

# A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin

Matthew C. Hansen<sup>a,\*</sup>, David P. Roy<sup>a</sup>, Erik Lindquist<sup>a</sup>, Bernard Adusei<sup>a</sup>,  
Christopher O. Justice<sup>b</sup>, Alice Altstatt<sup>b</sup>

<sup>a</sup> *Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007, United States*

<sup>b</sup> *Department of Geography, University of Maryland, College Park, United States*

Received 10 April 2007; received in revised form 18 November 2007; accepted 21 November 2007

## Abstract

In this paper we demonstrate a new approach that uses regional/continental MODIS (MODerate Resolution Imaging Spectroradiometer) derived forest cover products to calibrate Landsat data for exhaustive high spatial resolution mapping of forest cover and clearing in the Congo River Basin. The approach employs multi-temporal Landsat acquisitions to account for cloud cover, a primary limiting factor in humid tropical forest mapping. A Basin-wide MODIS 250 m Vegetation Continuous Field (VCF) percent tree cover product is used as a regionally consistent reference data set to train Landsat imagery. The approach is automated and greatly shortens mapping time. Results for approximately one third of the Congo Basin are shown. Derived high spatial resolution forest change estimates indicate that less than 1% of the forests were cleared from 1990 to 2000. However, forest clearing is spatially pervasive and fragmented in the landscapes studied to date, with implications for sustaining the region's biodiversity. The forest cover and change data are being used by the Central African Regional Program for the Environment (CARPE) program to study deforestation and biodiversity loss in the Congo Basin forest zone. Data from this study are available at <http://carpe.umd.edu>. © 2007 Elsevier Inc. All rights reserved.

*Keywords:* Forest cover; Change detection; Deforestation; High spatial resolution; Monitoring; MODIS; Landsat

## 1. Introduction

### 1.1. Regional-scale mapping of humid tropical forests

Operational landscape characterization and monitoring of the humid tropics is important to studies concerning habitat and biodiversity, management of forest resources, human livelihoods and biogeochemical and climatic cycles (Curran and Trigg, 2006; Avissar and Werth, 2005; FAO, 2005; SCBD, 2001; IGBP, 1998; LaPorte et al., 1998). Studies quantifying humid tropical deforestation over large areas using time-series high spatial resolution satellite data sets have been prototyped (Skole and Tucker, 1993; Townshend et al., 1995). However, operational implementation of such methods for long-term monitoring are only now being operationalized (INPE, 2002;

Asner et al., 2005). The primary limitations to large area high spatial resolution monitoring include the development of generic and robust methods, overcoming data quality issues, and having the resources to purchase required data sets. Concerning methods, many land cover mapping activities rely on photo-interpretation, or other approaches that are labor-intensive, costly, and difficult to replicate in the consistent manner required for long-term monitoring. The primary data limitation for humid tropical forest monitoring is persistent cloud cover that confounds efforts to operationalize land cover and change characterizations (Asner, 2001; Helmer and Rufenacht, 2005; Ju and Roy, in press). Regarding the high cost of high spatial resolution data sets, researchers often use the data they can afford, not the data they truly need. For a region like the humid tropics, data needs are intensive in order to overcome the presence of cloud cover.

This paper presents an approach to address these limitations by employing a multi-resolution methodology for mapping forest

\* Corresponding author. Tel.: +1 605 688 6848.

E-mail address: [Matthew.Hansen@sdstate.edu](mailto:Matthew.Hansen@sdstate.edu) (M.C. Hansen).

cover and deforestation within the humid tropical forests of the Congo River Basin. The MODIS Vegetation Continuous Fields (VCF) algorithm (Hansen et al., 2003) is used to create a regional MODIS 250 m forest/non-forest cover map which is in turn used to drive high spatial resolution Landsat forest characterizations. The automated use of the MODIS forest characterization to pre-process (normalize) and label Landsat data inputs in generating regional-scale forest cover and change maps for the Congo Basin is demonstrated. Data cost limitations cannot be resolved algorithmically, although the method is developed so that additional imagery can be ingested and new products automatically derived upon acquisition.

The research strategy engaged herein focuses on the operationalization of large area forest cover and forest cover change monitoring. Information on where and how fast forest change is taking place can be integrated with other geospatial data on the types and causes of change to better inform resource managers and earth system modelers. The ability to accurately assess forest cover dynamics in a timely fashion will contribute to new applications, such as the Reducing Emissions from Deforestation and Degradation (REDD) initiative (UNFCCC, 2005). Synoptic measurements of change can quantify the displacement of deforestation activities within and between countries and lead to the harmonization of national-scale statistics. To achieve this end, automated or semi-automated procedures that work at regional scales will be needed; such methods must be accurate, internally consistent, produced in a timely fashion and rely on remotely sensed inputs. Spatially explicit forest cover and forest change maps derived from remotely sensed data will be integral to the future monitoring of forests in support of both basic earth science research and policy formulation and implementation.

## 1.2. Congo Basin forest monitoring

The CARPE program (Central African Regional Program for the Environment) is a long-term initiative by USAID to address the issues of forest management, human livelihoods, and biodiversity loss in the Congo Basin forest zone (<http://carpe.umd.edu>). CARPE works within the framework of the Congo Basin Forest Partnership (CBFP, 2005, 2006), an international association of government and non-government organizations with the goal of increasing communication and coordination between in-region projects and policies to improve the sustainable management of the Congo Basin forests and the standard of living of the region's inhabitants. The methods presented here are a contribution to these efforts by advancing the creation of internally consistent, rapid assessments of the forested landscapes of the Congo Basin.

Unlike the forests of the Amazon Basin and Insular Southeast Asia, Central Africa does not exhibit large-scale agro-industrial clearing. As such, MODIS data offer little value in a monitoring sense as change events occur typically at a finer scale than is detectable with 250 m MODIS data. Deforestation in the Congo Basin occurs at fine scales and is caused largely by shifting agricultural activities (CBFP, 2005) that are correlated with local populations (Zhang et al., 2005). Commercial logging is also present, but is highly selective and typically only detectable via the extension of new logging road networks into the forest domain (LaPorte et al., 2007). Current estimates of tropical forest change from the latest UNFAO Forest Resource Assessment (FAO, 2005) indicate Africa as having annual rates of deforestation in excess of 4 million hectares per year. However, past satellite-based surveys indicate much lower rates of

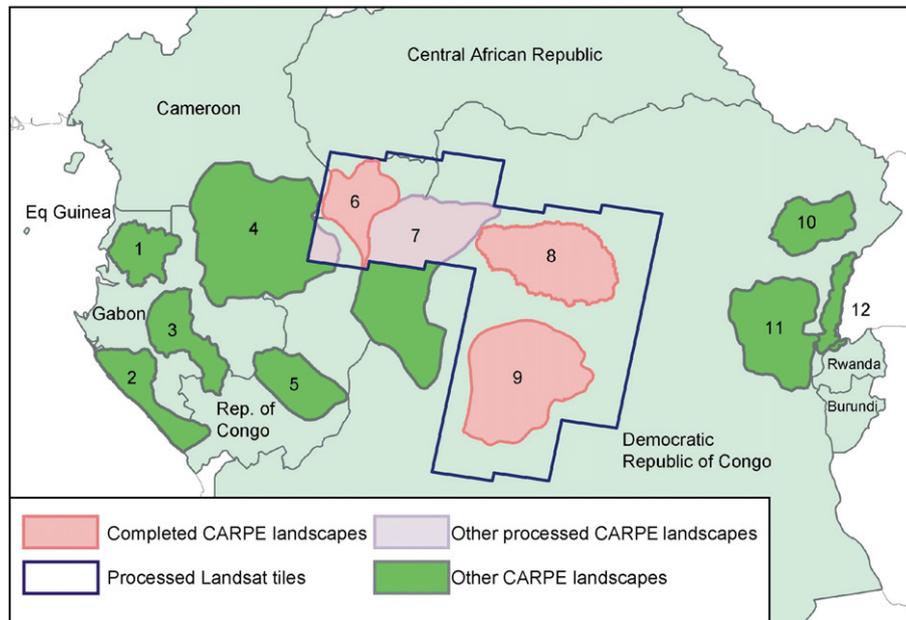


Fig. 1. Overview of CARPE landscapes with completed forest cover and forest change mapping areas highlighted in red. The blue outline delineates processed Landsat data to date, lavender other portions of processed CARPE landscapes, and green CARPE landscapes currently being analyzed. CARPE landscape boundaries are as follows: 1) Monte Alen–Mont de Cristal Inselbergs Forest, 2) Gamba–Mayumba–Konkouati, 3) Lopé–Chaillu–Louesse Forest, 4) Dja–Minkébé–Odzala Tri-National, 5) Leconib–Batéké–Léfini, 6) Sangha Tri-National Forest, 7) Lac Télé–Lac Tumba Swamp Forest, 8) Maringa–Lopori–Wamba, 9) Salonga–Lukenie–Sankuru, 10) Ituri–Epulu–Aru, 11) Maiko–Lutunguru Tayna–Kahuzi–Biega Forest, and 12) Virungas.

change for tropical Africa (Achard et al., 2002; Hansen and DeFries, 2004). Improved high spatial resolution monitoring is required to better determine the rates and spatial extents of forest cover change within the humid tropical forests of Central Africa.

## 2. Study area

The tropical forest ecosystems of the Congo Basin represent the largest and most diverse forest massif on the African continent and the second largest extent of tropical rain forest in the world, next to the Amazon Basin (Wilkie et al., 2001; CBFP, 2005). As defined by the Congo Basin Forest Partnership, the Congo Basin forests cover an area of nearly 2 million km<sup>2</sup>. The Congo Basin in this context is not defined strictly by the drainage area of the Congo River, but by the forest zone extending from the Atlantic Ocean in the west to the Albertine Rift Valley in the east, and spanning the equator by nearly 7° north and south (CBFP, 2005).

The Congo Basin forests, also known as the Lower Guineo–Congolian forests as defined by White (1983), consist predominately of humid evergreen broadleaf forests with seasonality increasing with latitude. Seasonality and corresponding cloud cover are related to the movement of the inter-tropical convergence zone. Cloud cover is typically high nearer the equator, and is persistent in the western Basin due to the warming and rising of moisture-laden air as it moves from the Gulf of Guinea onto the central African land mass. A recent global study of the availability of cloud-free MODIS data for compositing indicated that equatorial Africa was one of several regions affected by high cloud cover at the time of MODIS overpass (Roy et al., 2006).

The CBFP and CARPE have identified 12 priority landscapes for monitoring biodiversity, deforestation and other measures of disturbance within the remaining intact forest zones of the Congo Basin. The methodology presented here is being employed exhaustively across the basin, where high spatial resolution Landsat data are available, to determine rates of change. This paper reports results from three of the landscapes covering approximately one third of the basin that are broadly representative of those in the Congo Basin–Maringa–Lopori–Wamba, Salonga–Lukenie–Sankuru, and Sangha Tri-National (Fig. 1).

## 3. Data

Landsat data are available in 185 km×170 km scenes defined in a Worldwide Reference System of path (groundtrack parallel) and row (latitude parallel) coordinates (Arvidson et al., 2001). In this study we use as many viable Landsat data sets as possible. Ninety-eight Landsat acquisitions, including Landsat 4, 5 and 7 data sensed from 1984 to 2003 (just prior to the Landsat ETM+ scan line corrector failure) and covering 20 scenes (unique path/row) over the Maringa–Lopori–Wamba, Salonga–Lukenie–Sankuru, and Sangha Tri-National landscapes were obtained (Table 1). The reflective Landsat bands, 4 (0.78–0.90 μm), 5 (1.55–1.75 μm), 7 (2.09–2.35 μm), and the thermal band 6

Table 1  
Landsat data used in this study

Maringa–Lopori–Wamba landscape		
179059	178059	177059
January 21, 1987	September 08, 1986	December 22, 1986
December 10, 1994	January 14, 1990	December 25, 1990
August 18, 2002	January 20, 1995	October 20, 2001
March 05, 2000	January 28, 2001	December 26, 2002
	February 03, 2003	
	April 08, 2003	
179060		
January 05, 1987	178060	177060
January 21, 1987	January 14, 1990	August 26, 1984
October 07, 1994	January 20, 1995	December 22, 1986
March 05, 2000	December 27, 2000	January 04, 1986
December 24, 2002	April 08, 2003	October 20, 2001
		December 26, 2002
		February 25, 2002
Sangha Tri-National landscape		
182058	181058	180058
November 26, 1990	January 16, 1986	December 11, 1986
February 12, 1999	December 24, 1994	November 15, 1994
May 16, 2001	March 3, 2000	December 31, 2002
April 1, 2002	January 7, 2003	February 1, 2003
December 29, 2002		
February 15, 2003		
182059		
December 9, 1986	181059	180059
December 28, 1990	September 7, 1984	July 20, 1986
February 12, 1999	October 9, 1984	November 15, 1994
November 8, 2001	December 2, 1986	November 23, 2000
April 1, 2002	November 12, 1999	February 1, 2003
February 15, 2003	February 18, 2001	
	January 7, 2003	
Salonga–Lukenie–Sankuru landscape		
179061	178061	177061
September 9, 1984	January 14, 1990	January 4, 1986
February 6, 1987	January 20, 1995	December 12, 1994
October 7, 1994	February 27, 2000	February 25, 2002
May 3, 2000	December 27, 2000	August 20, 2002
May 3, 2002	April 8, 2003	
179062		
September 9, 1984	178062	177062
February 19, 1986	October 20, 1984	August 26, 1984
February 12, 1995	January 14, 1990	January 10, 1991
May 14, 2002	January 20, 1995	April 8, 2000
	February 27, 2000	February 6, 2001
	April 15, 2000	June 14, 2001
179063		
May 26, 1986	178063	
February 12, 1995	September 24, 1992	
October 15, 2000	December 3, 1994	
June 28, 2001	June 24, 2002	
August 15, 2001	April 8, 2003	
May 14, 2002		
May 17, 2003		

(10.4–12.5 μm) were used. The three shortest visible wavelength Landsat bands were not used due to their sensitivity to atmospheric contamination (Ouaidrari and Vermote, 1999).

The daily L2G 250 m and the 500 m MODIS land surface reflectance products (Vermote et al., 2002) and the 8-day L3 1 km MODIS land surface temperature product (Wan et al.,

2002) were used in this study. These MODIS land products are defined in the sinusoidal map projection in  $10 \times 10^\circ$  land tiles (Wolfe et al., 1998), with eight tiles covering the Congo Basin. All of the MODIS Collection 4 products available from 2000 to 2003 over these 8 tiles were used. The seven MODIS land surface reflectance bands were used: the two 250 m red (0.620–0.670  $\mu\text{m}$ ) and near-infrared (0.841–0.876  $\mu\text{m}$ ) bands, and the five 500 m bands: blue (0.459–0.479  $\mu\text{m}$ ), green (0.545–0.565  $\mu\text{m}$ ), mid-infrared (1.230–1.250  $\mu\text{m}$ ), mid-infrared (1.628–1.654  $\mu\text{m}$ ), and mid-infrared (2.105–2.155  $\mu\text{m}$ ).

#### 4. Methods

The methodology is illustrated in Fig. 2. First, a regional MODIS 250 m forest non-forest map is generated using a regression tree approach (Fig. 2 (1)), and after resampling to a common coordinate system (Fig. 2 (2)), is used to perform radiometric normalization and to reduce deleterious Landsat atmospheric and sun-surface-sensor spectral variations (Fig. 2 (3)). Each normalized Landsat image has standardized classification tree models applied to detect per-pixel clouds and shadows, (Fig. 2 (4)). Landsat-scale forest cover is then mapped using a classification tree approach with training labels derived from the MODIS 250 m forest map (Fig. 2 (5)). The Landsat forest estimates and associated spectral bands and cloud and shadow flags are composited to create decadal, pre-1996 and post-1996, composite forest cover products (Fig. 2 (6)). A deforestation classification tree model is applied to these composites to produce a per-pixel forest change assessment (Fig. 2 (7)). Before describing the methodology in detail, an overview of regression trees and their application for satellite classification is first described.

##### 4.1. Classification and regression trees

The algorithmic tool used to characterize forest cover and change, as well as to evaluate the presence of cloud and shadow artifacts, is the decision tree. Decision trees are hierarchical classifiers that predict class membership by recursively partitioning a data set into more homogeneous subsets, referred to as nodes (Breiman et al., 1984). This splitting procedure is followed until a perfect tree is created, if possible, composed only of pure terminal nodes where every pixel is discriminated from pixels of other classes, or until preset conditions are met for terminating the tree's growth. Trees can accept either categorical data in performing classifications (classification trees) or continuous data in performing sub-pixel percent cover estimations (regression trees). For classification trees, a deviance measure is used to split data into nodes that are more homogeneous with respect to class membership than the parent node. For regression trees, a sum of squares criterion is used to split the data into successively less varying subsets.

Tree-based algorithms offer several advantages over other characterization methods and have been used with remotely sensed data sets (Michaelson et al., 1994; Hansen et al., 1996; DeFries et al., 1997; Friedl and Brodley, 1997). They are distribution-free, allowing for the improved representation of training data within multi-spectral space. In addition, the tree structure enables interpretation of the explanatory nature of the

independent variables. A number of software packages are available, in this study we use the Splus package (Clark and Pergibon, 1992).

Multiple independent runs of decision trees via sampling with replacement allow for more reliable results. This procedure is called bagging (Breiman, 1996), and typically employs a per-pixel voting procedure based on  $n$  derived classification trees to label eventual outputs. Per node likelihoods, and not per node class labels, may also be used to derive mean class membership likelihood values for each pixel. By repeatedly sampling the training data to grow multiple tree models, isolated overfitting within any individual tree is reduced by calculating an averaged multi-tree output. Unless otherwise stated, the analysis presented here employs bagging procedures to derive thematic outputs.

##### 4.2. Generate regional MODIS 250 m forest non-forest cover map

The first step of the methodology is to generate a regional forest non-forest map at moderate spatial resolution. The MODIS Vegetation Continuous Field (VCF) method (Hansen et al., 2003) is used, modified for application to the Congo Basin, to create a MODIS 250 m percent tree cover map. The percent tree cover map is then thresholded into forest and non-forest classes.

Monthly composites were generated from four years of MODIS data (2000–2003), employing the same approach used to generate composites for the 250 m MODIS Vegetation Cover Change product (Zhan et al., 2002). In this compositing approach, the two MODIS 250 m bands are composited based on the MODIS land surface reflectance quality assessment flags and an observation coverage criterion to select the highest quality, nearest-nadir 250 m observation for each month (Carroll et al., in review). The five 500 m surface reflectance band values are retained by selecting the 500 m observation lying closest to each composited 250 m pixel for the selected day. Normalized difference vegetation index (NDVI) values are computed for each pixel from the 250 m red and near-infrared composited values. Monthly 1 km land surface temperature values are selected as those with the maximum land surface temperature value over the month (Cihlar, 1994; Roy, 1997) and resampled to 250 m pixels.

Thirty-four MODIS image metrics, defined following the approach of Hansen et al. (2002a), Hansen and DeFries (2004), were extracted from the four years of monthly composites. The metrics are summarized in Table 2, and three examples are illustrated in Fig. 3a. The metrics illustrated in Fig. 3a are the mean of the three lowest MODIS land surface reflectance red (620–670 nm), near-infrared (841–876 nm), and middle-infrared (1628–1654 nm) monthly composite values. These three metrics provide a time-integrated multi-year data set with minimal cloud contamination, and correspond largely with local growing season conditions.

In creating this Basin-wide forest cover map, we employed the standard MODIS VCF algorithm (Hansen et al., 2002a) using over 2 million training pixels derived from 15 Landsat images. A single perfectly fit regression tree was created using 50% of the training data, and pruned using the remaining set-aside training data. The training data were the dependent



Table 2  
MODIS 250 m metrics derived from monthly composite imagery for 2000–2003

---

Mean of 3 darkest band 1 reflectance monthly composites
Mean of 3 darkest band 2 reflectance monthly composites
Mean of 3 darkest band 3 reflectance monthly composites
Mean of 3 darkest band 4 reflectance monthly composites
Mean of 3 darkest band 5 reflectance monthly composites
Mean of 3 darkest band 6 reflectance monthly composites
Mean of 3 darkest band 7 reflectance monthly composites
Mean of 3 greenest NDVI monthly composites
Mean of 3 warmest LST monthly composites
Mean band 1 reflectance of 3 warmest monthly composites
Mean band 2 reflectance of 3 warmest monthly composites
Mean band 3 reflectance of 3 warmest monthly composites
Mean band 4 reflectance of 3 warmest monthly composites
Mean band 5 reflectance of 3 warmest monthly composites
Mean band 6 reflectance of 3 warmest monthly composites
Mean band 7 reflectance of 3 warmest monthly composites
Mean NDVI of 3 warmest monthly composites
Mean band 1 reflectance of 3 greenest monthly composites
Mean band 2 reflectance of 3 greenest monthly composites
Mean band 3 reflectance of 3 greenest monthly composites
Mean band 4 reflectance of 3 greenest monthly composites
Mean band 5 reflectance of 3 greenest monthly composites
Mean band 6 reflectance of 3 greenest monthly composites
Mean band 7 reflectance of 3 greenest monthly composites
Mean LST of 3 greenest monthly composites
Ranked band 1 reflectances from single monthly composite
Ranked band 2 reflectances from single monthly composite
Ranked band 3 reflectances from single monthly composite
Ranked band 4 reflectances from single monthly composite
Ranked band 5 reflectances from single monthly composite
Ranked band 6 reflectances from single monthly composite
Ranked band 7 reflectances from single monthly composite
Ranked NDVI from single monthly composite
Ranked LST from single monthly composite

---

See text for MODIS band descriptions.

NDVI signifies Normalized Difference Vegetation Index, and LST is Land Surface Temperature.

variable and the 4-year metrics (Table 2) were the independent variables. Importantly, these training data were not used in the Landsat classification described later in this paper.

The resulting 250 m percent tree cover map was then thresholded into forest and non-forest classes. Pixels with percent tree cover greater than or equal to 60% were labeled as forest and those with less estimated tree cover were labeled as non-forest. This definition conforms with several physiognomic classification schemes including the International Geosphere Biosphere Programme land cover classification scheme (Rasool, 1992). In addition, two other categorical land cover classes, water and rural complex, were characterized by independently running a classification tree on the MODIS metrics (Table 2). Water training data were derived using the 15 Congo VCF classified Landsat images. Water pixels were treated as non-land and not used in the subsequent analysis. The rural complex class is a mosaic of tree regrowth, settlement, cropland and plantation (Mayaux et al., 1999) that includes significant areal tree cover but of an immature form. This class is essentially a disturbance category that is not reliably detected by the global MODIS VCF algorithm. Rural complex class training data were derived from the same 15 Congo VCF

classified Landsat images via photo-interpretation. The resulting 250 m rural complex classified MODIS pixels were labeled as non-forest. Again, the water and rural complex training information were not used in the Landsat classification described later in this paper.

#### 4.3. Georectification/resampling of satellite data

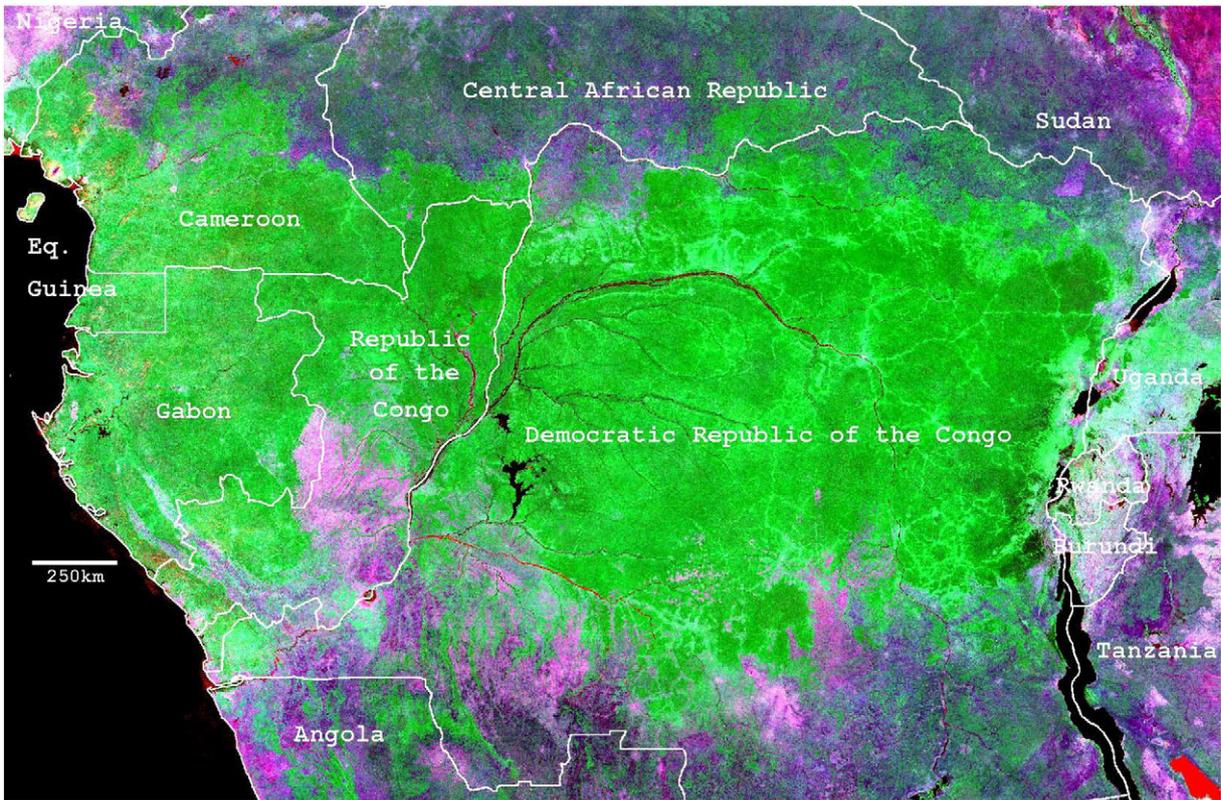
The MODIS 250 m forest non-forest map and the Landsat data (Table 1) were geometrically corrected into registration with the Landsat GeoCover data set. The GeoCover data set is defined in the UTM coordinate system with a geolocation accuracy of 50 m (Tucker et al., 2004). The Landsat data were georeferenced using an automated ground control point matching algorithm (Kennedy and Cohen, 2003) and by bilinear resampling. The MODIS 250 m forest map was reprojected from the MODIS sinusoidal projection to the Geocover projection by nearest neighbor resampling. To reduce spurious change detection due to residual misregistration effects (Townshend et al., 1992; Roy, 2000), the Landsat data were resampled to a spatial resolution of 57 m.

#### 4.4. Landsat normalization

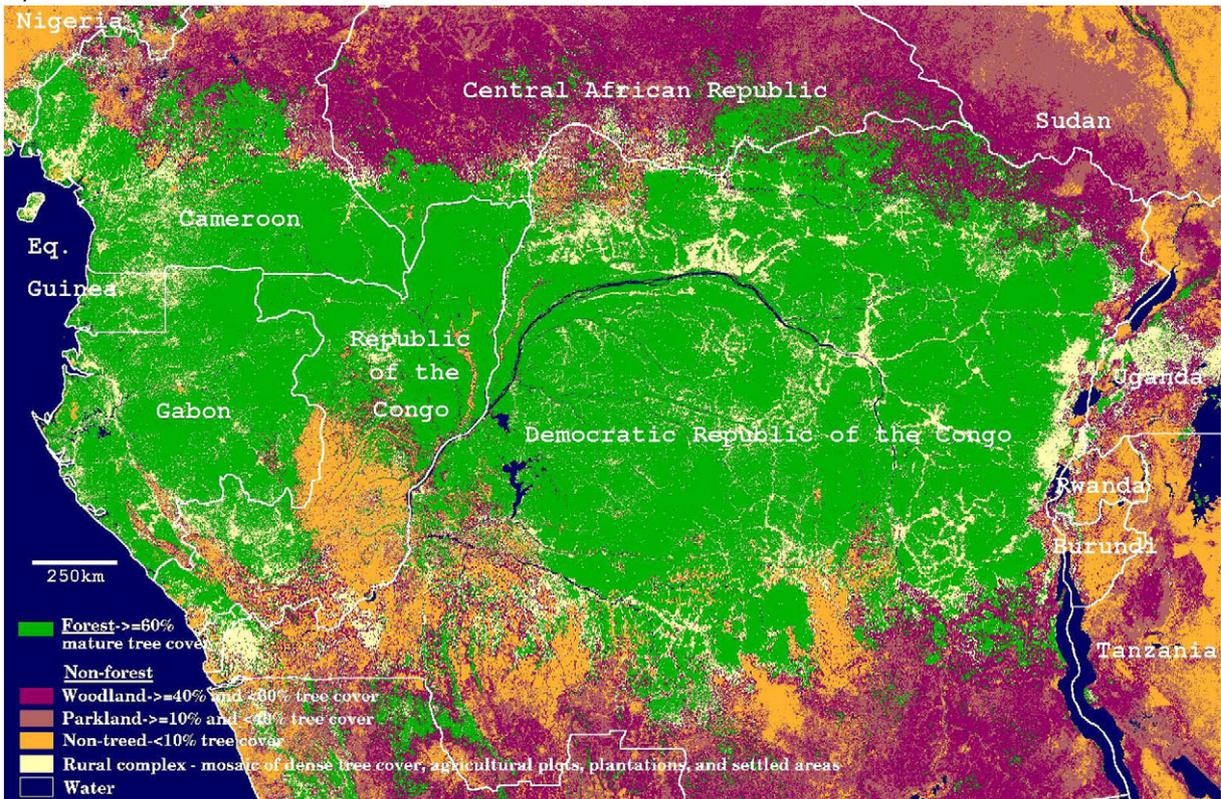
The radiometric consistency of Landsat data, such as used in this study (Table 1), may change due to sensor calibration changes, differences in illumination and observation angles, variation in atmospheric effects, and phenological variations (Coppin et al., 2004). Normalization of the Landsat data to remove or reduce the impacts of these effects is required prior to application of generic models for flagging the non-systematic presence of undesired cloud and shadow effects across all Landsat data, and to subsequently apply a regional deforestation mapping model.

Several methods of radiometric normalization have been proposed, with the dark-object subtraction (DOS) method widely used due to its methodological simplicity (Chavez, 1996). Since the approach focuses on the infrared Landsat bands 4, 5 and 7, more advanced methods that include haze corrections for the visible bands such as that of Carlotto (1999), are not employed. We implement a simple but robust normalization methodology that uses the 250 m MODIS forest map. MODIS forest/non-forest boundaries are spatially eroded by two 250 m pixels using the morphological erode operator (Serra, 1982) to identify interior forest pixels. These intact forest pixels are considered as dark-objects for DOS normalization and to model and remove linear variations occurring across each Landsat scene.

In the DOS normalization, uncontaminated land Landsat pixels are first identified as those with Landsat band 6 thermal brightness temperature values greater than or equal to 19 °C. This threshold was arrived at empirically by testing various candidate values. The flagged pixels are intersected with the unambiguous 250 m MODIS intact forest pixels. From this combined mask, a median forest value for each reflective band is computed. In this way, the forested lands are treated as a dark-object, and all the reflective bands are rescaled to the same reference forest value, defined for convenience as digital number value 100.



a)



b)

Fig. 3. a) Three example MODIS metrics derived over the study area from 4 years (2000 to 2003) of monthly composites. Red, blue and green are, the mean of the three lowest MODIS land surface reflectance red (620–670 nm), near-infrared (841–876 nm), and middle-infrared (1628–1654 nm) band monthly composited values respectively. Thirty-four such metrics were used to generate the Basin-wide forest cover map (Table 2). b) MODIS 250 m land cover for forest density classes made from VCF Landsat training data and 4 years of MODIS inputs. Rural complex and water classes were derived separately and superimposed on the VCF strata.

Remotely sensed variations may occur across the Landsat scene due to atmospheric scattering and surface anisotropy combined with variations in the viewing and solar geometry. The Landsat sensor, with its comparatively narrow field of view, is not as affected by surface anisotropic effects as wider field of view sensors, such as MODIS (Schaaf et al., 2002). However, systematic remotely sensed variations across the Landsat scene are sometimes evident and attempts to remove them using a variety of techniques have been implemented (Danaher et al., 2001; Toivonen et al., 2006). In our Landsat analysis, and in other Landsat research (Danaher et al., 2001; Toivonen et al., 2006), these variations appear greatest across scan rather than along track. For Landsat data, sensor view zenith angle, or scan angle, is the portion of the sun-sensor-target geometry that varies most across the scene. A simple linear regression relationship between the Landsat spectral response and the Landsat cross-track pixel location is estimated for each reflective band as:

$$y = b_0 + b_1 * x \quad (1)$$

where  $y$  equals the DOS normalized Landsat digital number for a given reflective wavelength band,  $x$  is the cross-track (column) pixel value,  $b_1$  is the slope of the linear regression function in digital numbers per cross-track pixel, and  $b_0$  is the intercept. Rather than compute this relationship over all the Landsat pixels, which may have different surface anisotropy and different atmospheric contamination characteristics, the relationship is computed only for Landsat pixels falling under the 250 m MODIS unambiguous intact forest pixels. All the pixels of the DOS normalized Landsat data are then adjusted using this relationship. This process is repeated independently for each reflective band.

#### 4.5. Landsat cloud, shadow and water flagging

Automated methods for flagging cloud and shadow effects are a requirement for large-volume Landsat processing (Helmer and Rufenacht, 2005). Landsat cloud fraction metadata are not spatially explicit and the cloud detection algorithm was not designed to generate per-pixel cloud masks (Irish et al., 2006). Consequently, in this study, a regional cloud and shadow masking classification tree was developed to classify clouds and shadows into low, medium and high-confidence categories. Water bodies must also be identified as they can be confused spectrally with dense dark vegetation and shadows.

Training data were developed from 9 Landsat images to differentiate cloud, shadow and land pixels. Landsat bands 4, 5, 6 and 7 and all combinations of possible 2 band simple ratios were used as inputs to two classification trees classifying these data independently into cloud and shadow classes. The trees were applied to each normalized Landsat scene and the class membership likelihood values used to define per Landsat pixel a low, medium or high cloud/shadow quality assessment state. Repeated shadow flags for a given pixel were used with topographical data (Rabus et al., 2003) to flag water. To reduce the impact of edge effects, a one-pixel (57 m) buffer around the high-confidence

cloud and shadow pixels was created using the morphological dilate operator (Serra, 1982) and made into an additional quality assessment state. In total, four cloud/shadow quality assessment states were defined for each geometrically corrected and normalized Landsat scene pixel: 1) high presence cloud/shadow/water, 2) buffered high presence cloud/shadow/water, 3) medium presence cloud/shadow, and 4) low presence cloud/shadow.

#### 4.6. Landsat forest mapping

A classification tree approach was used to estimate per-pixel forest likelihoods for each geometrically corrected and normalized Landsat acquisition (Table 1). Tree models were generated independently for each Landsat acquisition with the MODIS forest non-forest class labels as the dependent variable and the Landsat data as the independent variable. Training data were defined automatically by sampling from the interiors of the MODIS forest and non-forest mapped areas, derived by eroding the MODIS forest and non-forest classes by the equivalent of two 250 m pixels using the morphological erode operator (Serra, 1982). These training data were assumed to be applicable to the older (pre-2000) Landsat data as rates of forest change in Central Africa are sufficiently low that change is not readily reflected in the interiors of MODIS forest and non-forest mapped areas. Landsat bands 4, 5 and 7, and simple ratios of these bands were used as inputs. In addition, per-pixel local variances and local means for a 3 by 3 kernel were added to the input variable data set.

The per node likelihoods, not class labels, were retained from 30 independently generated decision tree runs. All trees were perfectly fit and run independently on a sample of approximately 30,000 training pixels per Landsat acquisition to generate an average forest likelihood per pixel. This has the effect of generalizing the relationship between the MODIS labels and the Landsat spectral measures, overcoming the frequent, yet minority occurrence of mislabeling. The forest and non-forest training categories were proportionately sampled according to their relative presence in the corresponding MODIS map in order to reduce training bias.

#### 4.7. Landsat compositing

Compositing is a practical way to reduce residual cloud contamination, fill missing values, and reduce the data volume of moderate resolution near-daily coverage sensor data such as AVHRR or MODIS (Holben, 1986; Cihlar, 1994; Roy, 1997). Compositing of higher spatial but lower temporal resolution satellite data, such as Landsat, is not normally undertaken however because of high data costs and because the land surface state may change in the period required to sense several acquisitions. In this study, the Landsat data were composited into two periods, pre-1996 and post-1996. The year 1996 was selected because for all but one scene (path 182 row 058, Table 1) there were at least two pre-1996 and two post-1996 Landsat acquisitions available.

A per-pixel Landsat compositing scheme based on selecting the date with the lowest cloud and shadow likelihood values was applied to the Landsat acquisitions (Table 1). When more than one acquisition date had the same cloud and/or shadow likelihood, the date that had a normalized digital number value

Table 3

Land cover derived from the MODIS 250 m Congo Basin Forest Cover Map (Fig. 3) by country (Eq. Guinea = Equatorial Guinea, R. Congo = Republic of Congo CAR = Central African Republic, and DRC = Democratic Republic of the Congo)

Country	Total area (in 1000's of square km)	Forest	Woodland	Parkland	Non-treed	Rural complex	Inland Water
Cameroon	466	197	68	91	61	44	5
Eq. Guinea	25	18	0	0	0	6	0
Gabon	265	210	4	1	15	31	5
R. Congo	342	208	24	7	66	32	4
CAR	620	82	301	194	25	18	1
DRC	2328	1105	463	279	187	246	47

closest to the 100 reference value was selected. In this way the number of cloud and shadow contaminated pixels was reduced and dates with forest pixels were preferentially selected. The date of the selected acquisition and the corresponding forest likelihood, cloud and shadow likelihood, and spectral band values were retained. This compositing approach was applied independently to pre-1996 and post-1996 Landsat acquisitions, providing two composited Landsat time periods or epochs.

#### 4.8. Landsat deforestation mapping

We employ a multi-date direct classification of change methodology (Bruzzone and Serpico, 1997; Coppin et al., 2004). This approach requires that training data are available at the same surface locations in all dates and that they reflect reliably the proportions of the change transitions across the landscape. Forest is directly characterized using both the pre-1996 and the post-1996 forest likelihood and spectral composited data as inputs.

Landsat training data were identified by photo-interpretation of the two composited periods to identify deforested and unchanged pixels. A total of 37,000 training pixels were defined from the equivalent of six Landsat path/rows across the three landscapes. The training data were selected without consideration of the forest likelihood values, although the cloud, water and shadow quality assessment flags were used to avoid selection of contaminated pixels. For each training pixel the composited pre-1996 and post-1996 forest likelihood and spectral band data, and their per-pixel differences and smoothed versions of the differences (generated using a  $3 \times 3$  averaging filter) were derived.

The same bagged classification tree methodology used to produce the forest likelihood results (Section 4.6) was employed. Per node likelihoods from 30 independently generated perfectly fit tree runs were used to generate an average per-pixel deforestation likelihood for each Landsat pixel. The proportion of deforested to unchanged training pixels was approximately 1:3, resulting in a significantly oversampled deforestation proportion, and a positive bias of the deforestation likelihood values.

## 5. Results

### 5.1. MODIS 250 m forest non-forest mapping

The MODIS 250 m forest cover map is shown in Fig. 3b. Tabular results per country are shown in Table 3. In general, the

product captures the mosaic of human disturbance within the Basin. Zones of disturbance (rural complex) include the belt of higher population densities along the southern forest fringe, 3–5° latitude south, and in the Albertine Rift Valley along the borders of Uganda, Rwanda, Burundi and the Democratic Republic of the Congo (DRC), where rich volcanic soils are present. Disturbance within the forest massif traces road networks and is generally spatially coherent. In the transition zones north and south of the forest, settlement patterns in the form of roads and towns are clearly evident within the mosaic gallery forests, secondary grasslands, parklands and woodlands.

Comparisons were made with the Global Land Cover 2000 Africa product derived using SPOT VEGETATION data (Mayaux et al., 2004). Fig. 4 illustrates the aggregate forest cover estimates and per-pixel agreement for the 6 most forested Congo Basin countries. Comparisons were made for forest, non-forest and rural complex, as this is the key disturbance category at coarse spatial resolutions. Overall per-pixel agreement for these three classes was 82.7%. Highest disagreement per pixel was related to the differing spatial resolution of the products. The GLC2000 map has a 1 km

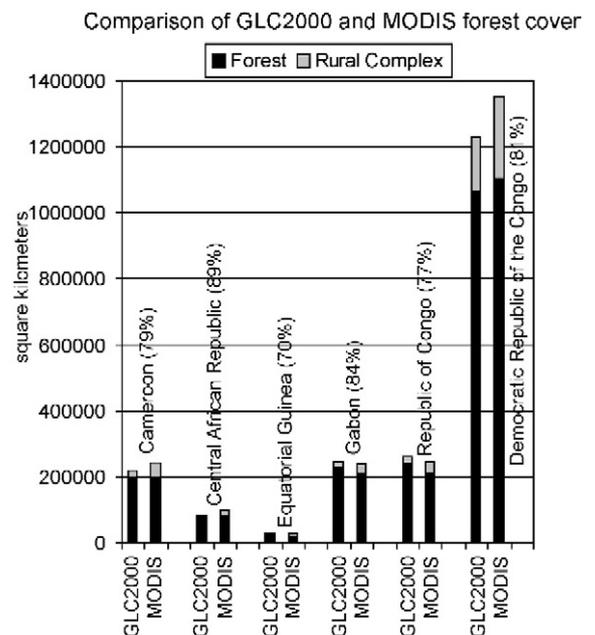


Fig. 4. Comparison of MODIS 250 m and GLC2000 1 km forest cover maps for aggregate forest and rural complex area estimates per country (black and gray bar scale) and per-pixel overall agreements for forest, non-forest and rural complex classes per country (percentages in parentheses).

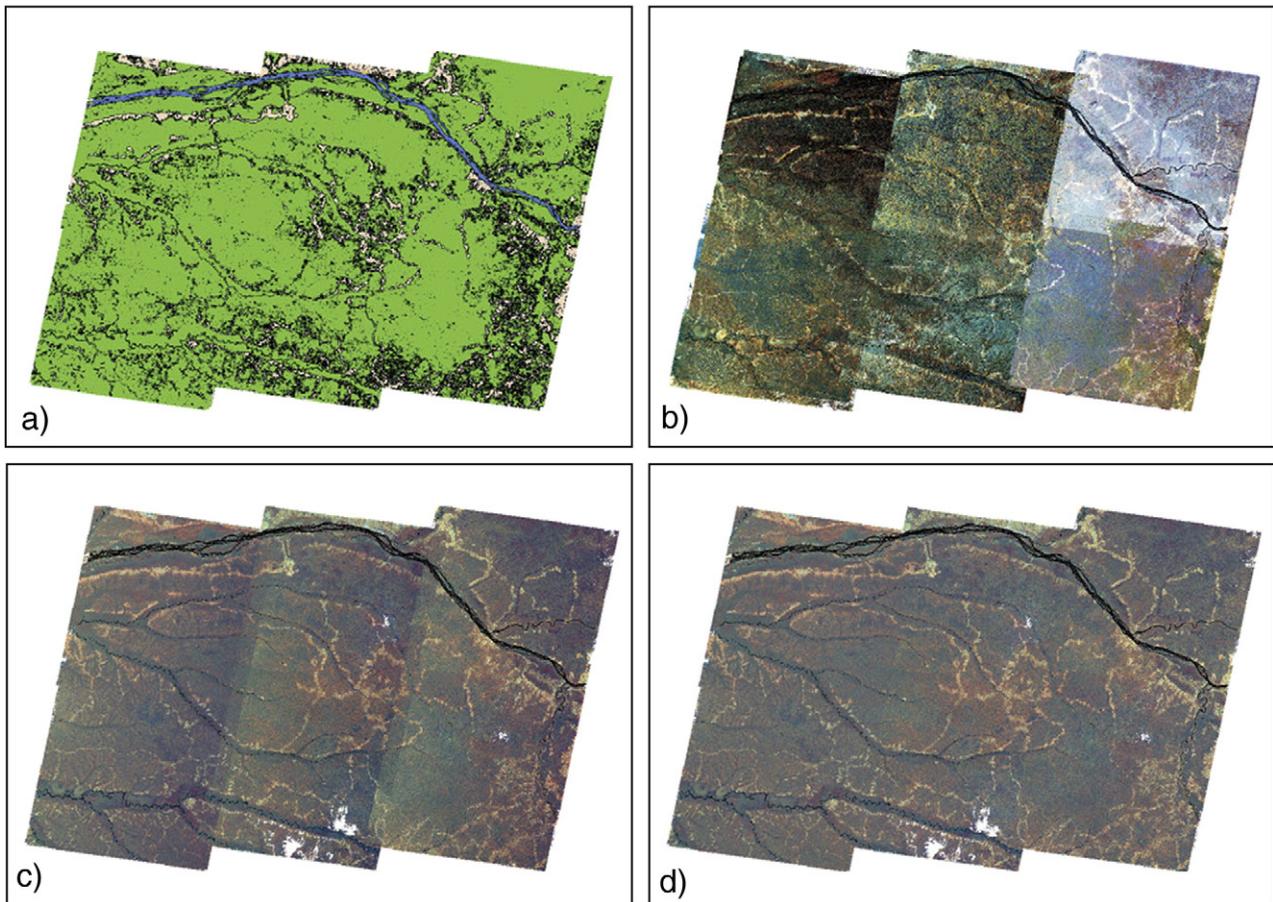


Fig. 5. Scene mosaicing after using MODIS VCF forest mask to drive Landsat interscene normalization where a) is the MODIS 250 m forest non-forest map (Green = Forest, Beige = Aggregated non-forest classes, Black = buffered border pixels, and Blue = water), b) is the Landsat digital numbers for 6 path/rows using post-1996 imagery (see Table 1), shown in bands 4 (0.78–0.90  $\mu\text{m}$ ), 5 (1.55–1.75  $\mu\text{m}$ ), and 7 (2.09–2.35  $\mu\text{m}$ ), c) is the dark-object subtraction (DOS) adjusted mosaic and d) is the DOS and anisotropy-adjusted mosaic.

spatial resolution while the MODIS map of this study has a 250 m spatial resolution. This leads to varying depictions of the rural complex class which is manifested at relatively finer scales. While forest and non-forest agreements were 85.7 and 82.2%, respectively, rural complex agreed only 46.7% of the time. Regionally, the area of greatest disagreement was in the heavily cloud-affected regions nearest the Gulf of Guinea in Cameroon, Equatorial Guinea, Gabon and the two Congos. Radar data would be a valuable alternative data source for overcoming the persistent presence of cloud cover within these areas (Saatchi et al., 2002; DeGrandi et al., 2002).

### 5.2. Landsat normalization

Fig. 5 illustrates mosaiced Landsat normalization results for six Landsat scenes covering the Maringa–Lopori–Wamba landscape (Landscape 8 in Fig. 1) acquired at different dates over a 3-year period. The MODIS 250 m forest non-forest map (Fig. 5a), top of atmosphere uncalibrated Landsat (Fig. 5b), dark-object subtraction (DOS) adjusted Landsat data (Fig. 5c), and DOS and anisotropy-adjusted data (Fig. 5d) are illustrated. Evidently, these processing steps incrementally improve the appearance of the data, providing a more coherent mosaiced

data set. To consider the quantitative impact of this processing on forest mapping capabilities, a test was applied to the forest and non-forest training data for the Maringa–Lopori–Wamba

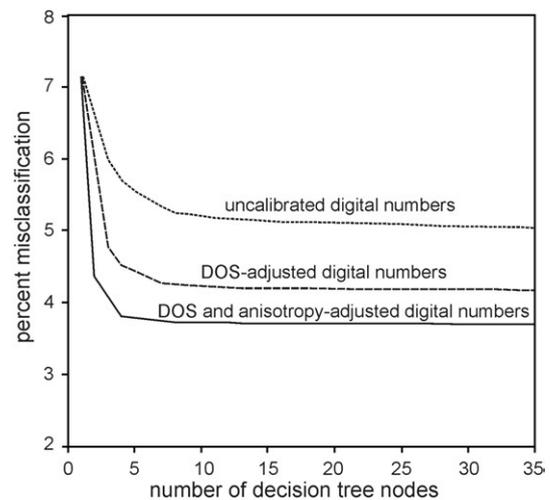


Fig. 6. Misclassification rate of MODIS labels as modeled by perfectly fit decision trees using different Landsat inputs for the Maringa–Lopori–Wamba landscape.

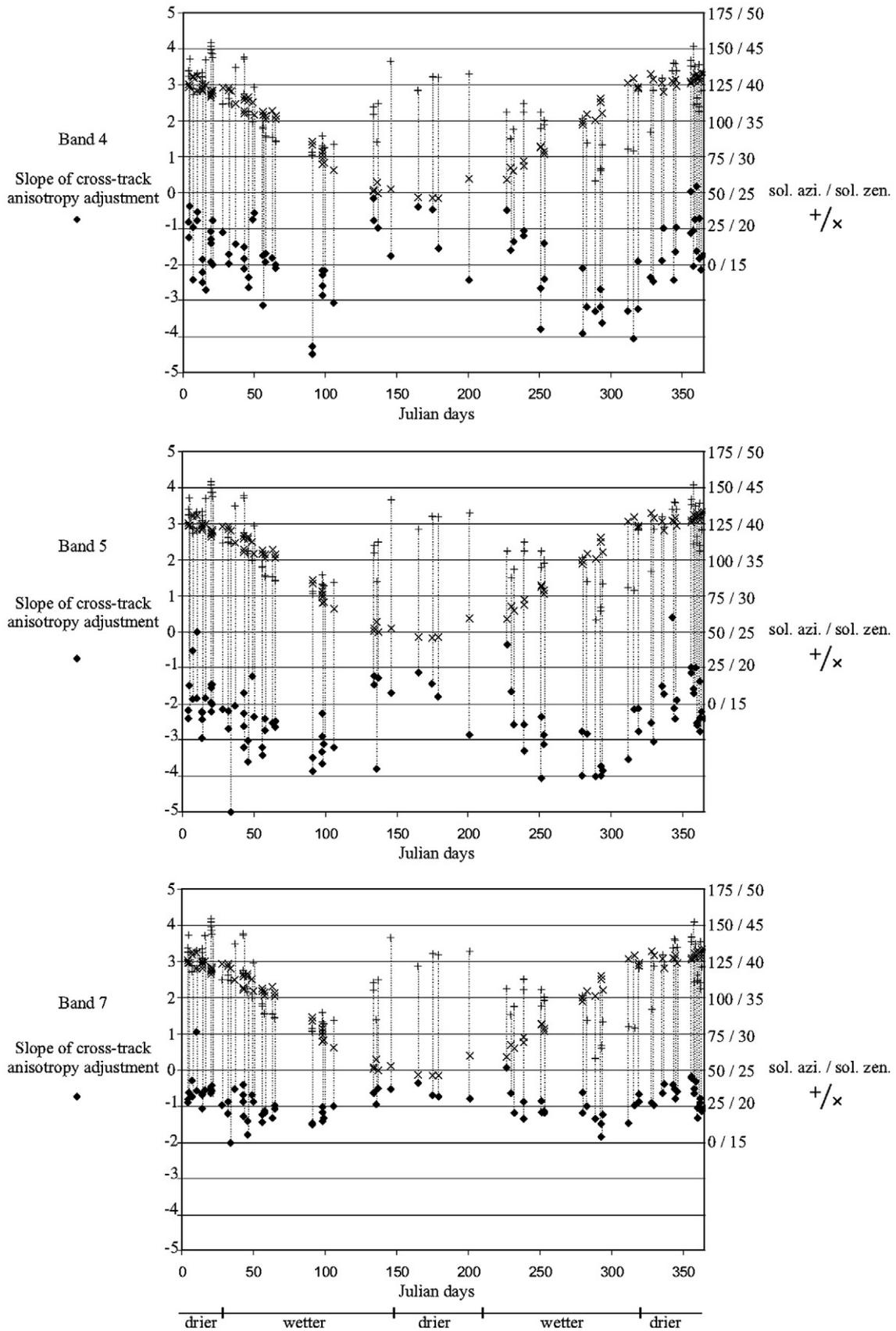


Fig. 7. Variation in the slope of the linear regression function (expressed in digital numbers per cross-track pixel), see Eq. (1), used to adjust the dark-object subtraction corrected Landsat data for anisotropic effects. Results for all the Landsat data used in the study are shown, plotting against the day of Landsat acquisition and showing the nominal solar geometry of each acquisition.

landscape. Over 400,000 MODIS labeled Landsat pixels were run through a perfectly fit decision tree algorithm to evaluate the different inputs of Fig. 5 for characterizing forest cover. The ability of the decision tree to discriminate the forest/non-forest cover categories is quantified and illustrates the increased generalization and internal consistency of the multi-spectral feature space achieved through the normalization process. Fig. 6 illustrates decreasing misclassification rates, respectively, for the top of atmosphere uncalibrated Landsat digital numbers, the DOS-adjusted, and the DOS and anisotropy-adjusted data. The DOS and anisotropy-adjusted inputs reduce the misclassification rate by one-quarter compared to the uncalibrated digital number inputs.

Fig. 7 shows the slope of the linear regression function (expressed in digital numbers per cross-track pixel location), used to adjust the dark-object subtraction corrected Landsat data for anisotropic effects, for all the Landsat data used in the study. The slopes are all significant for all Landsat scenes, with  $p$ -values of less than 0.001. The slopes are plotted as a function of the day of Landsat acquisition and with the corresponding solar geometry. There appears to be, for the reflective bands used, a seasonal variation in the strength of the relationship, with more pronounced cross-track adjustments during and near the solar equinoxes. These temporally varying effects may be related to changes in shadowing with changing solar geometry, seasonal atmospheric effects, and/or seasonal phenological variations. More study is required in order to attribute their cause(s).

### 5.3. Landsat cloud, shadow and water flagging

In generating the cloud classification tree, simple ratios of band 6 to bands 4 and 7 were the most important discriminatory variables explaining the majority of the decision tree deviance. The ratios are analogous to an albedo versus temperature measure that identifies bright, cool targets that are most likely to be clouds in the humid tropics. For the shadow model, band 5 was the primary discriminatory variable explaining more than 86% of the tree variance. Upon visual inspection, the results were found to be generally robust, save for two of the 1998 Landsat scenes, where it was necessary to manually remove shadows that remained undetected over highly reflective secondary grasslands. The vast majority of composited pixels were rated with high quality assessment states. For example, Maringa–Lopori–Wamba featured the following pre-1996 QA distribution: 93.75% low clouds/shadows, 2.04% medium clouds/shadows, 1.67% water (derived from repeated high shadow flags), 1.91% buffered high clouds/shadows, and 0.23% high clouds/shadows.

### 5.4. Landsat forest mapping

Each Landsat scene in Table 1 was processed independently using the classification tree bagging approach. Results from this mapping process were used as inputs to the subsequent compositing and change mapping procedures. As it is not feasible to

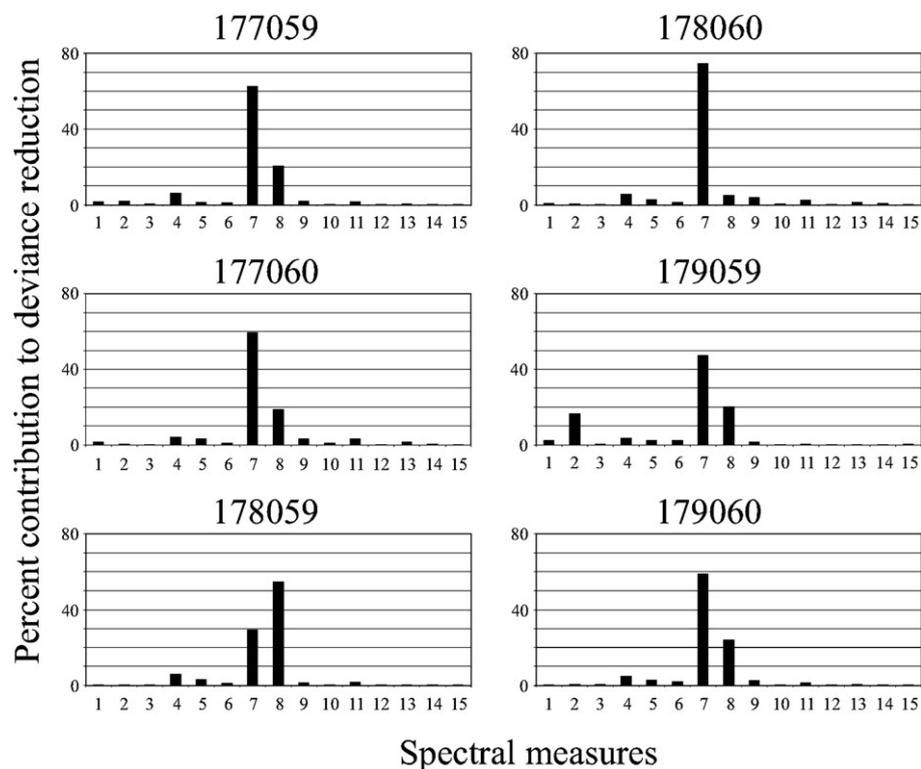


Fig. 8. Comparison of discriminatory power of spectral inputs to mapping forest likelihood per Landsat pixel as a percentage of total explained deviance from the bagged classification trees per path/row for Maringa–Lopori–Wamba. Inputs are as follows: 1–3) dark-object subtraction (DOS) adjusted bands 4, 5 and 7, 4–6) local variance of DOS-adjusted bands 4, 5 and 7, 7–9) DOS and anisotropy-adjusted bands 4, 5 and 7, 10–12) ratios of band4/band5, band4/band7, and band5/band7 of DOS and anisotropy-adjusted bands 4, 5 and 7. 13–15) ratios of local variances for band4/band5, band4/band7 and band5/band7 of DOS-adjusted bands 4, 5 and 7.

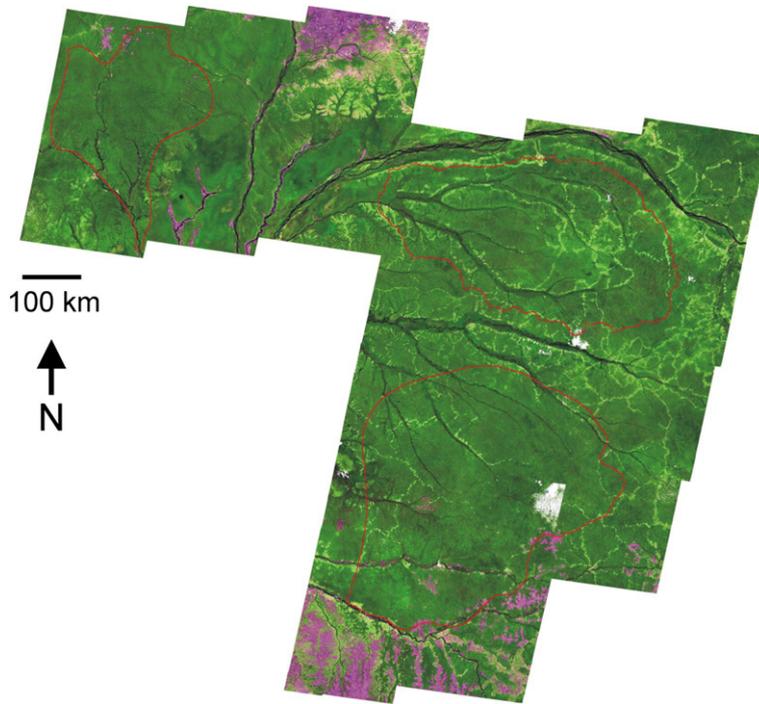


Fig. 9. Landsat composite of all post-1996 Landsat imagery for bands 5 (1.55–1.75  $\mu\text{m}$ ), 4 (0.78–0.90  $\mu\text{m}$ ), and 7 (2.09–2.35  $\mu\text{m}$ ) covering the three CARPE landscapes included in this study (outlined in red).

illustrate the results for 98 Landsat scenes, an assessment of which spectral inputs contributed most to discriminating between the forest and non-forest categories is presented here. For each spectral input, the deviance reduction per-split, summed over all bagged trees for all images per path/row, was derived to provide a relative measure of discriminatory strength. Fig. 8 summarizes for the Maringa–Lopori–Wamba landscape which spectral information drove the forest characterizations. These results reveal again the utility in performing the DOS and anisotropy adjustment and

the importance of the near-infrared (band 4) and mid-infrared (band 5) in characterizing forest/non-forest. Band 5 is most sensitive to the simple presence/absence of tree canopy, while band 4 provides additional information between regrowing/disturbed and intact tree stands. As the objective here is to map the mature forest lands and their modification through time, band 4 is the information source that largely drives the classification tree algorithm.

5.5. Landsat compositing

The Landsat forest likelihood images were composited following the methodology described in Section 4.7 into pre-1996

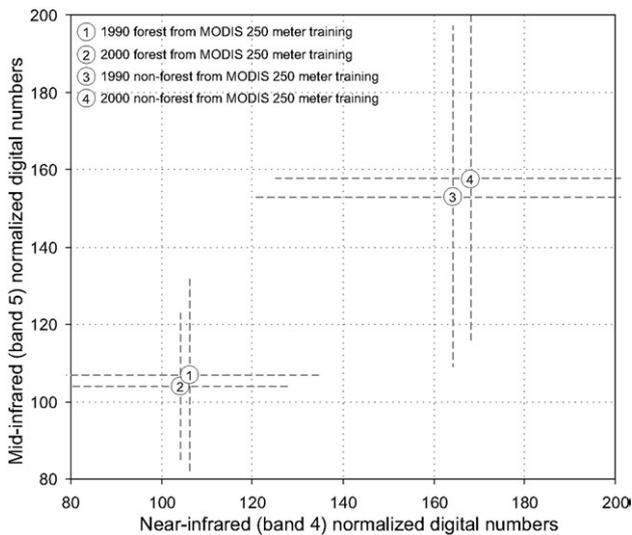


Fig. 10. Spectral plots from Landsat data for epochal 1990 and 2000 thematic classes as labeled by MODIS 250 m forest and non-forest masks. Bars represent  $\pm$  two standard deviations around mean spectral values.

Table 4

Confusion matrix indicating the level of agreement in square kilometers of MODIS-derived forest/non-forest training labels and derived Landsat forest likelihood classes thresholded at  $\geq 50\%$  likelihood for forest and  $< 50\%$  for non-forest

	Landsat non-forest	Landsat forest	Percent agreement per class	
<i>a)</i>				
MODIS non-forest	8369.6	6009.5	2360.2	71.8%
MODIS forest	125,538.1	2082.7	123,455.4	98.3%
		8092.1	125,815.6	
<i>b)</i>				
MODIS non-forest	8369.6	6199.6	2170.0	74.1%
MODIS forest	125,538.1	1878.2	123,659.9	98.5%
		8077.8	125,829.9	

Results are for the Maringa–Lopori–Wamba landscape. a) is for 1990 epochal Landsat imagery, b) is for 2000 epochal Landsat imagery. This table reflects only areas not identified as change. See Fig. 11 for a spatially explicit example.

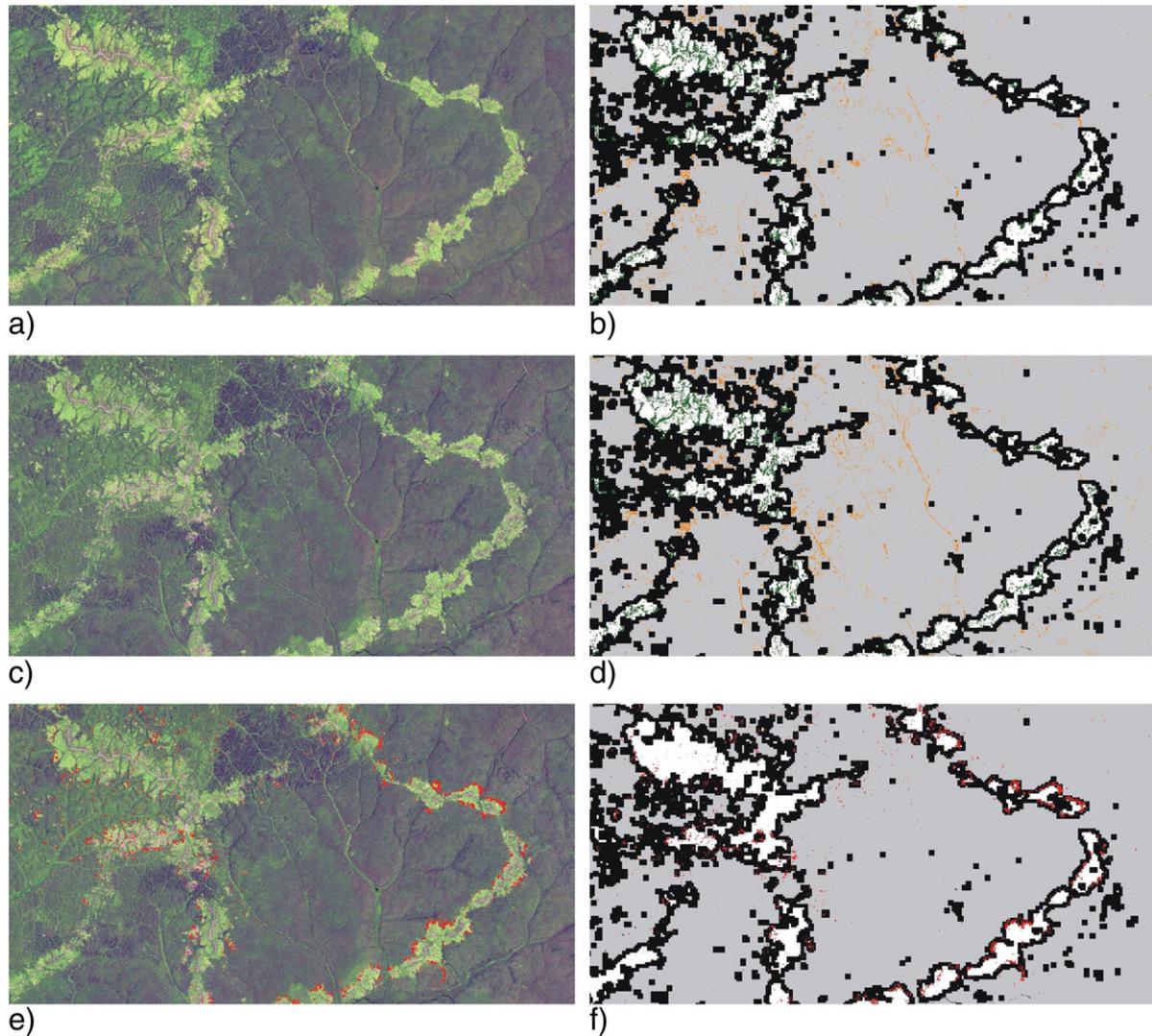


Fig. 11. Example Landsat (left column) and MODIS/Landsat cover comparison (right column) data for a 70 km by 35 km subset of the Maringa–Lopori–Wamba landscape centered on 22.73° E and 0.36° N. Landsat bands 5 (1.55–1.75  $\mu\text{m}$ ), 4 (0.78–0.90  $\mu\text{m}$ ), and 7 (2.09–2.35  $\mu\text{m}$ ) are shown in the left column where a) is 1990 composite and c) is 2000 composite imagery. The MODIS/Landsat cover comparison data in the right column are a visual confusion matrix for b) 1990 and d) 2000, colored as follows: Black — Eroded MODIS map pixels not used in training; Gray— MODIS forest/Landsat forest likelihood  $\geq 50\%$ ; White— MODIS non-forest/Landsat forest  $< 50\%$  likelihood; Dark green on white background— MODIS non-forest/Landsat forest  $\geq 50\%$ ; Orange on gray background— MODIS forest/Landsat forest  $< 50\%$  likelihood. At bottom, e) is 2000 composite Landsat imagery with identified forest clearing change pixels superimposed in red and f) is MODIS forest (white) and non-forest (gray) training labels with Landsat change pixels superimposed in red.

and post-1996 epochal data sets. An example of a composite epochal image product is shown in Fig. 9 for the region including the three landscapes presented in this study (Fig. 1). This figure reveals the regional consistency achieved by the MODIS-driven per-pixel Landsat normalization and compositing procedures. Radiometric inconsistencies along Landsat scene boundaries are largely absent and the data are largely cloud free except for those locations that were persistently cloudy in every Landsat scene acquisition. The patterns of human settlement and road infrastructure through the forest massif (bright green—high reflectance) and the savannas and grasslands to the North and South (magenta) are clearly evident. The most significant deleterious result from the compositing process is the rare preferential selection of unflagged shadows occurring over non-forest cover types (Section 5.3).

Fig. 10 shows the Landsat forest and non-forest cover training spectral signatures (means and standard deviations) derived under the MODIS 250 m forest/non-forest labels for the pre-1996 and post-1996 Landsat epochs covering the 3 landscapes. Although this description of the spectra greatly oversimplifies their actual distributions, it is evident that the 250 m MODIS labels do capture spectrally distinct forest and non-forest cover types at the Landsat scale across the region. To check this, a confusion matrix summarizing the level of agreement between the MODIS 250 m forest/non-forest classification training labels and the corresponding composited Landsat forest likelihood values was generated. Table 4 summarizes confusion matrices for the pre-1996 and post-1996 composites of the Maringa–Lopori–Wamba landscape. Each confusion matrix was generated by comparing every 57 m Landsat pixel with the corresponding 250 m MODIS pixel sub-

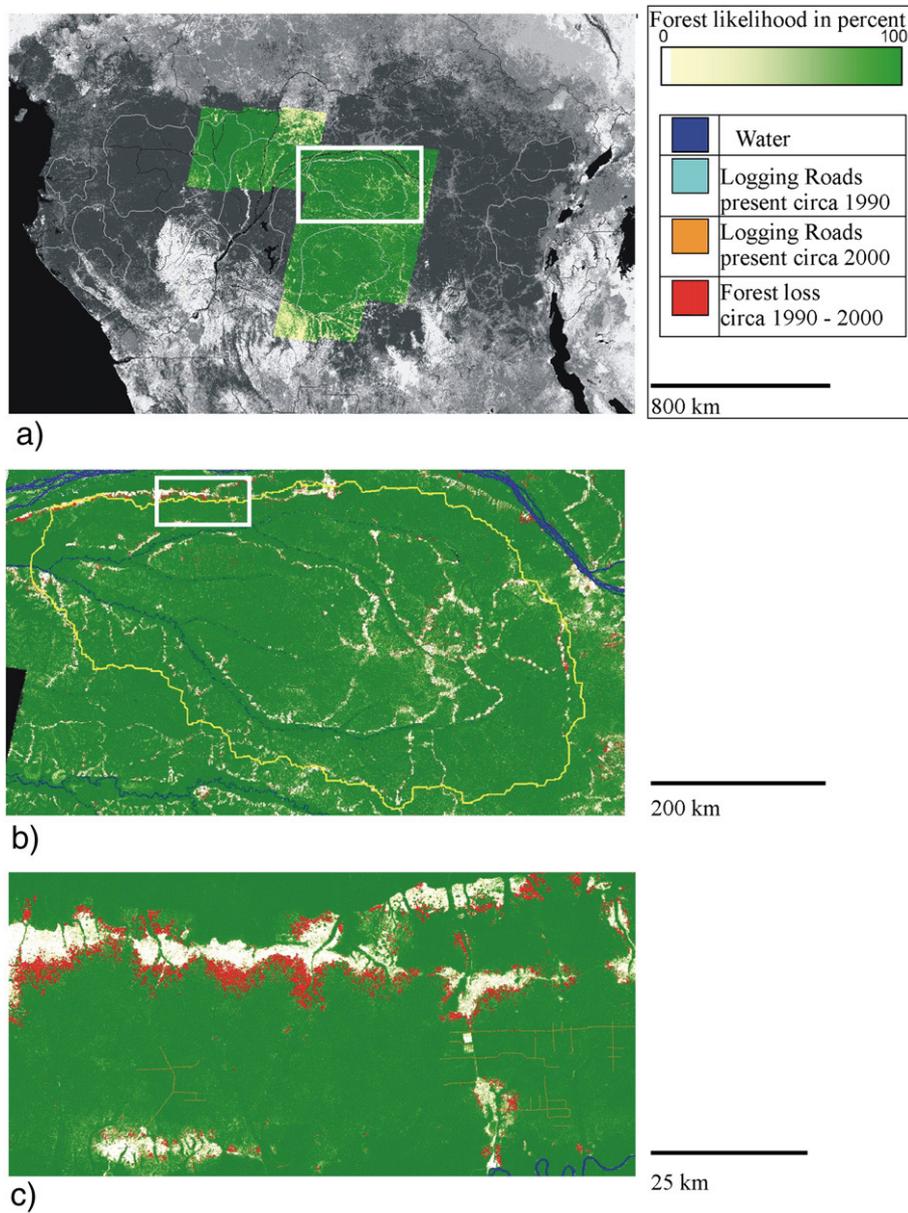


Fig. 12. Forest likelihood and forest change map for Maringa–Lopori–Wamba, Sangha Tri-National, and Salonga–Lukenie–Sankuru CARPE landscapes, where a) is Congo Basin overview with MODIS map in grayscale, b) is a zoom on the Maringa–Lopori–Wamba product, and c) is a full-resolution zoom on a locale in the north of this landscape.

sampled to 57 m, ignoring all cloud, shadow, water and change pixels. Landsat pixels were considered as “forest” if their forest likelihood was  $\geq 50\%$ .

In the confusion matrices, (Table 4) it is evident that more than a quarter of the MODIS non-forest labels have a Landsat forest likelihood greater than 50%. An explanation of this apparent confusion is seen in Fig. 11, which illustrates the degree of agreement between the MODIS 250 m and Landsat 57 m forest/non-forest training and classification results for a  $70 \times 35$  km subset of the Maringa–Lopori–Wamba landscape. The top and middle rows of Fig. 11 show imagery and forest training/mapping results for pre- and post-1996 periods respectively. The MODIS 250 m forest training labels (Fig. 11, b and d, gray and orange) perform

well as a reference source for classifying the 57 m Landsat pixels (Fig. 11, a and c), largely due to the homogeneous nature of the forest tracts. However, this is not the case for the non-forest due to the heterogeneous nature of the settled areas which include remnant riparian forests and mature tree cover within long-settled village boundaries (Fig. 11, a and c). The decision tree bagging procedure is insensitive to the significant portion of non-forest training error (apparent mislabels in the non-forest class). Consequently, at the Landsat scale, forest canopies falling within MODIS non-forest training pixels generally are labeled correctly as forest (Fig. 11, b and d, green). The decision tree classifier’s ability to not be adversely affected by this scale-related training confusion is critical to the integration of the Landsat and MODIS data.

### 5.6. Landsat forest change mapping

A final binary deforestation map was generated by thresholding the 57 m Landsat pixel deforestation likelihood values. Due to the positive bias in the frequency of deforestation training data, the binary deforestation map was generated by thresholding the deforestation likelihood values with a 30% rather than 50% threshold. This threshold was found by interactive analysis and evaluation of the model on the entirety of the input data. The composited dates for each deforested pixel were retained to enable subsequent calculation of deforestation rates.

Fig. 11, e and f, shows the pre-1996 to post-1996 Landsat forest change results (red) for the 70×35 km subset of the Maringa–Lopori–Wamba landscape. Most new change occurs adjacent to prior clearings as subsistence agricultural systems expand into primary forest zones. This expansion along roads and at settlement boundaries is a well known phenomenon in tropical systems. Small and isolated change sites are also prevalent across the landscapes, and are primarily related to resource extraction activities including mining and hunting. These findings are seen across all three mapped landscapes. Fig. 12 shows the pre- to post-1996 forest change results (red) superimposed on post-1996 forest likelihood data at three scales: all three landscapes (top), the Maringa–Lopori–Wamba landscape (middle) and a full-resolution subset of the Maringa–Lopori–Wamba landscape (bottom). The patterns of forest change in Fig. 12 are similar to those observed in the bottom of Fig. 11. In addition, linear commercial logging roads are evident (bottom Fig. 12). However, logging roads are not consistently detected as change because the road construction usually does not occur coincident with Landsat data acquisition, roads are usually considerably narrower than a 57 m Landsat pixel, and are quickly obscured by regrowing forest canopy.

Table 5 summarizes the Landsat mapped forest area and change summary statistics for the three Congo Basin landscapes; pre-1996 to post-1996. Rates of change were calculated using the composite change products and represent change over the average per change pixel time interval recorded from the composite date layers (Section 4.8). The change rates are relatively low compared to other humid forest change hot spots such as Insular Southeast Asia and the Legal Amazon (Achar et al., 2002; Hansen and DeFries, 2004). The Maringa–Lopori–Wamba landscape exhibits the most marked change of the three landscapes studied, a loss of 693 km<sup>2</sup>, equivalent to about 1% of the landscape over an approximately 13.5 year period, which is most likely related to the greater degree of human settlement in and around this landscape. The dominant change dynamic is an expansion of rotational agricultural activities along the primary trunk roads. Forest cleared for hunting camps is also depicted, and all three landscapes exhibit commercial logging activities.

Forest boundary length (edge) was calculated by creating a forest/non-forest interface using a 50% threshold of the pre- and post-1996 forest likelihood maps where only flagged change pixels are used to identify new forest edge. The forest edge criterion reveals the most dramatic change associated with logging to be found in the Sangha Tri-National landscape. While the proportion of forest area cleared from 1990 to 2000 is not as great as

Table 5

Landsat mapped forest area and change summary statistics for the three Congo Basin landscapes (Fig. 1); pre-1996 to post-1996

	Maringa–Lopori– Wamba	Salonga–Lukenie– Sankuru	Sangha Tri-National
Landscape area (km <sup>2</sup> )	74,973	101,820	36,241
Forest area circa 1990 (km <sup>2</sup> )	70,610	97,900	35,507
Forest area circa 2000 (km <sup>2</sup> )	69,918	97,522	35,357
Non-Forest area circa 1990 (km <sup>2</sup> )	4041	3415	546
Non-Forest area circa 2000 (km <sup>2</sup> )	4734	3793	696
% forest area loss 1990–2000	0.98	0.39	0.42
Forest edge 1990 (km)	75,508	66,573	16,403
Forest edge 2000 (km)	81,938	69,723	21,226
% forest edge increase 1990–2000	8.52	4.73	29.40

Mean interval for change pixels in Maringa–Lopori–Wamba is 13.54 years, Salonga–Lukenie–Sankuru is 12.18 years and Sangha Tri-National is 12.53 years. Areas are actual mapped quantities without regard to per-pixel time one and time two observation dates. Percent forest area loss has been normalized using the per-pixel observation dates to a 1990 to 2000 interval.

the other two landscapes, the wide-ranging incursion of road networks and associated extraction activities of commercial logging are clearly manifested. For this landscape, large areas of previously inaccessible forest have been opened up for exploitation. This change dynamic is captured in the forest edge metric of Table 5, where the measured forest edge within the landscape has increased by nearly 30% between 1990 and 2000.

## 6. Discussion

By relying on MODIS thematic maps as reference information, several limitations of past regional mapping procedures have been addressed, as the MODIS data allow for radiometric normalization of the Landsat inputs, and for the use of MODIS labels as regional training data. Using MODIS to normalize and label Landsat inputs allows for a standard processing stream that increases the internal consistency of the regional-scale cover characterizations. This multi-resolution methodology is portable, but requires study areas where the moderate and high spatial resolution thematic classes are the same and few in number, and where the land cover classes are spatially homogenous at scales larger than several moderate resolution pixels. In this study, these conditions are met as the Congo Basin region encompasses forest and non-forest areas that are homogeneous at the MODIS pixel scale. The standard error of the global MODIS Vegetation Continuous Fields product for sites tested to date is 11.5%, indicating an accuracy sufficient for identifying the broad tree cover strata required to map forest cover and change within the Congo Basin (Hansen et al., 2002b; Carroll et al., in review).

Aside from the relative coarseness of the moderate spatial resolution data, limitations for forest cover mapping with MODIS data typically relate to three considerations. First, persistently cloudy areas impact the quality of the MODIS input data and

resultant mapping quality. To overcome this problem, four years of inputs were used to map the Congo Basin. Second, the accuracy of the MODIS map is reduced as tree cover decreases from dense forests to woodlands and parklands (Hansen et al., 2000). Our expectation is that for densely covered intact forest biomes, such as the tropical forest realm of Central Africa, the MODIS map will be robust and sufficient as a reference data set. Third, herbaceous and shrub covered wetlands are inconsistently mapped, and can introduce error, particularly errors of change commission in herbaceous wetlands. Ongoing work is aimed at generating a wetlands mask to remove these confounding areas from the change analyses.

While this study has presented a new methodological approach to mapping Landsat-scale data sets using the MODIS percent tree cover products for calibration, the accuracy of the results have not been fully assessed. Limitations of Landsat-scale mapping relate to errors of change commission within herbaceous wetlands and to isolated residual cloud and shadow effects. However, many aspects of the study provide confidence in the methods. The increased separability and training accuracy of forest and non-forest classes due to the dark-object subtraction (DOS) and anisotropy adjustments indicates an improved mapping capability with Landsat data. The additional fact that the classification tree models primarily employ the DOS and anisotropy-adjusted bands to map forest and non-forest at the Landsat scale also validates the MODIS labels and pre-processing steps. The ability to generate regional seamless mosaics points the way towards mass processing of Landsat-scale images for land cover characterization applications.

## 7. Conclusion

This paper presents a multi-resolution methodology to map forest cover and deforestation at Landsat scale, demonstrating it for 0.56 million km<sup>2</sup> of the Congo River Basin. Typical Landsat-scale studies use a single “best” image to map forest cover state for a given year or decade. This is arguably due to high data costs and the difficulty of combining multi-temporal Landsat acquisitions. The method described here demonstrates an ability to automatically process multi-resolution, multi-date imagery. Initial results indicate that the Congo River Basin is an ecosystem absent of the large-scale clearing found in other humid tropical forest zones, such as the Legal Amazon and Insular Southeast Asia (Achard et al., 2002; Hansen and DeFries, 2004). However, the mapped region indicates that forest clearing is spatially pervasive and fragmented, with significant implications for sustaining the region’s biodiversity. Frontier forests absent of human impact are not widespread in the Congo Basin, unlike some remaining forest tracts in the interior of the Amazon or New Guinea highlands. Given this fact, identifying new incursions into remaining intact forests is important for CARPE project partners. To this end, the current approach is being extended to 2005 using Landsat 7 Scan Line Corrector-off data.

The method described here is meant to be an operational alternative to large area deforestation mapping at high spatial resolutions. Full automation is primarily a function of the quality of the input imagery. If there is a limitation regarding the usability of the results that is directly related to image quality, another

image is purchased and placed in the processing stream. The approach raised a current major limitation of large area monitoring with high spatial resolution data, namely the cost of imagery. Archives of high resolution data, whether SPOT HRV, IRS-LISS, or especially Landsat, are extensive but underutilized for this purpose due primarily to cost constraints. Operational environmental monitoring, a term often used as an objective within the remote sensing science community, is not feasible given the current cost structure for high spatial resolution data. For remote areas such as the Congo Basin, earth observation data are the only source of information for documenting land cover and land use change. Improving data access, minimizing data costs and operationalizing sensor missions is the only way to ensure timely and accurate monitoring of such areas.

## Acknowledgments

Funding from the National Aeronautics and Space Administration supported this research under grant NNG06GC416.

## References

- Achard, F., Eva, H. D., Stibig, H. -J., Mayaux, P., Gallego, J., Richards, T., et al. (2002). Determination of deforestation rates of the world’s humid tropical forests. *Science*, 297, 999–1002.
- Arvidson, T., Gasch, J., & Goward, S. N. (2001). Landsat 7’s long-term acquisition plan — an innovative approach to building a global imagery archive. *Remote Sensing of Environment*, 78, 13–26.
- Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, 22, 3855–3862.
- Asner, G., Knapp, D., Broadbent, E., Oliveira, P., Keller, M., & Silva, J. (2005, October 21). *Selective logging in the Brazilian Amazon*. *Science*, Vol. 310 (5747) (pp. 480–482).
- Avissar, R., & Werth, D. (2005). Global hydroclimatological teleconnections resulting from tropical deforestation. *Journal of Hydrometeorology*, 6, 134–145.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 26, 123–140.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Monterey, California: Wadsworth.
- Bruzzone, L., & Serpico, S. B. (1997). Detection of changes in remotely-sensed images by the selective use of multi-spectral information. *International Journal of Remote Sensing*, 18, 3883–3888.
- Carlotto, M. J. (1999). Reducing the effects of space-varying, wavelength-dependent scattering in multispectral imagery. *International Journal of Remote Sensing*, 20, 3333–3344.
- Carroll, M., Townshend, J., Hansen, M., DiMiceli, C., Sohlberg, R., & Wurster, K. (in review). Vegetative Cover Conversion and Vegetation Continuous Fields. In B. Ramachandran, C. Justice & M. Abrams (Eds.) *Land Remote Sensing and Global Environmental Change: ASA’s EOS and the science Of ASTER and MODIS*. New York: Springer Verlag.
- (CBFP) Congo Basin Forest Partnership. (2005). The forests of the Congo Basin: a preliminary assessment, Eds. Justice, C.O., et al. USAID, CARPE.
- (CBFP) Congo Basin Forest Partnership. (2006). In D. Devers, & Vande weghe (Eds.), *The forests of the Congo Basin: State of the Forests 2006*: CBFP.
- Chavez, P. S., Jr. (1996). Image-based atmospheric corrections-revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62(9), 1025–1036.
- Cihlar, J. (1994). Detection and removal of cloud contamination from AVHRR images. *I.E.E.E. Transactions on Geoscience and Remote Sensing*, 32, 583–589.
- Clark, L. A., & Pergibon, D. (1992). Tree-based models. In T. J. Hastie (Ed.), *Statistical Models in S*. Pacific Grove, CA: Wadsworth and Brooks.
- Coppin, P., Jonckheere, I., Nackaerts, K., & Muys, B. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, 25(9), 1565–1596.

- Curran, L., & Trigg, S. (2006). Sustainability science from space: Quantifying forest disturbance and land-use dynamics in the Amazon. *Proceedings of the National Academy of Sciences*, 103(34), 12663–12664.
- Danaher, T., Wu, X., & Campbell, N. (2001, July). *Bi-directional reflectance distribution function approaches to radiometric calibration of Landsat TM imagery*. Proceedings of IGARSS 2001.
- DeFries, R., Hansen, M., Steiner, M., & Dubayah, R. (1997). Subpixel forest cover in Central Africa from multisensor, multitemporal data. *Remote Sensing of Environment*, 60, 228–246.
- DeGrandi, G. F., Mayaux, P., Malingreau, J. P., Rosenqvist, A., Saatchi, S., & Simard, M. (2002). New perspectives on global ecosystems from wide-area radar mosaics: Flooded forest mapping in the tropics. *International Journal of Remote Sensing*, 21, 1235–1250.
- FAO (Food and Agriculture Organization). (2005). *State of the World's Forests*, Rome. 320 pp.
- Freidl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399–409.
- Hansen, M. C., & DeFries, R. S. (2004). Detecting long term global forest change using continuous fields of tree cover maps from 8km AVHRR data for the years 1982–1999. *Ecosystems*, 7, 695–716.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21, 1331–1364.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Carroll, M., & Dimiceli, C. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sensing of Environment*, 83(1&2), 303–319.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Marufu, L., & Sohlberg, R. (2002). Development of a MODIS percent tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, 83(1&2), 320–335.
- Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, 7(10) 15 pp. [online journal].
- Hansen, M., Dubayah, R., & DeFries, R. (1996). Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, 17(5), 1075–1081.
- Helmer, E. H., & Ruefenacht, B. (2005). Cloud-free satellite image mosaics with regression trees and histogram matching. *Photogrammetric Engineering and Remote Sensing*, 71, 1079–1089.
- Holben, B. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7, 1417–1434.
- IGBP (Terrestrial Carbon Working Group). (1998). The terrestrial carbon cycle: Implications for the Kyoto Protocol. *Science*, 280, 1393–1394.
- INPE (Instituto Nacional de Pesquisas Especiais). (2002). *Monitoring of the Brazilian Amazonian Forest by Satellite, 2000–2001* São José dos Campos, Brazil: Instituto Nacional de Pesquisas Especiais.
- Irish, R. I., Barker, J. L., Goward, S. N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric Engineering & Remote Sensing*, 72, 1179–1188.
- Ju, J., & Roy, D. P. (in press). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment*.
- Kennedy, R. E., & Cohen, W. B. (2003). Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, 24, 3467–3490.
- LaPorte, N., Heinicke, M., Justice, C. O., & Goetz, S. (1998). A new land cover map of Central Africa derived from multiresolution, multi-temporal satellite data. *International Journal of Remote Sensing*, 19, 3537–3550.
- LaPorte, N., Stabach, J., Grosch, R., Lin, T., & Goetz, S. (2007). Expansion of industrial logging in Central Africa. *Science*, 316, 1451.
- Mayaux, P., Bartholomé, E., Fritz, S., & Belward, A. (2004). A new land-cover map of Africa for the year 2000. *Journal of Biogeography*, 31, 861–877.
- Mayaux, P., Richards, T., & Janodet, E. (1999). A vegetation map of Central Africa derived from satellite imagery. *Journal of Biogeography*, 26, 353–366.
- Michaelson, J., Schimel, D. S., Friedl, M. A., Davis, F. W., & Dubayah, R. O. (1994). Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *Journal of Vegetation Science*, 5, 673–696.
- Ouaidrari, H., & Vermote, E. F. (1999). Operational atmospheric correction of Landsat TM data. *Remote Sensing of Environment*, 70, 4–15.
- Rabus, B., Eineder, M., Roth, A., & Bamler, R. (2003). The shuttle radar topography mission— A new class of digital elevation models acquired by spaceborne radar. *Photogrammetric Engineering and Remote Sensing*, 57, 241–262.
- Rasool, S. I. (1992). Requirements for terrestrial biospheric data for IGBP Core Projects. IGBP-DIS Working Paper #2, International Geosphere-Biosphere Programme — Data and Information Systems, Paris, France.
- Roy, D. (1997). Investigation of the maximum normalised difference vegetation index (NDVI) and the maximum surface temperature (Ts) AVHRR compositing procedures for the extraction of NDVI and Ts over forest. *International Journal of Remote Sensing*, 18, 2383–2401.
- Roy, D. (2000). The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 2017–2032.
- Roy, D. P., Lewis, P., Schaaf, C., Devadiga, S., & Boschetti, L. (2006). The global impact of cloud on the production of MODIS bi-directional reflectance model based composites for terrestrial monitoring. *IEEE Geoscience and Remote Sensing Letters*, 3, 452–456.
- Saatchi, S. S., Nelson, B., Podest, E., & Holt, J. (2002). Mapping land cover types in the Amazon Basin using 1 km JERS-1 mosaic. *International Journal of Remote Sensing*, 21, 1201–1234.
- Schaaf, C., Gao, F., Strahler, A., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, albedo and nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1&2), 135–148.
- SCBD (Secretariat of the Convention on Biological Diversity). (2001). *The Value of Forest Ecosystems*, Montreal. 67 pp.
- Serra, J. (1982). *Image analysis and mathematical morphology* New York: Academic.
- Skole, D., & Tucker, C. (1993). Evidence for Tropical deforestation, fragmented habitat, and adversely affected habitat in the Brazilian Amazon:1978–1988. *Science*, 260, 1905–1910.
- Toivonen, T., Kalliola, R., Ruokolainen, K., & Malik, R. N. (2006). Cross-path DN gradient in Landsat TM imagery of Amazonian forests: a challenge for image interpretation and mosaicking. *Remote Sensing of Environment*, 100, 550–562.
- Townshend, J. R. G., Bell, V., Desch, A., Havlicek, C., Justice, C., Lawrence, W. L., et al. (1995). The NASA Landsat pathfinder humid tropical deforestation project. *Land satellite information in the next decade, ASPRS conference, Vienna, VA, 25–28 Sept. 1995* (pp. IV–76–IV–87).
- Townshend, J. R. G., Justice, C. O., Gurney, C., & McManus, J. (1992). The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 1054–1060.
- Tucker, C. J., Grant, D. M., & Dykstra, J. D. (2004). NASA's global orthorectified Landsat data set. *Photogrammetric Engineering and Remote Sensing*, 70, 313–322.
- UNFCCC. (2005). *Reducing Emissions from Deforestation in Developing Countries: Approaches to Stimulate Action—Draft Conclusions Proposed by the President* (Bonn: UNFCCC Secretariat) available at <http://unfccc.int/resource/docs/2005/cop11/eng/102.pdf>
- Vermote, E. F., El Saleous, N. Z., & Justice, C. O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: First results. *Remote Sensing of Environment*, 83, 97–111.
- Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. -L. (2002). Validation of the land surface temperature products retrieved from Terra moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, 83, 163–180.
- White, F. (1983). *The vegetation of Africa, a descriptive memoir to accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa (3 Plates, Northwestern Africa, Northeastern Africa, and Southern Africa, 1:5,000,000)*. UNESCO, Paris.
- Wilkie, D., & Laporte, N. (2001). Forest area and deforestation in Central Africa: Current knowledge and future directions. In W. Weber, L. White, A. Vedder, & L. Naughton-Treves (Eds.), *African Rain Forest Ecology and Conservation* (pp. 119–139). Yale University Press.

- Wolfe, R., Roy, D., & Vermote, E. (1998). The MODIS land data storage, gridding and compositing methodology: L2 Grid. *IEEE Transactions on Geoscience and Remote Sensing*, 36, 1324–1338.
- Zhan, X., DeFries, R., Townshend, J. R. G., DiMiceli, C., Hansen, M., Huang, C., & Sohlberg, R. (2002). The 250m global land cover change product from the moderate resolution imaging spectroradiometer of NASA's Earth Observing System. *Remote Sensing of Environment*, 21, 1433–1460.
- Zhang, Q., Devers, D., Desch, A., Justice, C. O., & Townshend, J. R. G. (2005). Mapping tropical deforestation in Central Africa. *Environmental Monitoring and Assessment*, 101, 61–83.