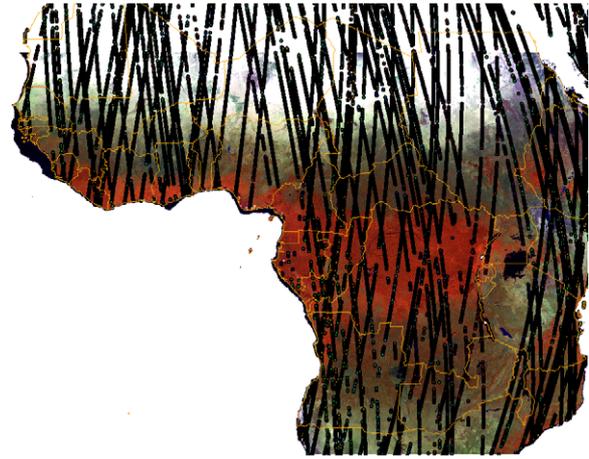
The GLAS instrument on board the Ice, Cloud, and Elevation Satellite (ICESAT) is a waveform sampling lidar sensor originally designed for observation of ice sheets (Zwally *et al* 2002). Lidar metrics have been extensively used to characterize vegetation structure (Sun *et al* 2008, Lefsky *et al* 2005, 1999) and to link structure metrics to biodiversity (Goetz *et al* 2007). Drake *et al* (2003) found a strong relationship between AGB and the height of median energy (HOME), i.e. 50% of the returned energy between the leading and trailing waveform edges. Because these HOME metrics are partly determined by the amount of lidar energy that reaches the ground surface, they are sensitive to both vegetation vertical structure and horizontal canopy density (canopy cover) (Drake *et al* 2002). As a result, they are useful for forest biomass estimation, either at plot locations (with satellite sampling instruments like GLAS) or mapping across spatial domains (using one of several operating aircraft imaging sensors).

In this work we derived the average vegetation height and the HOME for about 1.3 million observations recorded by GLAS Laser 2 (L2A) (figure 2). Observations recorded from Laser 2, acquired between Oct–Nov 2003, are considered to be best-quality data based on transmitted power levels (Sun *et al* 2008), and these data were also closest in time to the MODIS observations used for this study.

### 2.4. Biomass data

Field biomass data sets were derived from forest inventories carried out in Republic of Congo (ROC), Cameroon and Uganda. The forest inventories provided timber volume or biomass information at the plot level, or as averages associated with specific vegetation types. Because of time differences between field data collection and MODIS observations, the MODIS pixels used for the analysis were screened using high resolution orthorectified Landsat GeoCover imagery (Tucker *et al* 2003) to verify that major land cover transitions had not occurred in the interim. Specifically, we used the GeoCover imagery to determine visually if significant degradation or land cover change had occurred between 1990 and 2000. If land cover change or degradation was detected, the field data was not considered for further analysis.

**2.4.1. Republic of Congo.** We analyzed a set of forest inventory measurements collected over the period 2001–2003, covering four forest management units in the northern Republic of Congo. The forest inventory was done by a commercial logging company and was based on a systematic sampling design, with parallel transects of 200 m length separated by 2.5 km. The sampling intensity was 1% for large trees (trunk diameter 40 cm and above), 0.5% for small trees (in the 20–40 cm range) and 0.2% for ‘regenerating trees’ (5–20 cm range) (CIB 2003, Wilks 2003). For large and small trees, all individuals were counted. For regenerating trees only commercial species were counted. Above-ground biomass for each plot was derived as a function of the total number of trees within a range of stem diameter classes, using allometric equations developed for moist tropical forest based on 172

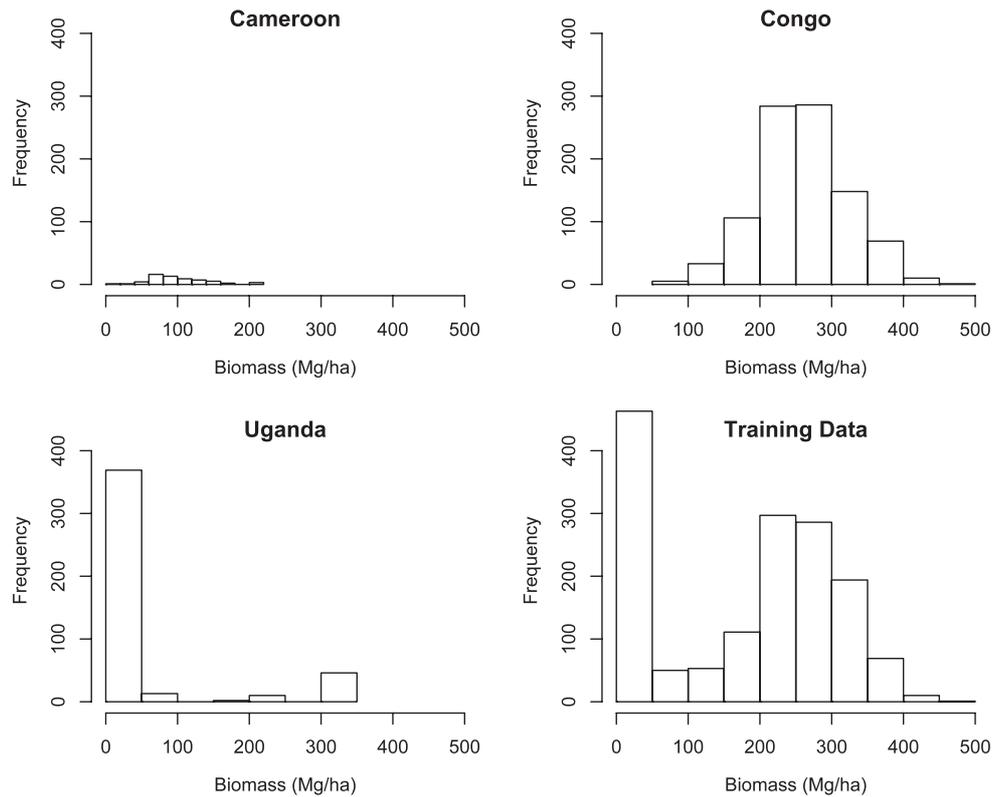


**Figure 2.** The black dots show about 30% of the GLAS L2A (year 2003) shots after screening procedures and used in the comparison analysis with the predicted biomass.

trees with DBH ranging from 5 to 148 cm (Brown *et al* 2005). We recognize that logging companies are primarily interested in timber volume and this can result in an underestimate of total biomass because foliage and branches are not being included. But this effect should be minimized because we used DBH to convert to biomass using Brown *et al* (2005) allometric equations that derive total biomass as a function of DBH.

To minimize the effects of subpixel variability and errors due to mismatches in resolution between field data and satellite observations, we overlaid the field plots on the MODIS imagery and computed average biomass for only those 1 km pixels with at least three field plots. Our assumption was that three or more samples were sufficient to characterize the spatial variability of biomass within the 1 km MODIS pixel. This reflected a compromise between better characterization of the pixels and the need for a sufficient sample size of training data. Setting more stringent criteria resulted in less than a usable number of training data, whereas relaxing it substantially and unacceptably increased the error in the estimates. If there were less than three plot per pixel those data were excluded. Using this approach we identified a total of 942 pixel locations containing more than 3 field plot inventories. Figure 3 shows the frequency distribution of the biomass data aggregated to the 1 km pixel.

**2.4.2. Cameroon.** Forest inventories were collected over an area of dense humid forest extending about 200 km north–south and 700 km east–west (Honzak 1997). As with the ROC data set, Landsat Geocover images were used to screen areas that had experienced forest cover change between the time of field inventories (1994) and the MODIS acquisitions. The sampling design was originally optimized to capture spatial variability within 1.1 km<sup>2</sup> AVHRR Local Area Coverage (LAC) observations, which is comparable in size to the 1 km MODIS products we used. The measurements were converted to biomass using the allometric equations of Brown (1997). A total of 61 sample locations were retained.



**Figure 3.** Frequency distribution of field measurements aggregated to the MODIS resolution for Cameroon, Congo, and Uganda. In the bottom right corner the distribution of training data used is shown.

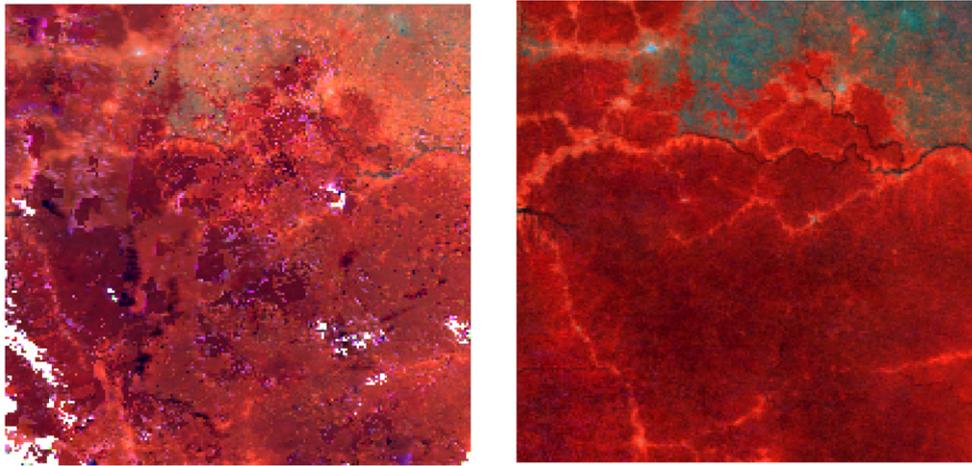
**2.4.3. Uganda.** We derived training data sets from a biomass map produced as part of the Ugandan national biomass inventory. The field measurements were collected between 1995 and 1999, and were associated with a high resolution land cover type map (Drichi 2003). We computed the area weighted average biomass per 1 km<sup>2</sup> pixel. Following screening with the Landsat GeoCover data, we retained 442 sample locations for our analysis.

**2.5. Analysis**

The MODIS product provides quality ranking information for each of the NBAR surface reflectance measurements. We analyzed these quality control flags and selected only reflectance values derived using full model inversion or ‘magnitude inversion’ based on at least 3 observations during the 16 day compositing period (Schaaf *et al* 2002). When more than one observation passed the quality checking, the average surface reflectance was computed. Thus we used an average of all good-quality observations during the ten 16-day periods over the 4 year period. We selected dates that minimized the effects of fire and that enhanced the difference between herbaceous and woody vegetation. We did this by selecting 16 day periods before and after the fire seasons to minimize the effect of burn scars. The best periods to separate the grassy vegetation from the woody vegetation were dates at the beginning of the dry season (early so that the fire season had not yet started) and just before the beginning of the rainy season (when trees start greening before grasses).

Tree-based models have been used in many contexts to predict both categorical (Hansen *et al* 1996, Friedl and Brodley 1997, Saatchi *et al* 2007, Friedl *et al* 2002) and continuous variables (Michaelsen *et al* 1994, Prince and Steininger 1999, Baccini *et al* 2004). The basic theory behind such models is reported in Breiman *et al* (1984). Tree-based models make no assumptions regarding the distributional properties of the input data, are able to capture nonlinear relationships between the response and predictor variables, and provide easily understandable output. For the work reported here, the specific methodology is referred to as a regression tree, because we predict continuous values. Tree-based algorithms perform recursive partitioning of a data set such that each partition results in greater homogeneity relative to the unpartitioned data. The tree is composed of a root node (comprised of all of the data), a set of internal nodes (splits), and a set of terminal nodes (leaves). The splitting procedure stops when the variability within a node is considered sufficiently low (based on the deviance within the node), or when a prescribed minimum number of cases is reached. The mean value of the response variable in each leaf node then serves as the prediction for all cases within that node.

Bootstrap aggregation (bagging), is a method for averaging predictions from a collection of bootstrap samples. The main goal of bagging is to reduce the variance of the predictions. Bagging produces a model for each bootstrap sample, and the final prediction provided by bagging is the average prediction across all models. If small changes in the training set result in different predictions, bagging provides



**Figure 4.** On the left: MODIS 16 day NBAR composite. On the right: MODIS NBAR data resulting from the best-quality pixel mosaic. Artifacts due to cloud cover residuals and shadows are not present in the best pixel mosaic.

a valid tool for tuning and improving the accuracy of the predictions (Breiman 1996). Breiman (2001) proposed a novel extension of tree-based model called Random Forest, in which random feature selection is used in addition to bagging. In Random Forest, a large number of trees are grown with the root node containing a different bootstrap sample of the data with the same number of cases as the original data. At each node, splitting is performed using a randomly selected subset of the predictor variables. To predict unseen cases, Random Forest is provided a new set of predictors, and the final prediction is the average of the values predicted by all the trees. Compared to standard tree-based model, Random Forest is less sensitive to noise in the training data and tends to result in more accurate models.

The set of field biomass training data (figure 3) and the MODIS observations were used to develop the Random Forest model. Biomass predictions for the entire area were then generated by incorporating reflectance measurements from the first seven MODIS spectral bands (the land bands) into the Random Forest model, effectively extending the model based on field training data to the entire region. Because one of the characteristics of tree-based models is the inherent ability to stratify data into homogeneous subsets (in this case different ecological regions) by decreasing the within-class entropy, the model identifies initial strata representative of the broad ecological domains present in the study area, and there is no need to stratify *a priori*.

To assess the accuracy of the predictions, a subset of the field data not used in model development were reserved for a cross-validation analysis (Friedl and Brodley 1997). We used 10% (154 samples) of the field data, which were extracted using a random sampling design (Cochran 1977).

GLAS data were screened on the basis of: (1) the number of peaks in the Gaussian waveform determined by GLAS post processing; (2) the presence of geographic coordinates for the shot; (3) the difference in elevation between Shuttle Radar Topography Mission (SRTM) and GLAS measured surface elevation; (4) the maximum signal never exceeding twice the noise level. We then analyzed lidar metrics relative to the

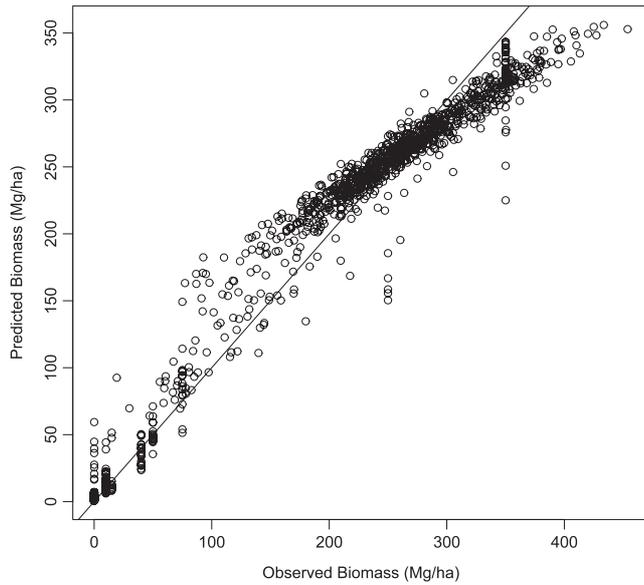
predicted biomass values with the assumption that as biomass increases the lidar height and HOME metrics would also increase (Drake *et al* 2002). Finally, we screened any GLAS shot on terrain exceeding 10% slope, using the SRTM gridded elevation data set.

### 3. Results

#### 3.1. Distribution of biomass density

The temporal compositing of the three years of MODIS NBAR products resulted in a high-quality cloud-free data set for each of the seven MODIS spectral bands (figure 1). Although NBAR products are cloud screened, high thin cirrus and clouds edges can be difficult to detect, thus artifacts may be evident in some regions (figure 4). The temporal compositing approach, long used in AVHRR data products (Holben 1986), provided a useful solution to the problem of optical remote sensing in areas of persistently high cloud cover. This was particularly evident in the coastal areas of tropical west and central Africa. We found the process also effectively reduced the effects of seasonal burning and associated smoke.

The field data used in the study extend across a relatively narrow latitude band (2° N and 6° N) but they cover a very large range of biomass values in a wide range of cover types, ranging from savannas to the dense humid forest (as compared to a Koppen ecological map (Koppen 1936) and the GLC2000 (Mayaux *et al* 2004) land cover map). This assured good representation of the range of African ecosystems, despite the relatively narrow latitude range of the sites, and this observation was supported by the integration of the field measurements and MODIS reflectance in the Random Forest model. More than 96% of the variance in above-ground biomass density was explained, with a root mean square error (RMSE) of 23.5 Mg ha<sup>-1</sup> (figure 5), when the same data set was used for the training and cross-validation. The model explained 82% of the variance in above-ground biomass density, with a RMSE of 50.5 Mg ha<sup>-1</sup> when tested against the 10% of reserved data that was not used for training. The



**Figure 5.** Relationship between observed biomass ( $\text{Mg ha}^{-1}$ ) and Random Forest predictions. A 1:1 line is indicated.

**Table 1.** Regression coefficients, standard errors and fit statistics showing relative sensitivity of the MODIS land reflectance bands.

MODIS bands	Coefficients	Std error	<i>t</i> value	Pr(>  <i>t</i>  )
B1	0.260	0.059	4.434	$1.00 \times 10^{-5}$
B2	-0.132	0.025	-5.233	$1.93 \times 10^{-7}$
B3	-0.362	0.091	-3.988	$7.00 \times 10^{-5}$
B4	-0.411	0.075	-5.440	$6.31 \times 10^{-8}$
B5	0.491	0.039	12.700	$<2 \times 10^{-16}$
B6	-0.705	0.049	-14.451	$<2 \times 10^{-16}$
B7	0.432	0.043	9.920	$<2 \times 10^{-16}$

range of observed (figure 3) and predicted biomass was 0–454  $\text{Mg ha}^{-1}$  and 0–359  $\text{Mg ha}^{-1}$ , respectively.

The utility of using tree-based models compared to more traditional multiple regression analysis is shown in table 1, where we report results from linear regression models using the same set of data. When applied to the same validation data set used for Random Forest, the explained variance is 71% compared to 82% from Random Forest. This suggests substantial improvement in using a non-parametric statistical model such as Random Forest. It is also interesting to note that the short wave infrared band (B6) had the largest coefficient, thus the largest contribution to explained variance in biomass.

Using the Random Forest model, we produced the first spatially continuous biomass density map of tropical Africa using remote sensing observations (figure 6). The map shows the distribution of AGB across Central Africa as well as the spatial variability of AGB. The map indicates that the above-ground biomass in the region varies from 0 to 356  $\text{Mg ha}^{-1}$  at 1 km spatial resolution and that most of the high values biomass are concentrated in the Democratic Republic of Congo.

Table 2 shows how the biomass values were related to land cover type classes, as provided by the GLC 2000 map (Mayaux *et al* 2004). The biomass values, which were

**Table 2.** Average biomass by GLC2000 land cover type.

GLC2000 land cover type	Mean biomass ( $\text{Mg ha}^{-1}$ )
Submontane forest (900–1500 m)	238.1
Swamp forest	251.0
Closed evergreen lowland forest	216.3
Degraded evergreen lowland forest	121.2
Montane forest (>1500 m)	169.6
Mosaic forest/croplands	91.5
Mosaic forest/savanna	77.4
Closed deciduous forest	85.0
Swamp bushland and grassland	32.7
Deciduous woodland	35.2
Open deciduous shrubland	12.7
Deciduous shrubland with sparse trees	11.5
Closed grassland	7.0
Croplands (>50%)	5.3
Open grassland with sparse shrubs	1.0
Sparse grassland	2.3
Open grassland	1.9

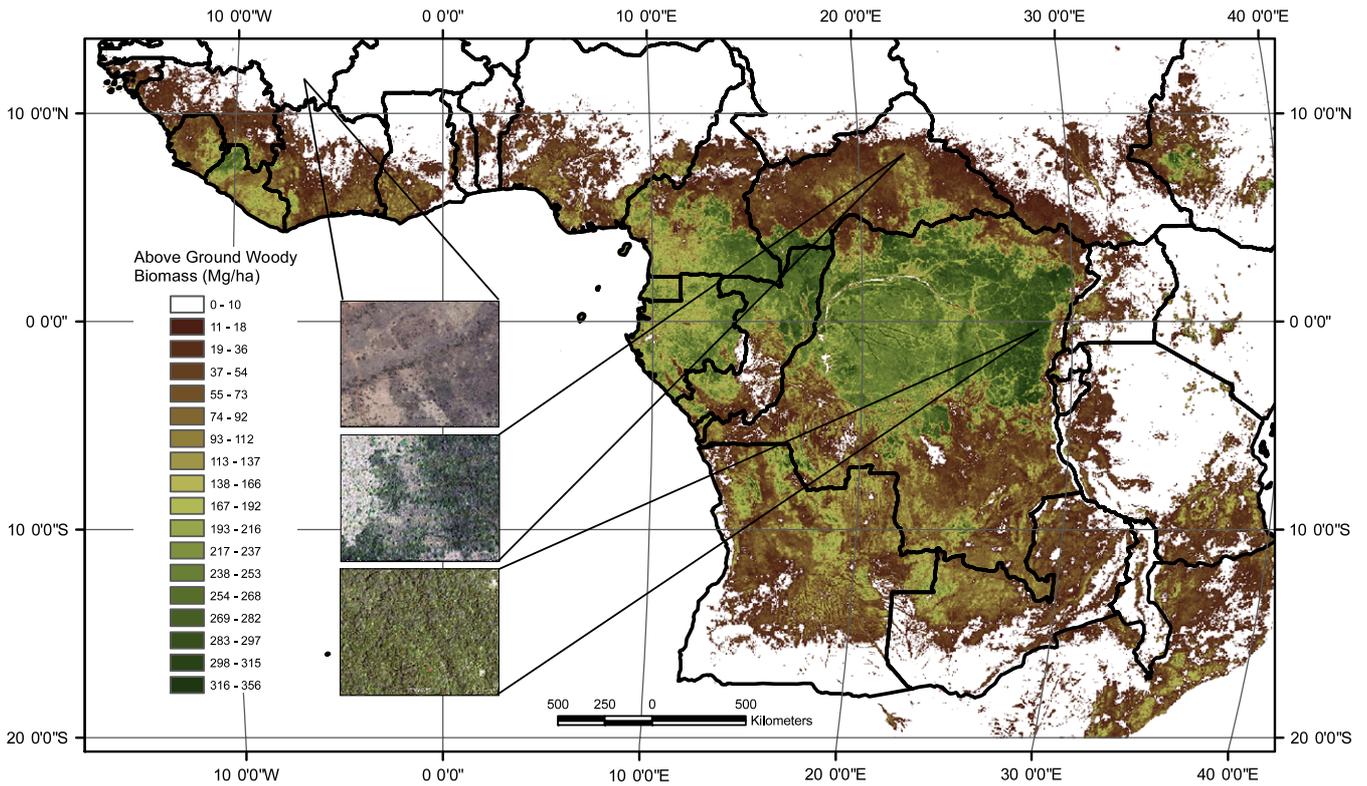
produced without the use of land cover information, partitioned into values that were reasonable and expected in terms of mean values. For example, the high value (238  $\text{Mg ha}^{-1}$ ) was associated with submontane forest and the lowest values (less than 10  $\text{Mg ha}^{-1}$ ) were associated with grassland classes.

### 3.2. Comparison with other data sources

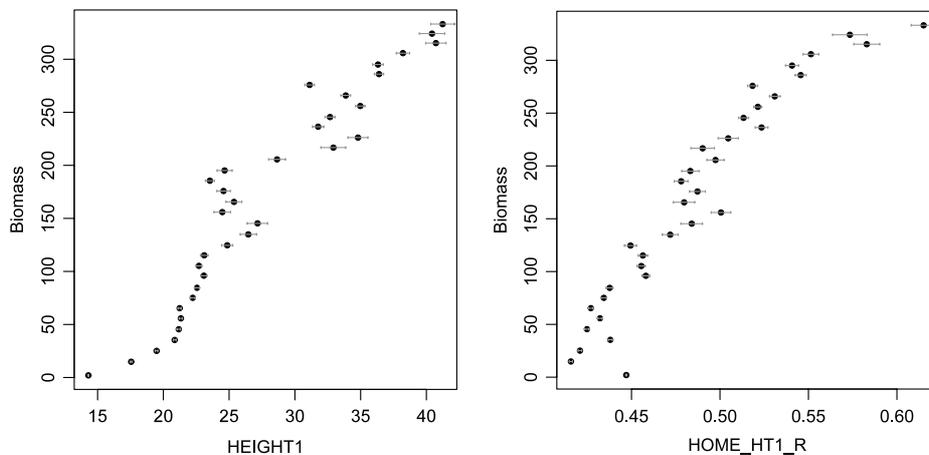
We computed the total amount of above-ground standing biomass for the Democratic Republic of Congo (DRC) to be 34.7 Gt (billion tons). The estimate is consistent with the Food and Agriculture Organization of United Nations (FAO) Forest Resource Assessment (FRA), which reports a value of 37.8 Gt for the year 2000.

Gibbs *et al* (2007) report a total of 20.4 Gt carbon (equivalent to 40.8 Gt of biomass) for the DRC, including the below-ground component. By adding the below-ground component to our estimate using an average ratio for tropical rainforest (0.37) Eggleston *et al* (2006), and converting the biomass into carbon (as 0.5 units C per unit biomass), we arrive at a value of 23.7 Gt C (equivalent to 47.4 Gt of biomass) in the DRC. Using a modified estimate of 0.33 for below-ground allocation (Mokany *et al* 2006) for tropical rainforest we get a total of 23.0 Gt C.

The analysis of the GLAS data (figure 2) showed a strong relationship between MODIS biomass predicted and the GLAS metrics (figure 7). We also found a strong positive relationship between MODIS biomass aggregated in classes of 10  $\text{Mg ha}^{-1}$  with the average vegetation height ( $r^2 = 0.90$ ) and the ratio of HOME and height ( $r^2 = 0.90$ ). Because forest biomass is mainly a function of tree size (DBH and height) and the number of trees per unit area, lidar metrics are useful for biomass estimation and our results are consistent with those of other lidar studies of tropical forest biomass (e.g. Drake *et al* 2003, Lefsky *et al* 2005, Drake *et al* 2002). These comparisons provide strong support for the validity of the approach and associated map.



**Figure 6.** Biomass map derived from MODIS imagery and field data. The three high resolution figures show the vegetation formations with different AGB. Each figure has a length of 1 km.



**Figure 7.** Relationship between GLAS derived height, ratio of HOME and height and predicted biomass ( $\text{Mg ha}^{-1}$ ) aggregated in classes of  $10 \text{ Mg ha}^{-1}$ . The horizontal bars show the standard error for the GLAS metric within biomass bin.

#### 4. Discussion

The results of the biomass mapping demonstrate the utility of satellite data sets, including optical imagery, for estimating above-ground carbon stocks even in persistently cloudy areas of the world. The frequent temporal coverage of MODIS imagery increases the likelihood of capturing cloud-free acquisitions, and the sensitivity of the composited reflectance to canopy density and structure provides the means to link canopy reflectance to above-ground biomass. We note here that we tested combinations of spectral bands and other

MODIS standard products including the NDVI, EVI and LAI, as well as climate data (precipitation, temperature, and evapotranspiration) and topography. The gain in adding these variables was quite limited, and carried with it some negative attributes including the emergence of spatial artifacts in the resulting biomass distribution map. As a result, we decided to use the simplest approach and model based on the 7 MODIS spectral bands designed for land studies.

The comparisons with independent GLAS lidar energy metrics confirm this sensitivity. Moreover, the Random Forest models are powerful for mining relationships in intensively

sampled data sets. There are, however, limitations to the Random Forest model in the prediction phase (figure 4). The model tends to over predict low biomass values and under predict high biomass values. This trend is intrinsic of regression tree-based models whose predictions are the average of the values within the terminal node. Although the model tends to overestimate in the small biomass classes, the biomass map indicates very low biomass in the sub-Saharan region of Mali, Burkina Faso, and Sudan. In this region the vegetation is characterized by sparse trees and low brush that is highly fragmented and dominated by bare soil reflectance.

Furthermore, the model seems to significantly reduce the predicted accuracy when tested on an independent set of data, as the increase in RMSE and decrease in explained variance indicate. These caveats should be kept in mind, as with those of any other technologically-based monitoring approach, in the context of the current political discussions on REDD. Despite efforts to expand field measurement efforts, particularly via the FAO, there are currently limited high-quality field biomass estimates available at sufficient spatial extent to develop and independently validate maps of AGB across tropical regions. Thus there is a need to expand these efforts along with improved field estimates of deforestation and degradation rates.

A limitation of empirical models, including regression tree models that are strongly influenced by the distribution of the training data, is the availability of field measurements representative of the biomass variability of the region. We used the most extensive field biomass data sets we could assemble in a consistent fashion, but additional field data collection could improve the resulting biomass map. Also, a common problem in the use of remotely sensed data in combination with field measurements is the mismatch between the area sampled on the ground and the resolution of the satellite observations. We minimized this effect by specifying a minimum number of field plots within each 1 km<sup>2</sup> MODIS pixel, but areal weighting using high resolution satellite data may permit improved spatial scaling from the plot to the pixel resolution (Baccini et al 2007).

## 5. Conclusion

The new role of Africa in the global economy, particularly the demand for new land for agro-industry, has the potential to significantly increase pressure on existing natural resources. It is therefore critical to have reliable and current information on the spatial distribution of AGB.

We describe methods to map above-ground biomass over tropical Africa using multi-year MODIS satellite observations and a wide range of field measurements. The results indicate that the MODIS data sets, used in a cross-validated regression tree model, captured the amount and spatial distribution of above-ground biomass across tropical Africa. Comparison with GLAS lidar energy height metrics, particularly HOME, showed strong positive correlations with the mapped MODIS biomass density values, and low standard errors across the full range of predicted AGB. This is the first biomass map of Africa based on satellite observations, and it provides not

only important information on carbon stocks but an essential baseline for monitoring and modeling carbon exchange in tropical Africa at relatively high spatial resolution.

Our future work will focus on fusion of lidar observations describing vegetation vertical structure with multi-temporal MODIS and other remotely sensed data products to further improve above-ground biomass and extend the results to additional regions.

## Acknowledgments

This work was funded under NASA contract number NNG05GD14G and NNS06AA06A, the Roger and Victoria Sant, Joseph Gleberman, and The Linden Trust for Conservation. The authors thank the Congolaise Industrielle du Bois (Olivier Desmet and Dominique Paget) in the ROC, the Uganda National Forest Authority, and Miro Honzak for sharing their data sets; and Claudia C Carabajal and Marc Simard for sharing their knowledge about GLAS data.

## References

- Baccini A, Friedl M A, Woodcock C E and Zhu Z 2007 *Photogramm. Eng. Remote Sens.* **73** 945–54
- Baccini A, Friedl M, Woodcock C and Warbington R 2004 *Geophys. Res. Lett.* **31** L10501
- Blackard J et al 2008 *Remote Sens. Environ.* **112** 1658–77
- Breiman L 1996 *Mach. Learn.* **24** 123–40
- Breiman L 2001 *Mach. Learn.* **45** 5–32
- Breiman L, Friedman J H, Olshen R A and Stone C J 1984 *Classification and Regression Trees* (Belmont, CA: Wadsworth)
- Brown S 1997 Estimating biomass and biomass change of tropical forests. A primer *Technical Report 134* Food and Agriculture Organization of United Nations (FAO), Rome
- Brown S, Pearson T, Moore N, Parveen A, Ambagis S and Shoch D 2005 Impact of selective logging on the carbon stock of tropical forests: Republic of Congo as a case study *Technical Report 6 Winrock International* <http://www.winrock.org>
- CIB 2003 Congolaise industrielle des bois; ressources forestieres de l' unite forestiere d' Amenagement de kabo *Rapport d' Inventaire d' Aménagement Technical Report* Ministere de l' Economie Forestiere et de l' Environment and Congolaise Industrielle des Bois (CIB)
- Cochran W G 1977 *Sampling Techniques* (New York: Wiley)
- Dobson M C 2000 *J. For.* **98** (6) 41–3
- Drake B J, Dubayah R O, Knox R G, Clark D B and Blair J B 2002 *Remote Sens. Environ.* **81** 378–392
- Drake B J, Knox R G, Dubayah R O, Clark D B, Condit R, Blair J B and Hofton M 2003 *Glob. Ecol. Biogeogr.* **12** 147–59
- Drichi P 2003 National biomass study 2003 *Technical Report* Forest Department, PO Box 1613, Kampala, Uganda
- Eggleston H S, Buendia L, Miwa K, Ngara T and Tanabe K 2006 Guidelines for national greenhouse gas inventories *Technical Report 4* Intergovernmental Panel on Climate Change (IPCC) IGES, Japan Prepared by the National Greenhouse Gas Inventories Programme
- Friedl M A and Brodley C E 1997 *Remote Sens. Environ.* **61** 399–409
- Friedl M et al 2002 *Remote Sens. Environ.* **83** 214–31
- Gemmell F M 1995 *Remote Sens. Environ.* **51** 291–305
- Gibbs H K, Brown S, Niles J O and Foley J A 2007 *Environ. Res. Lett.* **2** 045023
- Goetz S, Steinberg D, Dubayah R and Blair B 2007 *Remote Sens. Environ.* **108** 254–63
- Gurney K R and Raymond L 2008 *Carbon Balance Manag.* **3** 2

- Hansen M C et al 2008 *Proc. Natl Acad. Sci. USA* **105** 9439–44
- Hansen M, Dubayah R and DeFries R 1996 *Int. J. Remote Sens.* **17** 1075–81
- Holben B N 1986 *Int. J. Remote Sens.* **7** 1417–34
- Honzak M 1997 Mapping carbon pools in the regenerating forests of brazil and cameroon using remote sensing techniques *PhD Thesis* University of Wales, Swansea
- Houghton R 2007 *Annu. Rev. Earth Planet. Sci.* **35** 313–47
- Houghton R A 2005 *Glob. Change Biol.* **11** 945–58
- Houghton R A, Butman D, Bunn A G, Krankina O N, Schlesinger P and Stone T A 2007 *Environ. Res. Lett.* **2** 045032
- Koppen W 1936 *Handbuch der Klimatologie* ed W Koppen and G Geiger (Berlin: Borntraeger)
- Laporte N, Goetz S, Justice C and Heinicke M 1998 *Int. J. Remote Sens.* **19** 3537–50
- Laporte N T, Stabach J A, Grosch R, Lin T S and Goetz S J 2007 *Science* **316** 1451
- Lefsky M A, Harding D, Cohen W, Parker G and Shugart H 1999 *Remote Sens. Environ.* **67** 83–98
- Lefsky M A, Harding D J, Keller M, Cohen W B, Carabajal C C, Espirito-Santo F D B, Hunter M O and de Oliveira R 2005 *Geophys. Res. Lett.* **32** L22S02
- Lu D 2006 *Int. J. Remote Sens.* **27** 1297–328
- Mayaux P, Bartholome E, Fritz S and Belward A 2004 *J. Biogeogr.* **31** 1–17
- Michaelsen J, David S S, Friedl M A, Davis F W and Dubayah R C 1994 *J. Veg. Sci.* **5** 673–86
- Mokany K, Raison J and Prokushkin A 2006 *Glob. Change Biol.* **12** 84–96
- Mollicone D et al 2007 *Clim. Change* **83** 477–93
- Prince S D and Steininger M K 1999 *Glob. Change Biol.* **5** 1–22
- Puhr C B and Donoghue D N M 2000 *Int. J. Remote Sens.* **21** 633–46
- Saatchi S S, Houghton R A, Alvala D S, Soares J V and Yu Y 2007 *Glob. Change Biol.* **13** 816–37
- Schaaf C B et al 2002 *Remote Sens. Environ.* **83** 135–48
- Shugart H H, Chavez L B and Kasischke E S 2000 *For. Sci.* **46** 478–86
- Stephens B B et al 2007 *Science* **316** 1732–5
- Sun G, Ranson K J, Kimes D S, Blair J B and Kovacs K 2008 *Remote Sens. Environ.* **112** 107–17
- Tucker C, Grant D and Dykstra J 2003 *Photogramm. Eng. Remote Sens.* **55B** 378–90
- White F 1983 *The Vegetation of Africa* (LaChaux-de-fonds, Switzerland: UNESCO)
- Wilks C 2003 An outside look at the CIB management inventory in northern Congo *Technical Report* Congolaise Industrielle des Bois (CIB)
- Zhang X and Kondragunta S 2006 *Geophys. Res. Lett.* **33** L09402
- Zheng D, Rademacher J, Chena J, Crowe T, Bresee M, Moined J L and Ryu S 2004 *Remote Sens. Environ.* **93** 402–11
- Zwally H J et al 2002 *J. Geodyn.* **34** 405–45